Analysis of the Indonesian Cyberbullying through Data Mining:
The Effective Identification of Cyberbullying through Characteristics of Messages

Hendro Margono

This thesis is presented in fulfilment

of the requirements for the degree of Doctor of Philosophy

College of Engineering and Science

Victoria University

2019

Abstract

The use of social networks sites such as Facebook, Twitter, YouTube, Instagram, and LinkedIn has increased rapidly in the last decade. It has been pointed out in the international data that more than 83% of people between the age of 18 and 29 have used social networking sites (Best et al., 2014). Social networks are also a powerful medium that can be used for positive purposes, such as communication and information sharing, and can provide easy access to fresh news. On the other hand, social network sites can be used for negative purposes such as harassment and bullying. Bullying on social networks is usually called cyberbullying.

Cyberbullying has emerged as a significant issue and become an important topic in social network analysis, as more than 10% of parents globally have stated that their child has been cyberbullied (Gottfried, 2012). Ipsos reported that in Indonesia 91% of parents stated their children were bullied on social media in 2012 (Gottfried, 2012). Moreover, 58% of Indonesian adolescents ranging in age from 12 to 21 reported that they often suffered online harassment and humiliation (Dipa, 2016). Therefore, to be able to understand this phenomenon, the use of machine learning methods in data mining techniques can potentially assist in analysing cyberbullying issues. However, there are several points to be taken into consideration in the rapid use of various vocabularies for cyberbullying, the patterns of harmful words used in cyberbullying messages, and the scale of the data.

The purpose of this research is to identify the indicators of cyberbullying within the written content, and to propose and develop effective models of analysis with the goal of detecting the incidence of cyberbullying activities on social networks. Therefore, this research has addressed concerns about the measurement of cyberbullying and aimed to develop a reliable and valid measurable tool. Through developing systematic measurement and techniques, this research has enhanced an effective analysis model to discover the patterns of insulting words which can assist in accurately detecting cyberbullying messages.

The research in this thesis has developed the analysis model using association rules and classification techniques. These techniques have been used for effective identification of cyberbullying messages on social networks. Furthermore, this research has discovered interesting patterns of insulting words which can assist in identifying cyberbullying messages. The experimental results have also indicated that the proposed method can predict the messages precisely into cyberbullying or non-cyberbullying. Moreover, 80.37% of the total data has been detected as cyberbullying.

Overall, this thesis makes a significant contribution in identifying new characteristics for cyberbullying recognition, in developing the analysis method for social issues and in advancing the parameters to determine the strength of the relationship between data in relation to data mining techniques. The research in this thesis presents the analysis results and contributes to our understanding of various cyberbullying patterns. Also, the results can be developed further in future research.

Declaration

I, Hendro Margono, declare that the PhD thesis entitled “Analysing Indonesian Human Rights Issues Using data Mining Techniques” is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the awards of any other academic degree or diploma. This thesis is my own work.

Signature Date : January 2019

Herndro Margono

Acknowledgment

This thesis has been completed after a long journey with the research process and admirable support from my supervisors. As an international student for whom English is not a first language, I have experienced huge challenges in terms of reporting the results. Nevertheless, the blessings from God and the support from my supervisors have enabled me to overcome these challenges.

First and foremost, I would like to express my deep sincere gratitude to my principal supervisor Dr. Gitesh Raikundalia, and to Prof. Xun Yi who have assisted me during the hardest times, taught me many lessons and provided advice in developing my English skills and the content of this thesis so that I was able to write it scientifically and academically. I am fortunate to have such outstanding supervisors who have provided me with support in completing this thesis. They are also truly great teachers. More importantly, they have been great supporters, mentors, and companions in my PhD journey. They have always listened, reviewed my English, and given me excellent guidance throughout the difficult research process. It has been such a pleasure to work with them.

I would like to thank Russell Craig Paulet, who has helped me to develop this research through various discussions. I am also very thankful to Dr. Zion, the Victoria University Human Research Ethics Committee Chair, for approving both my human ethic applications (HRE13-077 and HRE15-211). Next, I would like to thank Angela Rojter, Petre Santry, Bruna Pomella and Dr. Kate White who have proofread and edited this thesis in accordance with English academic standards.

Special thanks go to all my family members, especially my beloved wife and my children who have been patient enough and faithful to me in completing my studies

in Melbourne. I would also like to thank my Indonesian companions who also students are undertaking different courses at Victoria University, Melbourne, who have spared their time to share their knowledge in small group discussions organised each month. Lastly, I would like to thank my PhD sponsor, the Indonesian Directorate General of Higher Education (DIKTI).

In addition, this thesis is expected to be useful not only for the author, but also for the readers. For this reason, constructive and thoughtful suggestions are welcome.

January, 2019

Hendro Margono.

Papers Published during the Author's Candidature

Conference

- H Margono, X Yi, G Raikundalia, "Mining Indonesian cyber bullying patterns in social networks", Thirty-Seventh Australasian Computer Science Conference. 20t-23 January 2014, Auckland, New Zealand.

Journal

- H Margono, X Yi, G Raikundalia, "Using Association Rules Mining to Analyse Human Rights Violations in Indonesia", *International Journal of Computer Science and Electronics Engineering* (IJCSEE) Volume 1, Issue 1, 2013. ISSN 2320–4028 (Online).
- H Margono, X Yi, G Raikundalia, "Clustering Indonesian Cyberbullying Words in Social Network". *International Journal of Computer Science and Electronics Engineering* (IJCSEE) Volume 3, Issue 4, 2015. ISSN 2320–4028 (Online).

Table of Content

List of Figure

List of Table

Chapter 1
Introduction

In recent decades, people have used the Internet and social networks as a part of their daily communication as a medium to exchange information (Subrahmanyam et al., 2008; Lenhart et al., 2009; Duggan et al., 2015). By using social networks, users can create their personal profiles for other users to view either privately or publicly (Ellison, 2007). With the social networks, users can interact with one and another.

As Information and Communication Technology is developing rapidly, the number of people using media communication technologies is increasing globally, including in Indonesia. The Association of Indonesian Internet Service Providers (APJII) is one of the telecommunication companies in Indonesia that provide Internet services. APJII affirmed that more than 139 million Indonesians accessed the Internet in 2015 (Marius and Anggoro, 2016) and that over 129.2 million Indonesians used the Internet for communication in 2015, specifically through social media (Marius and Anggoro, 2016). This indicates that Indonesia has increased its use of social media as a means of communication and sharing information.

As a medium of online communication where people interact with each other over short or long distances, social networks convey information in the form of texts, images and videos as an expression and communication of their users' ideas. However, the social interaction among people on social networks can lead in different directions, either to the improvement of the relationships between them or to differences of ideas, thoughts and opinions (Ellison, 2007). The expression of different ideas and opinions may have positive outcomes in that it may enable people to become more accepting, respecting and tolerant of those who are different while

enriching knowledge, learning diversity and understanding from different perspectives. However, if people are intolerant of and cannot respect others' differences, this can give rise to social issues such as bullying, intolerance, harassment and so on. Social networks can also be utilised by people who often anonymously publish hateful or discriminatory comments, and intimidate and harass a targeted victim (Cortis and Handschuh, 2015). Hence, this misuse of social media for the purpose of harming others needs to be taken seriously.

Since social networks enable people to establish and expand their own network, which permits them to interact openly and freely, social networks have rapidly become the medium for cyberbullying. Cyberbullying is understood to be intentional actions that are usually carried out repeatedly against a defenceless victim by an individual or a group using media communication technologies (Smith et al., 2008). Moreover, many cyberbullies (predators) believe that such despicable acts are entertaining, unaware that their behaviour may eventually have repercussions for themselves (Campbell, 2005). All of their online posts may rebound and inflict their bad behaviour on the communities in which they live (Gillespie, 2006). Cyberbullying can also inflict great damage on the victims, mostly targeted teens. Anxiety, depression and suicidal thoughts (and actual suicide) can result from cyberbullying (Donegan, 2012). It should be noted that once things have been posted online they may never vanish and can re-emerge in the future to re-inflict pain (Dredge et al., 2014).

Cyberbullying messages can be identified from their characteristics: the anonymity of the perpetrator (O'Brien and Moules, 2010), the timing of sending and re-sending the harmful messages (Privitera and Campbell, 2009), the intention of re-sending the cyberbullying messages (Spears et al., 2009), and the actual content of

the messages (Mishna et al., 2009). The content of the messages usually contains offensive words that tend to be cyberbullying messages. Therefore, the detection of cyberbullying messages by the content has become a significant challenge for researchers and practitioners who have an interest in stopping the spread of cyberbullying.

Several researchers have attempted to curb the spread of cyberbullying by proposing ways to detect negative messages on social networks, such as sentiment analysis and opinion mining (Pang and Lee, 2008; Liu, 2012), negative opinion extraction (Lo and Potdar, 2009), sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining (Aggarwal and Zhai, 2012). These studies categorised message content as either negative or positive. Moreover, other studies developed machine learning to detect content messages that are potentially instances of cyberbullying. Their machine learning techniques include detecting sensitive messages sent via YouTube (Dinakar et al., 2011), detecting cyberbullying messages through sentiment analysis of the messages (Sanchez and Kumar, 2011), identifying the cyberbullying messages from Formspring.me (Reynolds et al., 2011; Kontostathis et al., 2013) and classifying sentiment messages from Twitter (Kasture, 2015; Nalini and Sheela, 2015). These research efforts have successfully detected cyberbullying messages using data mining classification techniques. However, this previous research has focused on messages sent using the English language. Few studies have considered the analysis of cyberbullying messages that are written in an Asian language, particularly in Indonesian (Naradhipa and Purwarianti, 2012).

Therefore, the main objective of this research was to develop and implement a cyberbullying analysis model to detect cyberbullying messages in respect of the Indonesian language. To achieve this, an Indonesian stem dictionary was created,

and an analysis model was developed using various data mining techniques according to established theories. The implementation of the analysis model involved applying association rules techniques of FP-growth and cosine similarity sequentially, to discover patterns of offensive Indonesian words in cyberbullying messages. Moreover, naïve Bayes, decision tree and neural network were sequentially employed to predict the occurrence of Indonesian cyberbullying messages on social networks.

1.1 Cyberbullying Issues in Indonesia

Bullying is understood to be a form of aggressive behaviour undertaken by perpetrators and performed repetitively with the aim of harming or disturbing vulnerable victims (Thomas et al., 2013). Bullying occurs when a person or a group of people intimidate and harass victims, thereby jeopardising their safety and wellbeing, both physically and psychologically. For example, in 2010 in Indonesia, a 14-year-old teenager was spotted by her parents perched on a window ledge at home, contemplating suicide. Struggling with weight problems, the teen had been constantly bullied and taunted by schoolmates. The relentless name calling and insults had caused the teen to fall into such a depressed state that she contemplated taking her life. Fortunately, she was seen by her parents who immediately sought psychiatric help (Setyawan, 2014). Years earlier, two other teens were not as fortunate – both took their lives at home due to being bullied at school. Linda, a 15-year-old junior high school student, committed suicide after being ridiculed and picked on by fellow students due to her failing junior high school classes. A 13-year-old girl hanged herself at home after being incessantly teased at school because her father was a street vendor (Setyawan, 2014).

Nowadays, bullying is also carried out through the medium of Internet, and is commonly referred to as cyberbullying (Thomas et al., 2013). Cyberbullying is a serious issue that has attracted the interest of many researchers seeking to investigate, in particular, the impacts of bullying on its victims. These impacts include social isolation, feelings of insecurity, shame and anxiety, fear, physical self-harm, loss of self-esteem and difficulties with concentration and learning (Kowalski et al., 2012; Thomas et al., 2013).

According to a survey of 40 countries conducted by Kaman (2007) in 2005-2006, Indonesia has one of the highest rates of cyberbullying. This finding was confirmed by the data obtained by Ipsos, an independent market research company which conducted a survey to estimate the rate of bullying occurring in several countries (Gottfried, 2012). They investigated bullying among approximately 200,000 school-aged children in 40 countries across the world. Ipsos found that 91% of Indonesian parents had discovered that their children were bullied on social media. In Australia, it was about 87%, Poland was around 83%, Sweden was about 85%, the United States was about 82%, and Germany was around 81%. Furthermore, most of the citizens in 24 countries stated that cyberbullying required more attention and consideration from the public, including parents, particularly in Japan (91%) and Indonesia (89%) (Gottfried, 2012).

The increasing number of cyberbullying incidents in Indonesia led Febrianti and Hartana (2014) to conduct online questionnaires with a revised cyberbullying Inventory scoring instrument and they found 77% of students from University of Indonesia were involved in cyberbullying. The dominant characteristics of perpetrators and victims involved in cyberbullying are students from the 20-25 year age group, predominantly female and having an internet usage of 21-24 hours per week. According to Febrianti

and Hartana (2014), harassment occurs more frequently in cyberbullying activities compared to flaming, stalking, outing and trickery, disguising, alienation, happy slapping and defamation.

Indonesian researcher Yulianti (2014) study of 251 Indonesian students from senior high public and private schools in Yogyakarta, ranging in age from 14 to18 years, found that bullying and cyberbullying were predominantly occurring among Indonesian adolescents. According to Yulianti (2014), there was no significant distinction between state and private schools regarding the incidence of cyberbullying. Safaria (2016) recorded that of her 102 seventh grade respondents in Yogyakarta, almost 80% had been victims of cyberbullying on Indonesian social networks. Even one publicised case in October 2010 involving a famous Indonesian actress revealed that her son had been receiving threats and insulting comments. These threats were perpetrated by the son of a military officer on Facebook (Akbar and Utari, 2014). Cyberbullying victims are not only young people. For example, the chairman of a committee responsible for organising music festivals in Jakarta deliberately crashed into a passing train. He believed that he was not strong enough to face cyberbullying on social networks, after failing to organize a particular music festival (Putra, 2014).

The United Nations Children's Fund (UNICEF) and the Indonesian Ministry of Communication's study of 400 adolescents (aged 10 to19) in 17 Indonesian provinces found that 58% of Indonesian adolescents were unaware of cyberbullying attacks (Razak, 2014). Another survey conducted by *Yayasan Cinta Anak Bangsa* noted that 58% of adolescents between the ages of 12 and 21 in Indonesia admitted that they often suffered online harassment and humiliation, and admitted not telling their parents about the cyberbullying (Dipa, 2016). The survey also found that, on average, one in eight young people were exposed to cyberbullying. This indicates that cyberbullying in

Indonesia has increased significantly each year. In addition, they have a low literacy level when it comes to using the Internet and are given inadequate protection by their parents and Government (Jong, 2016).

Based on the above mentioned research it is clear that cyberbullying cases in Indonesia have become a serious issue that requires comprehensive research. The research in this thesis explored the phenomenon of cyberbullying in more depth, specifically in terms of message content. This is due to the content of the messages containing insulting and offensive words which tend to be classified as cyberbullying messages. Also, the content of the messages can be explored in order to find meaningful objects, such as texts (Krippendorff, 2012) that eventually discover the effect of the messages. When the effect of the messages is understood, further categorisation in respect of the content of various types of message is conducted (Riff et al., 2014). This can be achieved by discovering patterns of Indonesian offensive messages found in the obtained data.

1.2 Research Aim, Questions, and Objectives

A. Research Aim

The main aim of this study is to address the problem that has been outlined above by exploring how the patterns of insulting terms on social network messages can be identified as cyberbullying messages. This is achieved through the development of an analysis method using data mining techniques. The use of existing techniques in data mining such as the association rules and classification techniques can assist in detecting cyberbullying patterns appearing in messages on social networks. Improving the design of an analysis model through the combination of existing data mining techniques can address

the shortcomings of current cyberbullying identification processes, particularly regarding messages sent in the Indonesian language.

B. Research Questions

The research aim was addressed by answering the research question. Therefore, to explore the aim of the research, the main research question in this study is:

> *Which characteristics and relationships are important for the effective identification of cyberbullying messages in Indonesian social media data?*

The main research problem is addressed by examining the issues extensively. The main question is divided into two sub-questions:

1. Sub-question 1. How to identify the patterns of insulting words in social network messages, and the techniques that can be used to do so? In this case, any message sent through a social network containing some insulting words may be related to other insulting words. The relationship of these words will form a pattern of word relationships that make up a negative sentence that ultimately indicates the content of cyberbullying message. For example, a user who sends bullying messages to a certain victim will often encourage his or her followers, friends and other users to send bullying messages. They are able to do this by re-sending the same message(s) or creating new messages that contain similar bullying words. Thus, the probability of one message being connected to another is significant.

2. Sub-question 2. How to detect the cyberbullying messages on social networks and what techniques to use for detection? The prediction of a class label for the messages is essential in understanding whether and how the messages can be classified. Although messages that are sent on social networks may contain insulting words, they may not necessarily be classified as cyberbullying messages. This is because messages may be expressing the sender's own feelings of anger and stress and may contain offensive terms that are self-deprecatory, and not directed at others. Such messages are often sent to close friends. For example, the messages containing offensive terms that are sent via a social network are not necessarily cyberbullying messages. The offensive terms may be nicknames, teasing, or complaints about oneself.

C. Research Objectives

To achieve the aim of this research and answer the research question, this study has three objectives, each of which has its own function. The first objective is that this research has applied the association rules techniques from Han et al. (2012) to provide theoretical and practical knowledge on how to analyse messages on social networks in order to discover trends or patterns of insulting words. Here, the goal is to discover the trend as an indicator of the patterns of past events related to cyberbullying messages. Second, since analysing the content of the messages involves the removal of affixes (suffixes and prefixes) from root words (Frakes and Fox, 2003) and reducing the number of words to obtain an exact matching word (Ramasubramanian and Ramya, 2013), this thesis has developed an Indonesian stem dictionary in order to

condense the common form of words to their morpheme root form by stripping each word of its derivational and inflectional suffixes. Finally, as a limited number of detection process has yet to be applied to social network messages written in the Indonesian language, the author of this thesis has applied classification techniques to accurately discover the probability values of the messages and accurately place them into cyberbullying or non-cyberbullying classes on a local database.

1.3 Motivation

Generally, the analysis of Indonesian cyberbullying issues is conducted to find hidden interesting relationships in the content of cyberbullying messages that can lead to identifying links between messages. The relationship between them will form a pattern of the message content which reflects the characteristics and features of the Indonesian cyberbullying messages, so that the messages containing these patterns will be recognized as being either cyberbullying or non-cyberbullying. Generally speaking, the analysis of cyberbullying messages on social networks is intended to enable the early detection of harmful cyberbullying messages. To achieve this goal, data mining techniques are used as an effective tool for the in-depth exploration of data collected from Indonesia's cyberbullying message patterns on social networks.

The types of data mining techniques to be used for the analysis of cyberbullying messages depend on the chosen techniques suitable for the task. For example, the association rules technique is used when one wants to discover relationships between data items and analyse data patterns. Hence, association rules techniques will be incorporated into the analytical model proposed in this research. The model also involves data input, cleaning of data, analysis process and the output of analysis. This

model includes several data mining techniques used, each with its own function. The purpose of using multiple techniques is to obtain interesting patterns, which are a part of knowledge discovery. The sequence approach primarily focuses on finding relevant patterns between data samples that are effective in obtaining frequently occurring patterns and data similarity comparisons.

The overview of the process of analysing cyberbullying messages is as follows:

- Input. The data that will be analysed originated from tweets sent via Twitter. In this case, the tweets downloaded from Twitter are messages containing Indonesian insulting words.
- The analysis process. The analysis process is a necessary stage in determining the cyber bullying patterns, the characteristics and the features of the cyber bullying messages, and the prediction of the message classifications of either cyber bullying or non-cyberbullying. The use of data mining techniques is required at this stage.
- Output. The results provide valuable information and will be presented in structures of trend data as data patterns, and predictions of class labels of cyber bullying messages which identify the messages as either cyberbullying or non-cyberbullying.

In respect of the analysis process outlined above, the research of this thesis will develop an analysis model and examine the model for various and actual cyberbullying messages in the Indonesian context.

1.4 Research Gap

There has been much recent interest in reports on cyberbullying via social network sites due to the increase in the incidence of cyberbullying, as discussed above

in section 1.1. The cyberbullying issue is in need of serious attention from the public because of its harmful and sometimes devastating impact on victims, particularly adolescents. According to research findings, cyberbullying of students may lead to lack of concentration, absenteeism (Beran and Li, 2005), aggression, depression (Gradinger et al., 2009), mental health issues, substance abuse and suicidal thoughts (Goebert et al., 2011; O'Brien and Moules, 2010).

Previous investigations have been conducted to identify the characteristics of cyberbullying messages; these are the anonymity of the perpetrator (O'Brien and Moules, 2010), the timing for sending and re-sending harmful messages (Privitera and Campbell, 2009), the intention of sending the messages (Spears et al., 2009), and the content of the messages (Mishna et al., 2009). The content of the messages comprising harmful, inappropriate and abusive words derived from the swear-keyword list is a form of cyberbullying known as signed cyberbullying. As the swear-keywords are present in the messages, this signifies an early detection of cyberbullying. A sentence example of "You're a no good spoiled b***h that needs to get a f***ing life because no one will ever like an idiot stupid like u so f***k yourself u no good ugly fatty wh**e" illustrates the presence of abusive words while tweeting or referring to another user that shows a direct action of cyberbullying (Nahar, 2014).

According to previous studies conducted by Ellis (2013), cyberbullying mostly occurred on Facebook, Twitter, and Ask.FM in the United Kingdom; these findings have also been supported by Hackett (2015). He found that most of his respondents had experienced cyberbullying on Facebook, YouTube, Twitter, Ask.FM. Instagram and Tumblr. This indicates that social networks sites are the more popular media for spreading cyberbullying messages. Ipsos claimed that the cyberbullying phenomenon also occurs in Indonesia where the incidence of cyberbullying reported by Indonesian

parents is considerably high. These surveys supported the findings of Gottfried (2012) survey that found most of the respondents (Indonesian parents) were aware that their children had been victimised from cyberbullying. Yulianti (2014) also confirmed that cyberbullying occurs in Indonesia and found a considerably high rate of cyberbullying cases within the Indonesian adolescent community. All these research findings suggest that the cyberbullying phenomenon has become an increasingly serious issue.

In their attempts to detect cyberbullying messages on social networks, previous researchers have successfully identified cyberbullying messages from Formspring.me using decision tree to categorise messages as either cyberbullying or non-cyberbullying (Reynolds et al., 2011). However, Reynolds et al. (2011) noted that there are numerous challenges to overcome when mining data from social networks using machine learning. Dinakar et al. (2011) successfully detected textual cyberbullying in YouTube video comments through the application of machine learning techniques. They used a range of binary classifiers to detect textual cyberbullying in YouTube comments. However, in their work, the huge corpus of data on YouTube comments still required the application of semi-supervised learning techniques.

In order to detect cyberbullying content from Formspring.me, Kontostathis et al. (2013) used the query terms which are often used by predators. Using supervised machine learning, they identified cyberbullying and accurately assigned a score to messages posted on Formspring.me. Nahar (2014) applied a fuzzy support vector machine learning approach to detect cyberbullying content on social networks based on the gathered data. However, these studies were concerned only with messages written in English; to the best of our knowledge, there are few studies that have been conducted on cyberbullying messages written in Asian languages, for examples

cyberbullying detection in Japanese (Nitta et al., 2013; Hatakeyama et al., 2016), analysis contents in Chinese social networks (Cao et al., 2014). Since Indonesia has one of the highest rates of cyberbullying among several Asian countries (Gottfried, 2012), the detection of cyberbullying messages in the Indonesian context is still a challenge and requires extensive research (Naradhipa and Purwarianti, 2012; Lunando and Purwarianti, 2013; Safaria, 2016).

1.5 Research Methodology

The methodology for this research involved quantitative measurement in order to understand the characteristics of cyberbullying messages. By implementing data mining techniques, and a direct focus on Indonesian Twitter messages, patterns of Indonesian cyberbullying messages can be discovered. The quantitative measurement used in the methodology of this research concerns the number and density of the insulting words in tweets. Reynolds et al. (2011) argue that the percentage of “bad” words in a post is an indication of cyberbullying. Therefore, the methodology of the research in this thesis has adopted a similar quantitative measurement in order to identify cyberbullying messages using Indonesian tweets as a research focus.

As the research of this thesis focuses on Indonesian tweets, messages written in other languages were excluded from the study. After the data extraction process from Twitter, the data were recorded in a local database. Several cleaning techniques such as tokenize, transform case, stop words, stem dictionary and n-gram were implemented to obtain a normalised data set that could be analysed. The data analysis were carried out using various data mining techniques (Han et al., 2012).

The methodology of this research is distributed into two stages. The initial stage involved discovering patterns of Indonesian insulting words that have the potential to be used in a cyberbullying message. This stage can be accomplished by employing association rules techniques of FP-growth and cosine similarity. The aim for the implementation of these techniques is to find interesting relationships between data items. The relationships between the obtained data formed frequent patterns that often appeared together in a data set (Han et al., 2012).

Following this first stage, the patterns discovered were further used to detect messages that tend to be cyberbullying. The second stage can be achieved through the use of naïve Bayes, decision tree and neural network techniques to predict the labels of data class. As a result, based on these analysis and the findings from the data, the final output generated a modified framework for the detection of Indonesian cyberbullying messages.

The research in this thesis has utilised relevant literature and appropriate techniques for the ultimate goal of cyberbullying detection. The implementation of relevant literature in this thesis enables the construction of an analysis model to discover patterns of Indonesian insulting words, thus cyberbullying identification.

1.6 Research Outcome

Given the explanation of the problems listed above, the results from this study will contribute to solving the problems described in the problem statement section. There are two aspects that must be considered in this research:

- Outcome 1. Discovering cyberbullying patterns from social networks. This research has created a model for analysing cyberbullying messages on social networks. Each insulting word that a predator uses in messages will be

identified through the relationships between insulting words in the messages using the association rules techniques. The relationships between the insulting words will illustrate the trend of insulting words often used by predators when bullying someone on social networks. These trends are referred to as data patterns. The patterns of insulting words are the patterns of the content in the cyberbullying messages. Hence, once discovered, these patterns can be used to identify the characteristics and features of cyberbullying messages, especially within the Indonesian context.

- Outcome 2. The prediction of class data labels from social networks as cyberbullying or non-cyberbullying. In this research, an analysis model has been designed to predict class data labels and classify them into cyberbullying or non-cyberbullying categories. This research also considered a data training set comprising data that had been labelled for their class as a reference when calculating the probability of data where the classification has been specified. This classification process calculates the probability of each message to be categorised according to its appropriate class.

1.7 Significance of the Research

The identified characteristics of cyberbullying words has a significant role in the wider implementation of methods within the detection of cyberbullying messages. Therefore, the research in this thesis makes a significant contribution in providing a broader knowledge and understanding of the characteristics of Indonesian cyberbullying messages through identifying their patterns. Therefore, it is anticipated that this research will generate a great deal of interest, not only among the research

community, but also among the general public. This study can benefit the research community as follows:

- It may encourage other researchers to develop new and more effective research methods that can be used to analyse social issues using various data mining techniques.
- It offers a significant contribution to the wider knowledge of cyberbullying characteristics in international contexts, which means languages besides Indonesian.
- This research provides interesting data that can be used in current and future cyberbullying identification processes along with the application of data mining techniques to various social issues which are linked to cyberbullying messages.

Furthermore, since the research in this thesis has employed data mining techniques, this can be seen as a significant endeavour in developing the measurements of finding relationships between data. Additionally, in one of the data mining tasks, frequent patterns is an essential result in measuring relationships between data, and thus becomes a significant theme in data mining research (Han et al., 2012) as these frequent patterns reveal the recurring relationships in a given data set.

Hence, to find frequent data patterns, the use of the association rules within their parameters is a crucial strategy to discover the associations and relationships among data in a large relational data set such as cyberbullying data from social networks. The massive amount of cyberbullying data continuously being collected and stored enables the mining of hidden information which is useful for research on the characteristics of cyberbullying. Thus, the association rules techniques are one of the appropriate data mining techniques that can be implemented to discover interesting

relationships between data (Han et al., 2012). The research in this thesis has significantly extended the measuring parameters to find interesting associations between data in a large relational data set.

The patterns discovered in this research involved the prediction task of data classes. Furthermore, developing an analysis model in this research contributes to predicting data classes, which utilise classification technique of data mining. The research in this thesis therefore has expanded the measuring parameters to determine data class from a large data set, such as cyberbullying data from social networks.

1.8 Change to the Field

The past decade has seen an unprecedented increase in the volume of data on social networks, which has significantly altered the behaviour of the online community. This social phenomenon has attracted the interest of researchers particularly regarding the analysis of message content in social networks data. Previous research has investigated and analysed the content of sentiment messages in one of the social networks sites and successfully classified these messages according to whether they are positive, negative or neutral (Agarwal et al., 2011). The positive text usually contains words that connote happiness or enthusiasm, for example, or are complimentary or kind. The neutral text contains no emotive words. However, the negative text tends to contain harmful and offensive words or sentences that indicate cyberbullying (Thomas et al., 2013).

Since cyberbullying is a social issue affecting millions of citizens worldwide, it has become an interesting and disturbing issue during the last decade, attracting much attention from researchers in the fields of social and computer science. Therefore, the

author of this research has explored the cyberbullying field by combining social and computer science disciplines.

In respect to exploring the cyberbullying field, the frequency of the messages sent to the victim is taken into consideration in identifying the characteristics of cyberbullying messages (Spears et al., 2009). Moreover, cyberbullying can be recognised through the content of messages (Mishna et al., 2009). Next, the flexible access of time and place in which direct contact and interaction with the victim are not required is one of the features of cyberbullying messages (Privitera and Campbell, 2009). In addition, the anonymity of the sender which is intended to prevent the perpetrator's identification is one of the ways to recognise cyberbullying messages (O'Brien and Moules, 2010). Therefore, identifying the characteristics of cyberbullying content in the huge volume of posts from social networks is essential as a means of revealing information and arranging the data into a neat format (Cortis and Handschuh, 2015). The research in this thesis has brought a change to the cyberbullying field by expanding the identification process of cyberbullying characteristics by insulting word patterns. This has led to the application of data mining techniques to analyse the content of cyberbullying messages, which significantly improves our knowledge of the diversity of these messages. Furthermore, the features identified in the message content of social networks has become an alternative means of detecting the characteristics of cyberbullying (Zhao et al., 2016). In respect to changes to the field, this research proposes to combine the appropriate data mining techniques that ultimately discover the insulting word patterns as cyberbullying characteristics, thus cyberbullying detection within the Indonesian context.

In order for cyberbullying data from the social networks to be transformed into interesting information, powerful and versatile tools are required such as machine

learning (Nahar, 2014). The research in this thesis has implemented machine learning from data mining techniques in analysing a social issue such as cyberbullying, which came into existence through rapid technological development in the fields of social and computer science. For example, machine learning classifications have previously been used in data mining research for detecting cyberbullying messages on social networks (Reynolds et al., 2011; Dinakar et al., 2011).

1.9 Contributions

The main contributions made by this thesis are explained below:

Contributions to theory:

- This research contributes to cyberbullying recognition frameworks by identifying new characteristics that can be used to recognise cyberbullying messages containing insulting words. The framework was adapted from an existing framework developed in a general context, and enhances our understanding of the analysis of cyberbullying content that has emerged in Indonesia.
- This research also contributes to the development of analysis models for social issues, especially cyberbullying messages on social networks. In other words, this research has developed new constructs for an analysis model that can be applied to detect cyberbullying messages. This new analysis model consistently applied several parameters to all messages in the Indonesian context.
- A further contribution is the extension of the measurement of relationships between data through the use of association rules techniques. These measurements are adapted from the existing techniques developed for the

purpose of discovering strong relationships between itemsets as interesting data patterns in the database.

- Another contribution is the expansion of the measurement of data class prediction by combining several classification techniques. This was accomplished by adapting the current techniques of naïve Bayes, decision tree, and neural network for the purpose of detecting and predicting cyberbullying messages more effectively and accurately.

Contribution to practice:

- This research contributes important information about cyberbullying especially by providing crucial data regarding cyberbullying in Indonesia. This research offers meaningful information both to the public and research communities who may be interested in analysing Indonesian cyberbullying messages from social networks sites. Moreover, the findings related to cyberbullying patterns in Indonesia can be used by public bodies or by the research community to detect harmful messages comprising insulting words within the Indonesian context.
- This research provides practical knowledge on how to analyse cyberbullying messages on social networks, along with the implementation of data mining techniques. Practitioners and future researchers can adopt the analysis model developed in this research which can be expanded and applied using other techniques and in a wider field of disciplines.
- As a further contribution, this research aims to extend the Indonesian stem dictionary that deconstructs words to their morphological root format, in particular to the Indonesian insulting words. Practitioners and future researchers can adopt a similar approach to the stem dictionary which can be extended and applied in various techniques and wider topics of words.

1.10 Thesis Organisation

This thesis has six chapters. Chapter 2 presents the literature review and chapter 3 describes the analysis method for the Indonesian cyberbullying messages on social networks using the association rules techniques. Chapter 4 explains the method of detecting cyberbullying messages using classification techniques. Chapter 5 discusses the findings in relation to the theory and research questions. Chapter 6 concludes the thesis. The structure of this thesis is intended to provide readers with an adequate background knowledge of the phenomenon of cyberbullying, enabling them to understand and appreciate the contribution of this thesis. The thesis is structured as follows:

- Chapter 2 offers an overview of the literature that is required for analysing the phenomenon of cyberbullying occurring on social networks. This chapter begins by explaining the characteristics of cyberbullying and describing cyberbullying cases in Indonesia. In addition, the chapter observes the patterns of Indonesian cyberbullying messages in social networks and explores the literature related to the various data mining techniques implemented to analyse cyberbullying on social networks.
- Chapter 3 describes the application of the association rules for cyberbullying analysis in the Indonesian context. This chapter explains how the association rules functions in the process. This is followed by an explanation of how the data related to Indonesian cyberbullying messages can be represented as itemsets.
- Chapter 4 explains the implementation of classification techniques for the prediction of the label data class in a database. In this chapter, three

classification techniques are adopted to ensure the accuracy and precision of the prediction of data classes.

- Chapter 5 presents the discussion of findings and their relevance to theory and the research questions. In this chapter, the practical application of the findings to future research is discussed. The development of the analysis model for this research is explained. In addition, the strengths and limitations of the research are explored in detail.
- Chapter 6 concludes the thesis by summarising and highlighting the main contributions of this study. This chapter also shows the link between chapters, beginning with a review of the problem statement, followed by the literature review revealing the main issues and leading to the solutions and contributions of this research. This chapter concludes with suggestions for future research that will further develop the findings of this study.

Chapter 2
Cyberbullying and Data Mining Review

2.1 Introduction

A new form of bullying has emerged in recent years via information technology devices, which is commonly known as cyberbullying (Campbell, 2005; Privitera and Campbell, 2009). This is a new phenomenon involving one person sending an intentionally harmful message to another person. The purpose of sending such messages is to express power, positioning the sender in a superior position over others (Dooley et al., 2009). E-mail, text, chat rooms, mobile phones, mobile phone cameras and websites are used frequently as mediums for sending insulting and offensive messages to victims of bullying. Previous research on this phenomenon has implemented three basic methods in the study of bullying; these are quantitative (e.g. the observation of motives, characteristics and types of cyberbullying (Sanders et al., 2009)); qualitative (e.g. measurements of cyberbullying perpetration and victimization (Savage, 2012)), and data mining techniques (e.g. detecting cyberbullying messages (Reynolds et al., 2011; Dinakar et al., 2011; Kontostathis et al., 2013)). Of these three methods, the data mining technique has been adopted in this research due to its superior ability to identify terms that mostly indicate cyberbullying through using language-based methods (Reynolds et al., 2011; Dinakar et al., 2011), as well as developing query terms (Kontostathis et al., 2013). Thus, the use of the data mining technique promotes the effectiveness of detecting cyberbullying messages.

This chapter starts with a discussion of cyberbullying definitions in Section 2.2. Sections 2.3 and 2.4 discuss the characteristics and impacts of cyberbullying. Following this, an overview of types of social networks, including Indonesian social

networks in particular, are presented in Sections 2.5 and 2.6. The content analysis of social networks is discussed in Section 2.7, and lastly, the techniques used for analysing the content in social networks are discussed in Section 2.8.

2.2 Definition of Cyberbullying

People can perform acts of intimidation using information technology devices, for instance, a mobile phone through the Internet to send offensive or threatening messages to others, either directly or indirectly in the forms of texts, pictures and videos (Slonje and Smith, 2008). In this way, one or more individuals can demean victims (Privitera and Campbell, 2009) by repeatedly targeting people who cannot easily defend themselves (Dooley et al., 2009). As a result, a new type of human rights violation known as cyberbullying has emerged (Thomas et al., 2013). This term is now used globally, and refers to a new phenomenon of bullying in society using information technology devices.

In ground-breaking research into cyberbullying in secondary schools, Smith, Mahdavi, Carvalho, Fisher, Russell, and Tippett (2008) stated: *"*cyberbullying is reported as an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend him or herself*"* (Smith et al., 2008). The repetitive and continuous cyberbullying in this case can be done in two ways. The first approach is by sending direct cyberbullying messages, where a person wishing to bully their target will deliver private messages straight to the target via e-mail (Langos, 2012). Here, the followers and friends of the target are expected to mirror the same act of sending cyberbullying messages by re-sending the messages or creating a similar new message. The second approach is by indirectly sending cyberbullying messages through a chain

whereby a person requires assistance from others in sending cyberbullying messages to the target via emails (Sleglova and Cerna, 2011; Snakenborg et al., 2011) Using this method, the perpetrators enlist the support of friends or followers to show their power. It is then expected that the followers and friends of the initial perpetrator will continue the harassment by sending more cyberbullying messages.

An executive board in the United Nation Development Programme further defined cyberbullying as *"*the situation in which a person is verbally abused or threatened via electronic media, such as social websites, email or text-messaging. The text can include derogatory remarks, insults, threats or harmful rumours*"* (UNDP, 2013). In respect to this, offensive actions in word form can be sent as a text message on social networks through the medium of information technology such as Facebook, Twitter, YouTube, and LinkedIn.

Both definitions indicate that cyberbullying is not merely one or more verbal attacks; it involves different media in electronic form in order to send potentially harmful messages to victims, and perpetrators acting alone or engaging in cyberbullying with the help of friends or followers repeatedly during a period of time. For the purposes of the research of this thesis, the following definitions are employed. The term 'cyberbullying' is defined as being repetitive acts of intimidation and aggressive harassment by an individual or group (followers and friends) with the intention of intimidating or harming a targeted person by means of insulting text messages or pictures or videos sent via communication technologies. The insulting text messages associated with cyberbullying in this research are the Indonesian insulting words engaged in such specific activity, which is cyberbullying on the social networks.

2.3 Characteristics of Cyberbullying

The characteristics of messages sent online are defined by their content, which can consist of either negative or positive words (Sanders et al., 2009). These contents are understood in relation to the meaning of every word and the combination of words. Every word has a precise meaning, while the word combinations will modify meaning to produce either positive or negative sentences.

According to Willard (2006), there are various forms of cyberbullying which are the following:

- Flaming: sending messages using rude or vulgar content in order to express anger through an online group, e-mail or other electronic means;
- Harassment: sending offensive, rude, and insulting messages repetitively via e-mail, social media or other electronic devices;
- Denigration: spreading derogatory and unjust criticisms via posts on a Web page. The distribution of messages to others is usually done via e-mail or instant messaging, or posting or sending digitally altered photos of a person;
- Impersonation: hijacking an e-mail or social networking account of the victims and using the victim's online identities to send or to post vicious or embarrassing materials to others;
- Outing and Trickery: sharing the victim's personal and shameful information which tricks the victim into revealing their personal information and forwarding it to others; and
- Cyberstalking: repetitively sending messages intended to be harmful threats or strong intimidation, or engaging in other online activities that make a person fear for his or her safety (Willard, 2006).

Although cyberbullying can fall into one of the above six categories, it can also have these characteristics: the anonymity of the perpetrator (O'Brien and Moules, 2010); the timing for sending and re-sending harmful messages (Privitera and Campbell, 2009); the intentionality of re-sending messages (Spears et al., 2009); and the actual content of messages (Mishna et al., 2009). These characteristics are explained as:

- The anonymity of the perpetrator. The purpose of anonymity is to disguise the name and identity of the perpetrator so that the victim does not know who is sending the message (O'Brien and Moules, 2010);
- The flexible access of time and place where messages were sent. Cyberbullying can be done anytime and anywhere, meaning that perpetrators can access the Internet at whatever time and place they choose in order to send the offensive messages without directly meeting their victim. This is different from traditional bullying practices where, in order to bully someone, the perpetrator has to meet face-to-face with the victim(s) and this usually occurs at school, in the bus, walking to or from school, in a public area, and in the workplace (Privitera and Campbell, 2009).
- Power differentials. Here an inherent need for greater power and influence over other people through the repetitive behaviour of sending offensive messages underlies the intention to harm (Spears et al., 2009). Repetitive bullying-type behaviour takes on a different meaning in cyberbullying, since the shared materials can continue to be distributed long after the original bullying incident.
- Spreading of rumours. The content of cyberbullying messages contains threats and derogatory comments that are intentionally passed on to deceive others (Mishna et al., 2009).

The characteristics and forms of cyberbullying help to determine whether or not the messages contain cyberbullying words. Moreover, these characteristics can be combined together as a guide to recognising the nature of received messages from social networks as well as other information technology devices. For example, such messages can be identified by their rude or vulgar content, sent continuously and repeatedly by followers and friends who have become involved. However, to accurately recognise the characteristics and forms of cyberbullying messages requires other measurement parameters such as the classification of cyberbullying messages (Bauman et al., 2012; Whittaker and Kowalski, 2015). Multiple variable measurements are needed to provide a more comprehensive idea of the means by which cyberbullying occurs, and to recognise the differences between cyberbullying types in different places. In previous studies, the usage of variable measures in identifying cyberbullying messages signifies that the information is less detailed. However, with two or more variables using specific characteristics will produce more comprehensive information (Bauman et al., 2012; Bauman, 2015; Whittaker and Kowalski, 2015).

To briefly summarise the forms presented previously, cyberbullying messages typically can be categorised as flaming, harassment, denigration, impersonation, outing, and cyberstalking.

2.4 The Impact of Cyberbullying

This section describes the impact of cyberbullying on victims and is based on previous studies. Several studies have been conducted to determine the effects of cyberbullying messages. Ybarra et al. (2006) found that victims often feel distressed. Beran and Li (2008) studies of the impact of cyberbullying found that it affects students' concentration, causes absenteeism, and results in difficulties related to academic

achievements. Other effects of cyberbullying were revealed by Gradinger et al. (2009) who noted that cyberbullying resulted in students having poor adaptation, aggression, depression and other somatic symptoms compared with those students who had not experienced any form of bullying.

Concerning its wider impact, the effects of cyberbullying may be detrimental to mental health, and lead to substance abuse and ideas of suicidal thoughts (O'Brien and Moules, 2010; Goebert et al., 2011) when the victims feel helpless (Gualdo et al., 2015). Cyberbullying also decreases the individual's level of confidence and self-esteem, as well as affecting mental health and emotional well-being (O'Brien and Moules, 2010). Moreover, after receiving cyberbullying messages, victims might perceive themselves to be worthless (Goebert et al., 2011), which eventually can increase their level of depression (Kowalski et al., 2012). Hence, research on detecting cyberbullying messages on social networks has evolved because of the devastating effects of this practice. It has attracted a great deal of attention and it is anticipated that such research will help in the understanding and early identification of messages.

2.5 Cyberbullying on Social Networks

According to Smith et al. (2008), cyberbullying has been carried out through the media of communication technology that includes the Internet and communication devices such as mobile phones. They explored cyberbullying though the Internet via chat rooms, e-mails, social networks, and websites. Cyberbullying using a mobile phone can take the form of phone calls, text messages, pictures and video clips including so-called 'happy slapping', where a victim is slapped or made to appear silly by one person, filmed by another, and the resulting pictures circulated on mobile phones.

Smith et al. (2008) have noted that social websites are one form of media often used to send cyberbullying messages. Social websites that provide multiple daily opportunities for connecting with friends, classmates and people with shared interests can include: social networking sites such as Facebook, Twitter, LinkedIn, and YouTube (Russell, 2011). As these social networks websites have grown exponentially in recent years (O'Keeffe and Clarke-Pearson, 2011), people's use of social networking sites to interact with each other has also increased dramatically (Cheung et al., 2011; Best et al., 2014).

Social networks have become a popular online medium of communication, enabling people to send their ideas as text, pictures, and video without any restrictions (Duggan et al., 2015; Lenhart, 2015). Since, social network sites allow people to freely send and receive messages, they have more opportunities to send not only positive messages but also negative ones as well. If the latter are abusive and harassing messages, they are considered to be cyberbullying (Ybarra et al., 2006; Livingstone et al., 2010).

Dredge et al. (2014) stated that most of their respondents have had negative experiences of receiving cyberbullying messages through social network websites. This finding supports that of Whittaker and Kowalski (2015) who indicated that the common venues of bullying have likely shifted to social networks. Ellis (2013) survey conducted in the United Kingdom (UK) found that cyberbullying mostly occurs via these three social networks: Facebook, Twitter, and Ask.FM. Ditch the Label surveyed teens in the UK in order to discover the social networks most often used and the frequency of the bullying that was experienced (Hackett, 2015). Ditch the Label's statistics regarding the use of social networks as a platform for cyberbullying are as follows:

- 75% of teens use Facebook and 54% experienced cyberbullying
- 66% of teens use YouTube and 21% experienced cyberbullying
- 43% of teens use Twitter and 28% experienced cyberbullying
- 36% of teens use Ask. FM and 26% experienced cyberbullying
- 24% of teens use Instagram and 24% experienced cyberbullying
- 24% of teens use Tumblr and 22% experienced cyberbullying

This report shows that in 2014, the next most cyberbullied respondents had used Instagram, Tumblr, Ask.FM and Twitter, respectively.

The primary focus of the social network in the research in this thesis is Twitter, which is an accessible site to analyse cyberbullying. For this reason, the wide access of the social network site, where post and short messages are limited to 140 characters or referred to as "tweets", has been selected for analysis (Diffen, 2016). In addition, the availability of numerous Application Programming Interface (APIs) in Twitter provides easy access for the data to be collected and analysed than for other social networks like Facebook (Sanchez and Kumar, 2011). Therefore, this research has had access to crawl data from the Twitter API in order to identify cyberbullying messages that have occurred in the Indonesian context.

2.6 Cyberbullying on Indonesian Social Networks

Cyberbullying has become a concern in developing countries like Indonesia (Sulianta and Hendrawan, 2015). An example is the case of a cyberbullying attack on a senior teacher in one of the junior high schools in Indonesia. A senior teacher stated that, “My Facebook account has been hacked by one of my students and used to send false and deceptive messages to another student” (Soeriaatmadja, 2011). Another case was a student who created a fake Facebook account under a teacher’s name

and declared that the teacher was in a sexual relationship with either a male or a female student. A further case of sexual harassment was hacking into a teacher's Facebook account, and then spreading indecent and obscene messages to other students. This incident caused a great deal of stress to the teacher who became deeply depressed (Soeriaatmadja, 2011).

Therefore, Safaria (2016) undertook research on Indonesian cyberbullying based on a questionnaire format assessed with Cronbach alpha measurement and content validity. The results indicated that of the total of 102 Indonesian junior high school students, 14.28% had never been exposed to cyber victimisation, 25.5% had experienced occasional cyberbullying, 20.6% were exposed to cyberbullying some of the time and 27.5% experienced it very often, while 12.7% were exposed to almost daily cyberbullying. In addition, there were significant disparities between males and females that conducted the act of cyberbullying. Males committed cyberbullying on social media slightly more than females, suggesting that gender is a determinant element in cyberbullying. There were various types of cyberbullying acts identified in Safaria (2016) research. Most of the participants had been exposed to name-calling harassment (45.1%), while 12.7% had been subjected to name calling and denigration as part of cyberbullying. Multiple acts such as name calling, exposure of sexual material, denigration, disclosure of private information and threats were experienced by 13.7% of the participants.

Previous research by Rasi and Huang (2016) adopted a content-based approach by proposing a model to discover cyberbullying vocabulary comprising of heavy usage of offensive language. However, their research did not focus on a particular language context as compared to the research of this thesis. They achieved their objective based on participant-vocabulary consistency using Twitter and Ask.fm

data sets, and detected cyberbullying vocabulary including the subjects involved, victims and bullies.

Thus, the content of the messages are examined to determine whether they contain insults or swear words that enable them to be classified as cyberbullying or otherwise (Syam, 2015). Therefore, cyberbullying detection-based research needs to be expanded (Riadi and Hariani, 2017). The research in this thesis has used a similar theme to the above mentioned research in discovering cyberbullying characteristics through insulting word patterns, thus cyberbullying detection on the social network within the Indonesian context.

2.7 Analysis of Contents on Social Networks

This section describes how data mining techniques are used to analyse the content of messages sent via social networks. Social network sites such as Facebook, Twitter and YouTube have been increasingly attracting the attention of researchers in recent years (Kossinets and Watts, 2006) because of their association with various social processes such as social interaction, diffusion of social influence, information exchange between users, and their dynamic nature (Ellison, 2007; Riff et al., 2014).

Social networks are defined as: *"web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system"* (Ellison, 2007). Social networks are also understood to be: "virtual communities which allow people to connect and interact with each other on a particular subject or to just hang out together online" (Murray and Waller, 2007; Cheung et al., 2011). In other words, social networks are a medium of communication between people by means of the Internet, whereby

users can establish social relations with others through sharing similar personal interests, activities, backgrounds, and career interests. Moreover, as a new communication and interaction platform, social networks can motivate users to learn something new, and they facilitate learning because users can communicate easily and interact (Mazer et al., 2007).

Thus, social networks can become an effective communication medium among users because they offer users a means of creating a virtual identity and network with friends and family as well as sending messages directly which contain texts, pictures, and video (Acemoglu et al., 2010). For example, in the past, people sent letters to their families and friends through postal services and it took several days for letters to arrive at their destination; now, people can directly send a letter through short messages services (SMS) and social networks, and can receive replies promptly.

Although, social media provide functions that allow texts, pictures, and videos to be sent easily and directly, nevertheless they do not offer any means by which positive and negative content can be cleaned (Pang and Lee, 2008). This means that on social networks, people can express their ideas freely without any restriction regardless of whether the content contains positive or negative messages which might affect the receiver (Liu, 2012). Therefore, the early detection of content on social networks has become an interesting issue in some research areas such as sentiment analysis and opinion mining (Pang and Lee, 2008; Liu, 2012), opinion extraction (Lo and Potdar, 2009), and sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc (Aggarwal and Zhai, 2012).

In the past decade, several types of research have been conducted to analyse content on social networks in a variety of contexts. For example, finding high-quality content in social media, and exploring methods to obtain the community's feedback

have been the focus of research by Agichtein et al. (2008). They used a general classification framework and a high-quality definition of content in order to automatically combine evidence from different sources of information for a given social media type. Research using content analysis by Riff et al. (2014) on social networks from the communication science perspective analysed media messages using quantitative content analysis. Their research used content analysis to describe the content of messages and to test the theory of communication.

Thus, the content analysis approach can be used to analyse meaningful matter such as texts, images and voices (Krippendorff, 2012). Analysing content on social networks is an important issue for a researcher who is exploring the effect of messages. Content analysis is also an essential tool for the categorization of all types of content in messages (Riff et al., 2014). However, one area that needs to be further explored is the use of data mining techniques for the analysis of content on social networks (Barbier and Liu, 2011; Aggarwal and Zhai, 2012; Han et al., 2012).

2.8 Techniques for Data Analysis

Data analysis is a method of processing, examining, cleaning, converting and modelling data with the purpose of discovering interesting hidden information that can assist with decision-making (Han et al., 2012). In data analysis, diverse techniques and approaches are used to extract meaningful information; the particular approaches that are used rely on the variables and science disciplines (Ott and Longnecker, 2015). In statistical applications, data analysis can be divided into:

- A descriptive statistic which describes the results in terms of a statistical model;
- Exploratory data analysis that concentrates on learning and discovering new features in large data; and

- Confirmatory data analysis that examines and confirms the existing hypotheses (Neuman and Robson, 2012).

One of the data analysis methods is data mining techniques. These techniques are used for modelling data analysis and knowledge discovery of large data, the purpose of which is to predict rather than purely describe data (Zhu, 2007). Therefore, data mining techniques can also be referred to as predictive analytic techniques since these techniques involve the application of statistical models (Han et al., 2012) for the purposes of forecasting and data classification. Furthermore, these data mining techniques can be used to analyse text and linguistics by extracting and classifying information from textual sources (Miner, 2012; Larose, 2014).

A sample case of implementing data mining techniques to find hidden information in the database of a bookstore is identifying and understanding the patterns of the relationship between variables from the database. Every purchase or sale of items in a bookstore is recorded in a database, whereby one of the tables stores transactional data that contains information about customer identity, age, address, credit card, and the purchased items. These variables are processed further by data mining techniques to find the relationship between them, so that patterns emerge from the data (Raorane et al., 2012). This can be achieved by combining each item with other purchased items. Then, the items that have been purchased in the same transaction will be determined, and the frequency of sets of item will be obtained. The frequent item set is commonly referred to as the frequent pattern (Chapman and Feit, 2015).

As a further example, data mining techniques have been applied to analyse data in texts on social networks, such as predicting opinions in messages (Bhagat et al., 2011), analysing current topics and issues on social networks (Lee et al., 2011),

and identifying the characteristics of users from social networks (Spiegel, 2011). Hence, data mining techniques can be used for classification purposes. This particular classification technique is used to predict the probability of data classification either in the form of texts or numbers. However, in this technique, data training is necessary for estimating data classification. Data training is a set of data values, often applied to distribute pre-defined class labels to new objects (Raschka, 2014).

In terms of functionality, data mining techniques can be categorised according to some common tasks. These tasks include finding the association rules that link objects together, clustering objects together based on some characteristics, classification of objects into predefined groups, detecting outliers as well as regression and summarization (Han et al., 2012). Several data mining techniques are described below:

- Association rules are used to discover the relationship between the variables that exist in the database (Raorane et al., 2012; Chapman and Feit, 2015).
- Clustering is applied to find groups and data structures of unknown classification in the database, where the data have similarities with each other (Jain, 2010; Shiells and Pham, 2010).
- Classification is applied to generalise data structures that its own group has previously known, into a new group of data. Classification involves supervised learning techniques that classify data items according to pre-defined class labels (Kotsiantis et al., 2007; Richards, 2013).
- Outlier detection is used to detect any unusual data that are very different from expectations. Any unusual data will be revealed and subjected to further in-depth investigation (Bakar et al., 2006; Angiulli, 2009).

- Regression is used to estimate the response variable value based on one or more predictor variables in which the variables are in numerical form (Witten and Frank, 2011; Han et al., 2012).
- Summarization is used to provide a report concerning data collection as calculated using the algorithm. The reports can be presented as graphs or regular reports (Kantardzic, 2011; Han et al., 2012).

The types of data mining techniques listed above can be used for a variety of purposes in research, business, and government. From the academic perspective, data mining techniques can be termed as statistical methods, since they need to be further investigated and refined according to the various types of data which include data streams, ordered or sequenced data, graph or networked data, spatial data, text data, multimedia data, and websites (Han et al., 2012). Similarly, in the business field, data mining techniques can assist businesses to increase their sales and company profits (Larose, 2014).

In Government organisations, such as those in the United States, data mining techniques are applied to identify terrorism networks throughout the world (DeRosa, 2004; Thuraisingham, 2004). The United States government employs several data mining techniques such as association rules and classification techniques to find the link between people suspected of terrorism and their followers and friends in their networks (DeRosa, 2004). These techniques are effective and helpful in obtaining information that can provide decision-making support to counteract terrorism in the United States.

Table 1 the Implementation Data Mining Techniques

No.	The use of data mining techniques	Authors
1.	A study of development data mining techniques	(Bakar et al., 2006; Zhu, 2007; Angiulli, 2009).
2.	Implementation of data mining in business	(Berry and Linoff, 2004; Hsieh, 2004; Yeh and Lien, 2009).
3.	Using data mining techniques by government	(DeRosa, 2004; Thuraisingham, 2004; Shaikh et al., 2007; Shacheng, 2012).
4.	Analysing social networks using data mining techniques	(Barbier and Liu, 2011; Aggarwal et al., 2011; Cooper, 2012; Liu, 2012; Nahar et al., 2012; Kontostathis et al., 2013; Nahar, 2014).

Data mining techniques can serve as tools for analysing the content on social networks (Barbier and Liu, 2011; Cooper, 2012); for instance, by detecting sentiment content through websites (Aggarwal et al., 2011; Liu, 2012), thus eventually identifying either positive or negative content that may lead to cyberbullying or non-cyberbullying messages (Nahar et al., 2012; Kontostathis et al., 2013; Nahar, 2014).

2.9 Stemming Data

Many types of language processing and text analysis methods use stemming techniques to remove affixes (suffixes and prefixes) from root words (Frakes and Fox, 2003), in order to reduce the number of words and to obtain an exact matching word (Ramasubramanian and Ramya, 2013) that closely approximates the root morpheme of a word. For example, the word “children” stems from the word “child”.

A stemming technique is a computational process the purpose of which is to reduce all common form words to the morpheme word root by stripping each word of its derivational and inflectional suffixes (Smirnov, 2008). Therefore, a stem technique is employed to maximise the usefulness of the subject terms.

Several stemming techniques are available for reducing and removing words including: Snowball, Porter, Lovins, German, Arabic, and Dictionary. Snowball stemming is a designated small string language process intended to construct stemming algorithms for information retrieval. The stemmers in Snowball stemming can be precisely characterised and can be generated by fast stemmer programs in Java (Porter, 2001).

Another stemming technique of Porter's algorithm was developed for the stemming of English-language texts. However, an increased amount of information retrieval has become very important since the 1990s, creating widespread interest in the development of conflation techniques. These strengthen the discovery process of texts written in other languages. Currently, the Porter algorithm is the standard technique for stemming English; therefore, it is a natural model within the area of language processes (Porter, 2005; Porter, 2006).

The first stemmer to be proposed, particularly in retrieval applications, is Lovins stemming. Apart from its retrieval application, this stemming includes stemming according to the dictionary of common suffixes, for instance, DES, ING or TION (Porter, 2005). Lovins stemming has initiated and promoted the improvement of prevalent algorithms, as well as being a basic, general application tool for information retrieval (Frakes and Fox, 2003).

In addition, stem dictionary is a processing stem based on the dictionary. When a word undergoes the stemming process in a dictionary-based stemming algorithm, a general search for the presence of any suffixes in the dictionary at the right-hand end of the word is executed. When the presence of a suffix is discovered, it is then extracted, and exposed to a range of context-sensitive rules. These may involve the removal of *ABLE from TABLE or of *S from GAS; additionally, a spectrum of recording

rules may be arranged and supplied to facilitate the conflation of variants such as “Forgetting” and “Forget” or “Absorb” and “Absorption” (Porter, 2006).

One of the several kinds of stemming techniques related to the Indonesian language stemmers is the approach used by Nazief and Adriani (Nazief and Adriani, 1996; Adriani et al., 2007). The Nazief and Adriani approach is based on comprehensive morphological rules in which the grouping and encapsulating processes allow and forbid affixes, as well as prefixes, suffixes, infixes (insertions) and confixes (combination of prefixes and suffixes) (Asian et al., 2005; Adriani et al., 2007). In the grouping of basic affixes, the following procedures are applied: the first step, the affix, which is the inflexion suffix that does not alter the root word. For instance, “makan” (to eat) may perhaps be suffixed with “-lah” to express “makanlah” (please eat). The second affix is a derivation suffix that is directly added to the root word. In this case, there can be one exclusive derivation suffix per word. An example is the word “baca” (to read) that can be suffixed by the derivation suffix “–kan” to become “bacakan” (please read). The third affix is the derivation prefix, which is employed either directly to the root words, or to the words possessing up to two other derivation prefixes. To illustrate, the derivation prefixes “mem-” and “per-” may be prepended to “jelek” (bad) to produce “memperjelek” (the act of ill-favouring) (Asian et al., 2005). Re-coding, an approach that locates and recovers an initial letter that was previously taken out from a root word before prepend of a prefix, is also provided by the algorithm. Moreover, the algorithm controls the use of an auxiliary dictionary comprised of root words applied in most of the steps used to determine whether or not the stemming has reached a root word (Nazief and Adriani, 1996; Adriani et al., 2007).

A stemming technique developed by Arifin and Setiono (2002) employs a dictionary to deliberately extract affixes and manage the re-coding process. The

purpose of their approach is to eliminate all prefixes and suffixes; the process halts when the stemmed word is discovered in a dictionary, or when the number of affixes that has been removed has reached a maximum of two prefixes and three suffixes. When the removal of prefixes and suffixes has been completed, if the stemmed word still cannot be searched, the affixes will then be revived in the word in every possible combination, so as to minimise the possibility of having stemming errors (Arifin and Setiono, 2002; Asian et al., 2005). For instance, the word “disamakan” (to be equal) contains the root word “sama” (equal). Following the first step of eliminating the prefix “di-”, the word that remains is “samakan” (to equalise).

A different stemming technique by Vega (2001) does not require a dictionary as compared to other approaches. Rules are applied to each stemmed word in order to divide the word into smaller units. For instance, the word “didudukkan”, which translates as ‘to be seated’, might be expressed by the following rule: (di) + stem (root) + (i | kan) (Vega, 2001; Asian et al., 2005).

Additionally, a stemming technique by Ahmad et al. (1996) has two distinctive characteristics: first, the approach was intended to be closely-associated with the Malaysian language, instead of Indonesian; and second, it is a straightforward approach. A list of all prefixes, suffixes, infixes and confixes in order and validity are maintained. Prior to the stemming process, the algorithm seeks the word in the dictionary, and successfully restores the original version of the word. If the word cannot be located in the dictionary, the next rule in the rule list will then be applied to the word (Ahmad et al., 1996; Asian et al., 2005). Say, when the infix rules are applied before the prefix rules, “berasal” (to originate) — for which the precise and appropriate stem is “asal” (an origin or source) — is stemmed to “basal” (basalt or dropsy) by extracting the infix “-er-” preceding the prefix “ber”.

Of these four approaches explained above, Nazief and Adriani's (1996) approach has been adopted in this research on account of its ability to eliminate and to embed prefixes and suffixes, which are able to decipher the words close to their true definition. Therefore, the research in this thesis has extended the Indonesian stem dictionary specifically to create and implement an Indonesian stem dictionary of insulting words. The purpose of this dictionary is to identify and remove the suffixes and prefixes, especially in Indonesian insulting words. Our stem dictionary will help to identify some Indonesian insulting words that usually appear in social network messages.

2.10 Association Rule Techniques

The association rules techniques are one of the data mining techniques which will be applied further in this study. The concept of the association rules techniques introduced by Agrawal et al. (1996)is defined as a method for analysing data in discovering the most important relationships between variables in a large database, using rule-based machine learning based on the minimum standards of rules support and confidence (Han et al., 2012).

Han explains that the methodology of association rules mining relates to the development of a market basket analysis system (Han et al., 2012). Every basket of purchases will be analysed for its consumption patterns representing the goods bought by the customer. To find the consumption patterns that associate one item with another, the first stage that must be completed is to find the frequent itemsets (Chapman and Feit, 2015). Frequent itemsets are sets of items that commonly and simultaneously appear in the database.

In the process of generating the association rules techniques, there are two requirements that should be considered. First, each item set that occurs in the database must appear frequently until it satisfies the minimum support count. Second, the relationships between the itemsets must be strong enough to create a pattern; therefore, the relationship calculation of itemsets must satisfy the minimum standards of rule support and confidence (Han et al., 2012). Han et al. (2012) define that the rules support and confidence can be expressed as follows:

$$Support\ (A \Longrightarrow B) = P(A\ \cup B)$$

If the calculated result from the relationship of itemsets does not satisfy the rules support and confidence, then the item will be ignored. But if the calculation results of frequent item sets satisfy both standards, then the item will be combined with the other items until the trend of frequent item sets is found (Han et al., 2012; Chapman and Feit, 2015). This trend represents the patterns of goods purchased by customers. The following is an example of a frequent item extracted from Table 2:

Table 2 Example of Market Basket Transaction in Association Rules

TID	Items
1	pencil, eraser pencil, pen,
2	book, clip eraser, book, pen,
3	tape pencil, eraser, pen,
4	book, clip pencil, eraser, pen,
5	tape, clip book, pen,
6	tape, eraser

The frequent data: (pencil) → (pen)

In this scenario, Transaction ID (TID) is a set of transactions obtained by means of association rules. The association rules find the relationships between different items in Table 2, in which all the frequently purchased items will be considered as the patterns of purchases. In this case, the outcome from the frequent items in Table 2

after extracting the data set is the relationship between a pencil and an eraser. This means that some customers who purchase a pencil also buy an eraser.

In research conducted by Nancy et al. (2013), association rules techniques have been applied in mining of association patterns in social networks data. Their research analysed the relationship between courses and gender attributes in 100 Facebook universities in the United States and found the influence of gender in studying courses. They confirmed that association rules had been successful in discovering the relationship between attributes, such as the possibility of gender influencing course preferences in these universities.

Association rules have also been used to discover the hidden correlations among the contents posted on social networking websites and detect trends of online users (Mahoto et al., 2014). Mahoto et al. (2014) reported that they analysed the data of social networks and found hidden correlations, and then built a taxonomy based on their corresponding relationships. They asserted that association rules techniques are helpful in understanding the hidden associations between the textual contents and contextual features of the user-generated content, and association rules are an effective technique in finding correlation among terms in tweets.

Therefore, the research in this thesis uses the association rules technique to analyse the messages sent via social networks to determine the relationship between the words in the messages. The use of association rules in analysing the relationship between the words is very precise and efficient (Mahoto et al., 2014), and can help to solve the problem of research in this thesis of analysing the relationships between texts within the special terms in social networks. However, choosing the most suitable techniques in association rules needs to be considered to obtain effectiveness and

efficiency in detecting the correlation between the words, especially the terms that have been used in cyberbullying messages.

Another reason for using association rules is that it is a well-researched method for discovering interesting relations between variables in large databases. These techniques can also discover and extract hidden correlations among the terms in large data texts (Han et al., 2012; Mahoto et al., 2014).

2.10.1 Apriori Algorithm

Proposed by Agrawal and Srikant (1994), Apriori is an algorithm for mining frequent itemsets including association rules learning in transactional databases. Apriori performs through an identification process over frequent single items in a database. Apriori algorithms have been used in previous research to analyse text document within various fields. For example, Lovinger and Valova (2015) implemented Apriori algorithm to find the frequent patterns of terms in text documents. They modified the Apriori algorithm to convert the text into rules and to eliminate the need for calculating support by effectively utilizing hash tables and scales for large database applications which cover virtually all blogs and other forms of social networks. Through the results of the modified Apriori algorithm, their research has shown that the Apriori algorithm has efficiently determined the link between words, which can predict one word appearing after another given word.

Maheshwari et al. (2014) also used the Apriori algorithm to find the identification between cyberbullying victims and predators. As a result, they established that the communication frequency between two people is based on a direct relationship, with a close probability of them being related to each other. This research successfully

applied the Apriori algorithm to investigate links between those involved in cyberbullying messages from social networks.

The research in this thesis has also employed the Apriori algorithm to solve the research problem of finding the relationship between words that refer to frequent patterns in Indonesian cyberbullying messages. Although the Apriori algorithm is simple, it can help to compute the relationship between words by reducing candidate generation to improve the efficiency of finding the frequent itemsets (Han et al., 2000; Chai et al., 2007; Wu et al., 2008; Abaya, 2012; Lovinger and Valova, 2015).

However, finding all frequent itemsets in a database requires a large space for temporarily storing data when the algorithm is generated. Applying the Apriori algorithm can help to reduce the usage of a large space in memory and to minimise candidate generation (Han et al., 2000; Chai et al., 2007; Wu et al., 2008; Abaya, 2012; Lovinger and Valova, 2015). To improve the efficiency of the amount of used space and to minimise candidate generation, the Apriori algorithm has a prerequisite that must be met. This is that the entire non-empty subsets must be frequent. If the entire non-empty subsets are not frequent, they must be infrequent. This is commonly referred to as the Apriori property (Agrawal and Srikant, 1994; Defu et al., 2008; Han et al., 2012).

The Apriori algorithm uses an iterative technique, which is popular for its level-wise search approach through searching candidate *k*-itemset (Srikant et al., 1997). In order to find candidate *(k+1)*-itemset in the Apriori algorithm, the first step is to browse the database to obtain the count for every item and investigate whether or not the items are frequent. After this step, the process gathers all the items which satisfy the minimum support as candidate 1-itemset (L1). L1 combines with itself to discover and identify the candidate 2-itemset (L2). This process repeats as L2 also combines with

itself producing a discovery of candidate 3-itemset (L3), and so on until there are no more frequent k-itemset to be found (Abaya, 2012; Han et al., 2012). According to Han et al. (2012), the listed steps of the Apriori algorithm in detail are:

- The first step is to find a candidate k-itemset, which is generated by joining Lk-1 with itself. The set of candidates will be named as Ck. For instance, the results from joining both (*k*-1) itemsets A= ($a_1,\ldots,a_{k-2}, a_{k-1}$) and B= ($a_1,\ldots,a_{k-2}, a_{k-1}$) are candidate *k*-itemset ($a_1,\ldots,a_{k-2}, a_{k-1}, a_k$).
- The next step is to prune the size of C_k by applying the Apriori property, since there are large amounts of computations. When the (*k*-1) subset of candidate k-itemset is frequent in L_{k-1}, the candidate cannot be taken out from C_k. Nevertheless, when *k*-itemset subset of candidate k-itemset is not frequent in L_{k-1}, removing the candidate from C_k is permissible. This means the entire candidate itemset that has been calculated has to satisfy both the minimum support and the minimum confidence value.

2.10.2 Frequent Pattern Growth (FP-Growth)

In finding the frequent itemsets in an efficient manner, FP-growth is one of the techniques in the association rules techniques that does not require candidate generation (Han and Pei, 2000). This method derives the frequent itemsets straight from the Frequent Pattern Tree (FP-Tree). In the order of the level of techniques in association rules, FP-tree is more advanced than the original Apriori algorithm (Defu et al., 2008).

A FP-tree is a compact data structure that grants access to discover the frequent itemsets in the form of a tree (Han and Pei, 2000; Han et al., 2012; Larose, 2014). This involves the process of scanning every transaction data and then creating

a path of FP-tree form. Next, this technique iterates until the entire transactional data has been scanned through and read. The different transactions in the common subsets permit the tree to stay solid due to their paths overlapping.

According to Han and Pei (2000), the stages of the FP-tree mining algorithm are listed as follows:

- The first step is scanning the entire transactions in the database, then constructing the FP-tree in a descending order of the support count.
- The frequent patterns from the constructed FP-tree in the previous stage are derived through obtaining the frequent itemsets from the FP-tree. The next step is adopting a divide-and-conquer method in bottom-up fashion from the leaves towards the branches. The intention of this approach is to explore the frequent itemsets along with specific ending item suffixes. In the direction of dividing the database into several sets of conditional databases, the suitable approach is to construct a conditional pattern base consisting of the set of prefix paths in FP-tree followed by the suffix patterns.

Yuan et al. (2007) have implemented the FP-growth algorithm for a quick search of the entire frequent sets from a grouped-words dataset. A significant visual phrases lexicon pattern has been detected efficiently in a boundary of 27 seconds from the database with more than 60,000 grouped-words. In addition, Ahmad and Doja (2013) have applied the FP-growth to mine frequent patterns of opinion in conjunction with appropriate features representing a variety of topics and subjects from data of unstructured texts. Through the application of FP-growth, they have found the entire features of opinion and an improvement of the features' recall value.

Furthermore, Alghamdi (2011) has efficiently implemented the FP-growth algorithm in the medical field to provide more referential text and understanding about

the relationship between a doctor and a patient. In other words, Alghamdi (2011) employed the frequent itemsets rules from the divide and conquers method using the National Bureau of Economic Research surveyed data to generate rules for hospital information, such as a doctor diagnostic system and disease recognition, and even more simple data like patients visiting doctors.

In similar research, Gadia and Bhowmick (2015) parallelised the FP-growth algorithm based on multicore machines in favour of a rapid search on frequently occurring keywords in data. Instead of positioning the algorithm as a measure of generated frequent itemsets, they used the algorithm to discover the frequent items in the database that are less frequently occurring, but are unique.

Research by Mhashakhetri and Sheikh (2016) in the field of business intelligence adopted the technique of FP-growth or what can be referred to as an Item set tree algorithm to generate frequent itemsets about product features that frequently occur in customers' reviews or comments. The derived frequent features served as criteria for which positive, negative, and neutral comments are measured for an overall product review.

This research has implemented FP-growth techniques to achieve three essential objectives; first, a decrease in effort in scanning the database in a range of two times, and a reduction in computational cost; second, no candidate generation is necessary; third, a minimisation of the search space, as the technique employs divide-and-conquer approach (Nasreen et al., 2014).

Moreover, the FP-growth algorithm is one of the approaches that rapidly mines frequent itemsets. In order to produce a concise representation of the database transaction and employees, using its FP-Tree data structure and a divide and conquer approach, this algorithm can eliminate issues arising during the mining process (Han

et al., 2012; Aggarwal and Han, 2014). Additionally, the FP-growth approach is efficient and expandable in mining both long and short frequent patterns (Mishra and Choubey, 2012; Yang et al., 2013).

2.10.3 Cosine Similarity

Cosine similarity is one of the data mining techniques described in this section. The relationship between texts in a context has been extensively researched and analysed for decades. The reason for this interest is the number of messages that have been sent on social networks that may contain positive and negative words. Therefore, identifying the relationship between words is required in order to understand the sentences in the messages and their positive or negative context (Gindl et al., 2010).

The analysis of text using similarity techniques has been adopted for research data text and tasks such as information retrieval using association rules, classification and clustering of text, and text summarization(Huang, 2008; Gomaa and Fahmy, 2013). Identifying the similarity between words is an essential part of connecting text which is then used as a key point in understanding sentence, paragraph, and document similarities. Usually, the words will be considered to be semantically similar when they have similar meanings, are used in the same way, and are applied in the same context (Gomaa and Fahmy, 2013). The similarity of words established by the semantic approach can be determined by Corpus-Based and Knowledge-Based similarity measurements. Corpus-based similarity is a type of measurement that determines the similarity between the words according to the data that appeared in large corpora in the text document. Knowledge-Based similarity is another type of

measurement used to determine the degree of similarity among words by collecting information from the semantic linkages (Sánchez and Batet, 2013).

A similarity measure is a crucial tool used to determine the degree of similarity between two objects. One type of similarity measurement is a cosine similarity measure which was proposed by Bhattacharya (1946). Cosine similarity can be described as a measure of the similarity in terms of the inner product space on two non-zero vectors according to the length of the division of their product (Bhattacharya, 1946; Ye, 2011). This similarity method is intended to measure the similarity between text and is an extensively reported measure of vector similarity (Mihalcea et al., 2006). Cosine is the angle located between the two vectors. When the angle between the two vectors is 0°, it is expressed as 1, while 90° is expressed as 0 (Han et al., 2012).

The formula for cosine similarity, popularised by Han et al. (2012) is:

$$similarity = \cos\theta = \frac{x.y}{||x||||y||} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

in which x_i and y_i components of vector *x* and *y*. x and y are the two vectors that are to be measured for the purpose of comparison. ||x|| and ||y|| are the Euclidean norm of vectors $x=(x_1,x_2,x_3,\ldots x_n)$ and $y=(y_1,y_2,y_3,\ldots y_n)$, outlined as $\sqrt{x_1^2 + x_2^2 + x_3^2, \ldots + x_n^2}$ and $\sqrt{y_1^2 + y_2^2 + y_3^2, \ldots + y_n^2}$. Through this representation, the measure of cosine angle between vectors x and y is computed. The cosine similarity is usually employed in high-dimensional positive spaces (Tata and Patel, 2007), in which the outcome will be accurately confined in [0, 1]. If the unit vectors are parallel, this means that they are maximally 'similar'; when orthogonal or perpendicular, they are maximally 'dissimilar'. It should be noted that this is analogous to cosine, which defines as a union (maximum value) when the segments extend to a zero angle and zero (uncorrelated) if the segments are orthogonal (Han et al., 2012). For instance, in retrieving information

and mining text, every term is allocated a distinct dimension notionally. Also, a document is expressed by means of a vector in which the value of every dimension correlates to the number of times that a term emerges in the document. Then, cosine similarity delivers an effective measure of the extent of similarity between two documents in relation to their subject matter (Singhal, 2001).

The cosine similarity technique is similar to the Euclidean distance technique, since both techniques are the similarity measures between two vectors. On the other hand, the distinction between cosine similarity and Euclidean distance is the method of calculation. Cosine similarity technique measures the degree between two vectors, and Euclidian distance techniques measure the distance between two vectors (Qian et al., 2004). However, both techniques are effective when used for different purposes. For example, cosine similarity is effective when used to measure the similarity between two vectors, while Euclidean distance is effective when used to measure the differences between two vectors (Qian et al., 2004; Kryszkiewicz, 2014).

Therefore, choosing an appropriate technique for finding the similarity between words is essential in the research in this thesis. For this reason, the cosine similarity technique is an appropriate technique for measuring the angle of two vectors as the represented words in a two-dimensional data space (Mihalcea et al., 2006; Huang, 2008; Gomaa and Fahmy, 2013). This thesis has extended the researchers' work by implementing the measurement and experiment in another dataset that comprises text documents from social networks.

This thesis applies the cosine similarity measurement because it can be used to compute similarity scores based on the number of words common to two sentences (Achananuparp et al., 2008; Huang, 2008). Moreover, the cosine similarity technique is a successful metric as it can be used to identify and measure the similarity between

two vectors. Increasingly, cosine similarity is being used for more complex queries (Tata and Patel, 2007). Moreover, cosine similarity is one of the few fascinating measures that have symmetry, triangle inequality, null-invariance (Omiecinski, 2003), and cross-support properties (Xiong et al., 2006). In particular, cosine similarity is suitable for proximity measurement within high-dimensional space (Zhu et al., 2011). Furthermore, cosine similarity is a very popular similarity measure for high-dimensional data when mining text (Zhao et al., 2005; Tagarelli and Greco, 2010) and retrieving information (Bayardo et al., 2007; Solskinnsbakk and Gulla, 2010). Searching for and identifying the correlation among the insulting words in the messages has been a major obstacle in this thesis because of the difficulty of finding the association between one message and another.

2.11 Classification Techniques

Classification is a process of data analysis with which to find the same data based on a set of objects within a database and to classify them into different classes according to the established classification models (Han et al., 2012). Han et al. (2012) define classification as a prediction process of labelling classes of objects into categories by mapping each set of attributes X (input) to one class of Y.

The classification models are used for:

- Descriptive modelling of data as an illustration to distinguish objects from different classes; and
- Modelling predictions in a process of labelling a class to data records that either have or have not had a known class label (Go et al., 2009).

Classification techniques are one of the data mining techniques that use previous data for data training to predict subsequent data grouping (Han et al.,

2012).The classification techniques are also known as supervised learning which can be defined as a process of learning tasks from previous data in providing a class label to the future data (Han et al., 2012; Richards, 2013). By using the previous data or data training, the categorised data which is still unknown or can be termed as data testing will be predicted for their class labels.

In general, there are two stages of the classification process: the learning process from data training and the classification of new cases. Within the learning process, classification algorithms use the data training to generate a particular classification model. After the classification model has been tested and accepted in the classification stage, that model will be used to predict the class of new cases to help the process of creating decisions (Han et al., 2012). The following figure is an illustration of a simple diagram for identifying class label data:

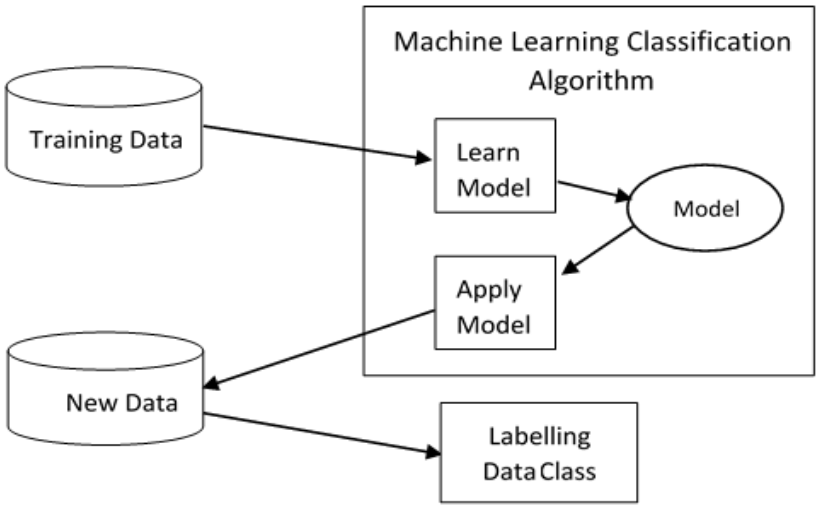


Figure 1 Simple Diagram of the General Model of Identifying Class Labels Data

Han et al. (2012) provided the following brief descriptions of the classification techniques, which can be referred to in Figure 1:

- The first step in classification is building a set of data that has been labelled as a class of data. This step, called a learning step, is a process of training in

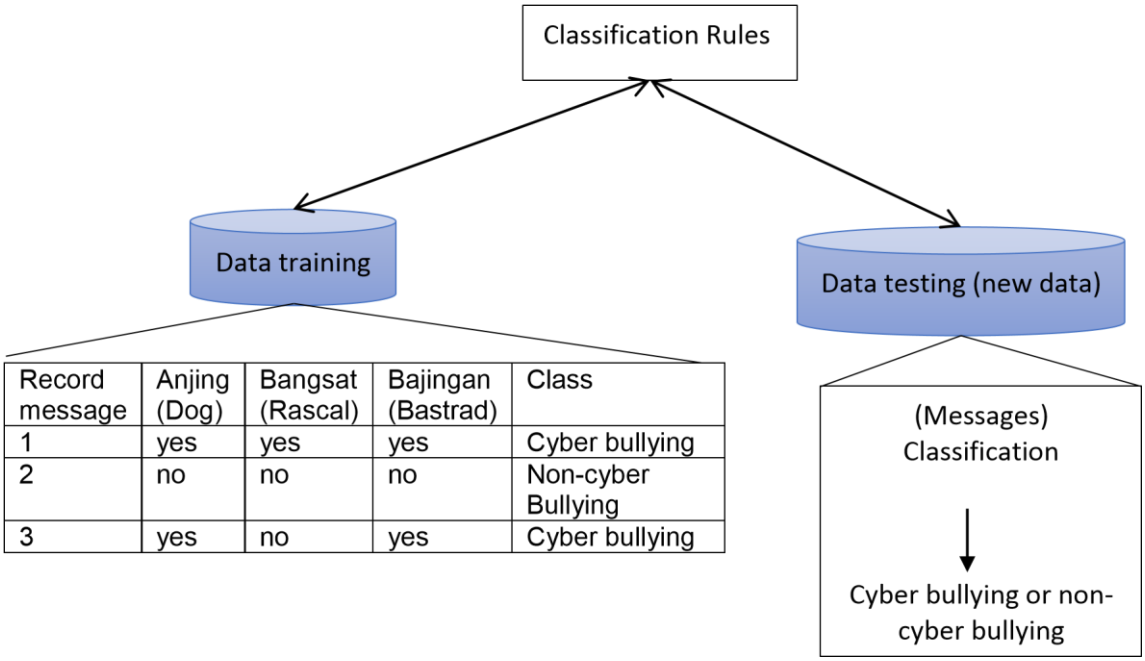
classification algorithms, for constructing a classifier along with analysing the data training consisting of tuples and class labels. This learning step can be illustrated in a simple example. Let A be a tuple, which can be represented in n-dimensional attribute vector, $A = (a_1, a_2, a_3, \ldots, a_n)$, illustrating that n measurements are created on the tuple based upon *n* database attributes, respectively, $C_1, C_2, \ldots, C_n$. The tuple A is considered to be located in a predetermined class in a regulation by a different database attribute, which is commonly referred to as the class label attribute. The values of the class label attribute are distinctive and disordered. Every value of the class label attribute facilitates a category or a class as they are nominal. Each tuple that constitutes the training set is specified as training tuples, which are sampled at random.

- The second step is the classification process, which can be referred to as supervised learning. This stage is perceived as learning about mapping data in regards to predicting the class label data for each tuple. Each tuple that consists of data will be calculated in every class. If the value from the calculated data is higher than the others, the data will be placed in class x. For example, there are classes of a, b, c, and d, therefore each tuple will be calculated in every class of a, b, c, and d. When the calculated data in a class has more values than the other classes, then that data will be addressed into that class.

The examples of the application of data training in a form of message, which contains several Indonesian insulting words, i.e.: *anjing, bangsat,* and *bajingan,* can be seen in Figure 2.

Figure 2 shows the values: *anjing*=yes, *bangsat*=yes, *bajingan*=yes, that can be referred to as a collection of attribute values. The attribute values are chosen to be a data reference in favour of verifying the class labels where the new data will be

inserted. Therefore, whenever there is an entry of new data, the data can be saved in a database. This data will later be calculated with the classification algorithms. After this process, the data will be organised into classes in accordance with its content of bullying words from the messages.



Record message	Anjing (Dog)	Bangsat (Rascal)	Bajingan (Bastrad)	Class
1	yes	yes	yes	Cyber bullying
2	no	no	no	Non-cyber Bullying
3	yes	no	yes	Cyber bullying

Figure 2 illustration of Simple Example Classification Indonesian cyberbullying messages

Dinakar et al. (2011) used classification techniques to detect sensitive topics and to categorise text from YouTube comments. They crawled a corpus of 4.500 YouTube comments and applied a range of binary and multiclass classifiers. They found that binary classifiers for individual labels outperform multiclass classifiers. They also discovered how to detect textual cyberbullying by tackling the building of individual topic-sensitive classifiers.

Kasture (2015) extracted 1.313 unique tweets collected from Twitter. They categorised every word in a tweet as individual based on the pragmatics of language and used Linguistic Inquiry and Word Count to gain that categorisation. Then, tweets

were converted into a multi-dimensional attribute relational numeric dataset. Next, this dataset was applied further in training machine learning classifiers in Weka in order to construct a predictive model for cyberbullying detection purposes. 66% of the data was segmented at random in the contribution to the predictive model, while the 34% was for testing. With a successful result, the predictive model classified the dataset as cyberbullying tweets based on a 0.97 precision value.

Nalini and Sheela (2015) presented machine learning algorithms for classifying the sentiments in Twitter messages either as a cyberbullying text or not. Through a weighting scheme of feature selection, they proposed an effective approach in detecting cyberbullying messages.

Adopting classification techniques is to be given due consideration in the research in this thesis, because these techniques are useful to solving the problem in prediction of class data (Han et al., 2012). Therefore, using classification techniques in this thesis can help to solve the research problem by predicting data from the Indonesian Twitter messages. The messages are classified into cyberbullying or non-cyberbullying class. Dinakar et al. (2011) reported that the classification techniques have successfully detected textual cyberbullying from YouTube comments. These techniques are also effective in predicting the dataset from Twitter as cyberbullying messages with high precision value (Kasture, 2015; Nalini and Sheela, 2015).

2.11.1 Naïve Bayes

Naïve Bayes is a supervised learning method in classification techniques. These methods predict future probabilities based on past experiences (Han et al., 2012; Raschka, 2014). Bayesian classifier is a statistical classifier that can estimate the probability of class membership, such as the likelihood of a given tuple being

owned by a certain class. Bayes theorem explains the relationship between the probabilities of the occurrence of event *A* conditioned with event *B* along with the occurrence of event *B* conditioned with event *A*.

This method is based on the principle of improving probabilities with additional information. Bayes method is useful to alter and update the probability that has been calculated by the availability of the data and additional information. In accordance with subjective probability, if one looks at the event *B* and believes that there is a possibility of *B* to appear, thus the probability of *B* can be called the prior probability. Once there is extra information that such events of *A* have emerged, perhaps there will be changes to the original estimation of the likelihood of *B* to appear. The probability of *B* is now the conditional probability of *A* and referred to as the posterior probability. The concept of the Bayes' rule that was formulated by Thomas Bayes (Han et al., 2012; Raschka, 2014) is written in simple words as follows:

$$Posterior\ probability = \frac{conditional\ probability\ x\ prior\ probability}{evidence}$$

$$p(A_k|x) = \frac{p(A_k)p(x|A_k)}{p(x)}$$

where $X = (x_1, x_2, x_3, \dots, x_n)$ is the probability of a specific combination of X, features (independent variables) in which the variables are allocated to this instance probabilities for each of k potential outcomes or classes A_k

There are two stages in the process of grouping or classifying data, which are training and testing. At the training stage, most of the data where the class has been known are used as data reference to build an estimation model, while in the testing stage the estimation model that has been formed in the previous stage is tested with some of the data.

In previous research, Go and Schneider explained that the naïve Bayes technique can also be applied in analysing messages on social networks (Go et al., 2009; Schneider et al., 2013), which can be both positive and negative. Every sent message will be stored in a database in which the appearance of both positive and negative words within the messages will be calculated. Go and Schneider use a list of positive and negative words as data training to analyse sent messages on blogs and on Twitter (Go et al., 2009; Schneider et al., 2013). If a message contains many appearances of positive words, that message will be classified in the positive class. On the other hand, if there are many appearances of negative words, then that message will be grouped into the negative class. If a message is well-balanced, which means that both the appearances of positive and negative words are equal, that message will be assembled into the neutral class.

Sánchez et al. (2009) used the naïve Bayes technique to predict the classification of cyberbullying messages from Twitter. They retrieved data by directly accessing the Twitter streaming API. Their training data set consisted of Twitter messages containing common used terms of abuse, which were non-classified data. The naïve Bayes technique successfully classified the data close to 70% accuracy (Sánchez et al., 2009).

Nandhini and Sheeba (2015) collected data from two different social networks: Formspring.me and Myspace. They downloaded 500 posts from Formspring.me and 600 posts from Myspace randomly selected. They used naïve Bayes technique to classify data from both Formspring.me and Myspace as a cyberbullying message or non-cyberbullying message. Their result reported that they identified the data which indicated cyberbullying messages with the 95% interval confident of validation on Myspace.com and Formspring.me data set.

This thesis has adopted the naïve Bayes technique which purposes to determine the conditional probability of data class. According to Raschka (2014), naïve Bayes technique is applied in different fields, since this technique is relatively robust, easy to implement, fast and accurate. For example, the naïve Bayes technique is used to analyse text documents by combining several words that appear. Naïve Bayes has also been successful in predicting cyberbullying messages from social networks such as Twitter (Sánchez et al., 2009), Formspring.me and Myspace (Nandhini and Sheeba, 2015). Therefore, the application of naïve Bayes technique in analysing Indonesian messages from social networks is employed because our study proposes to discover class label data that have the potential to be classified as cyber bullying or non-cyber bullying messages.

2.11.2 Decision Tree

Decision tree is one of the classification techniques and predictions that is famous for its application of data mining. Quinlan (1996) stated that decision tree *"is a formalism for expressing such mappings. A tree is either a leaf node labelled with a class or a structure consisting of a test node linked to two or more sub-trees"* (Quinlan, 1996). The computation of node in decision tree is based on the attribute values of data, where the probability for every outcome is linked to one of the sub-trees. When the node is tested, the data outcome determines the continuation of the process using appropriate next sub-trees. When the root is finally encountered, the prediction of data class is labelled.

Decision Tree can be referred to as a non-parametric analysis tool, designed to represent the decision rules in a form called a binary tree (Han et al., 2012). A decision tree is similar to a flowchart, however this uses the structure of a tree. Each

of the internal nodes denotes a test on an attribute while the leaf node holds a class label in which every branch represents an outcome of the test.

Many application fields such as medicine, manufacturing and production, financial analysis, astronomy, and molecular biology are established by the classification process from the decision tree algorithms. In other words, the decision tree is the foundation of several commercial rule induction systems. Developing a decision tree technique does not need domain knowledge or parameter setting, hence the reason why the algorithm is suitable and appropriate enough for the exploration of knowledge discovery as they are able to hold multidimensional data. Their structural 'tree' form that symbolizes their acquired knowledge is perceptive and accessible in general for humans to grasp.

The core of the algorithm in constructing a decision tree is referred to as Iterative Dichotomise 3 (ID3). Quinlan (1996), utilising a hierarchical and rapacious hunt, explores every potential space of branches without the backtracking process. This algorithm employs Entropy and Information Gain to build a decision tree. In detail, a decision tree is organised in hierarchy initiating from the root node and dividing the data into several subsets that are composed of similar-valued instances (homogenous). ID3 algorithm applies entropy to predict the similarity of the values or their homogeneity, based on a sample. Assuming that the sample is entirely homogeneous, the entropy is zero, however, if the sample is divided at equal, the entropy is one.

In the context of constructing a decision tree, Quinlan (1996) has proposed a formula of ID3, which is shown below:

a) Entropy:

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

Entropy is a measure that calculates the amount of uncertainty in the (data) set *S*; in other words, an entropy represents the (data) set *S*, where:

- *S* - The recent data set where the entropy is being measured, meaning that each iteration of the ID3 algorithm is altered.
- *X* - Set of classes in *S*
- *p(i)* - The ratio between the number of elements located in class i and the number of elements in set *S*

The set *S* is completely categorised when H(*S*) =0 (meaning that every element in *S* is based on the same class).

The entropy is calculated for every remaining attribute in ID3. When the attribute has the smallest entropy, it is applied to divide the set S on this iteration. In other words, the higher value of the entropy, the higher chance to advance the classification.

b) Information Gain:

$$IG(A, S) = \sum_{c \in T} P(c)E(c)$$

Information gain *IG(A)* measures the distinctness in entropy from the prior process to the after process of splitting set *S* on an attribute *A*: namely, the reduction in the amount of uncertainty in *S* following the process of dividing set *S* on attribute *A*.

Where,

- *H(S)* - Entropy of set *S*
- T - The subsets built from the division of set *S* by attribute *A*, as shown: $S = \bigcup_{t \in T} T$
- *p(t)* - The ratio between the number of elements located in class t and the number of elements in set *S*
- *H(t)* - Entropy of subset *t*

The information gain can be measured for every attribute that remains in ID3. When the attribute has the largest information gain, it is implemented to divide set S on this iteration.

Kontostathis et al. (2010) used a decision tree technique such as J48 classifier to construct and develop a decision tree that aimed to identify whether the coded dialogue that they crawled from their data was from a predator or a victim in the context of cybercrime and C4.5 (Quinlan, 1996) to distinguish between Perverted-Justice.com and Chat Track transcripts. Their research located several articles associated with approaches (J48 and C4.5) to cybercrime detection. Using those two approaches, they classified articles either into cyber predators or cyberbullies. Similar to previous research, Reynolds et al. (2011) further developed this process in applying decision tree to identify and locate messages to be categorised as cyberbullying or non-cyberbullying. They collected data from Formspring.me which had a high content of cyberbullying in the question-and-answers provided, and obtained 78.5% accuracy in their identification of true positive cyberbullying messages.

Similar work by Dinakar et al. (2011) focused on detecting cyberbullying content from the data gathered on YouTube video comments. Their research implemented two levels of classification: the first identified content of sensitive topics such as sexuality,

race or culture, intelligence, and physical attributes from YouTube comments, and the second defined what the topic discussed. With a successful experimental result of 61% accuracy, their research identified cyberbullying content from YouTube comment using decision tree, specifically J48.

Hence, determining an approach that satisfies classification of messages either as cyberbullying or non-cyberbullying is an important core of this research. Therefore, the decision tree is a suitable approach for constructing and developing a predictive model that calculates a target variable value through learning basic decision rules obtained from data features (Song et al., 2014).

The decision tree technique was employed in this research to validate the cyberbullying and non-cyberbullying classification of messages derived from the social networks, specifically Indonesian Twitter messages. This approach not only accomplishes analysis and prediction in the form of a tree diagram (based on the decision rules with no special statistical hypothesis attached), but has also been successful in verifying the variables relationship that has a significant impact on the dependent variables against a number of independent variables (Hoover and Miller, 2016). Furthermore, this approach is efficient in categorising texts from several articles (Kontostathis et al., 2010), along with highly accurate identification of true positive cyberbullying messages (Reynolds et al., 2011).

2.11.3 Neural Network

One of the data mining techniques that is based on the brain's method of processing information is known as an Artificial Neural Network (ANN) or a Neural Network. ANN is a computational process in a mathematical model based on a biological neural network, and is a simulation of the biological neural system (Singh

and Chauhan, 2009; Zhang, 2009; Gaur, 2013). This simulation involves a highly interconnected group of artificial neurones that compute processed information using a connectionist approach.

This section explains the concept of the neural network that is adopted in data mining techniques to compute information processes such as the prediction of data classification (Krizhevsky et al., 2012; Mather and Tso, 2016), pattern classification (Lippmann, 1989; Ripley, 2007), time series analysis (Gao and Er, 2005; Khashei et al., 2008), and cluster analysis (Su et al., 1997; Cao and Li, 2009). In quantitative modelling, the neural network is particularly used as a tool to calculate the fundamental relationship between a set of variables or patterns in data (Zhang, 2009).

Neural networks are also popularly used in data mining research due to their powerful modelling capability for pattern recognition and classification (Ripley, 2007). There are two reasons why neural networks are suitable and valuable for data mining. The first is that various unrealistic prior assumptions are not required in neural networks involving data generation processes and specific model structures. Moreover, the modelling process in neural networks is flexible and adaptable. Also, the model primarily learns from the occurrence of network patterns by the acquired data from the learning stage (Zhang, 2009). The second is that in approximating and representing various complex relationships, the mathematical method in the neural network has been proven to be highly accurate, so that this method can also assist with the development of the theoretical framework (Rojas, 2013). Therefore, the neural network approach is an appropriate data mining technique when there is a large amount of data and meaningful information needs to be discovered.

In practical terms, neural networks can be used as tools in statistical data modelling. The technique can be implemented to model complex prediction of data

classification, which goes from inputs to outputs (Singh and Chauhan, 2009) in which inputs can be text data, for example, in the transition from character to sentence level information. This performs a sentiment analysis of short texts using a new sophisticated convolutional neural network developed by dos Santos and Gatti (2014). They have created a neural network architecture that combines character-level, word-level, and sentence-level representations to conduct and to calculate sentiment analysis. Similar research conducted by Kim (2014) has applied a neural network loop positioned on the top of pre-trained word vectors for tasks requiring sentence-level classification. The results from Kim's study contributed to the well-established evidence whereby unsupervised pre-training of word vectors is an essential ingredient in deep learning for the neural learning process.

Many prominent approaches to text analysis have utilised machine-learning techniques to develop classifiers (Read, 2005; Kotsiantis et al., 2007; Zhang et al., 2008). Of the various machine-learning techniques, support vector machines and naïve Bayes are commonly applied (Pang and Lee, 2008), However, artificial neural networks have also been applied successfully for the analysis of texts (Ghiassi et al., 2013). In research conducted by Ghiassi et al. (2013), the artificial neural network technique was used to analyse messages from social networks, which was data-driven whereby the words in messages were presented as vectors of zeros and ones (Ghiassi et al., 2013; Severyn and Moschitti, 2015). Subsequently, in the generated input matrix, (number) one indicates the presence of the feature words in messages and zero indicates the non-existence of words. Meanwhile, the dependent variable in the neural network technique is the class designation of messages.

To determine the class label of a message according to its subject, a training data model is needed. The vector representation of each message must be generated

in the training process together with its class label (Han et al., 2012). The input matrix is created when the feature set of data is determined. The class designation of messages for the training dataset, however, can assist human evaluators to detect and to determine the class label of messages in a database. Similar to other class designations of messages, the training dataset is still being employed and a separate testing dataset is implemented to train the model and to assess its accuracy.

The advantages of using neural networks for data classification are listed below:

- High Accuracy: The ability to approximate complex non-linear mappings by the neural network (Lisboa et al., 2000).
- Independence from prior assumptions: No prior assumptions are made relating to the data distribution or the interacting forms between factors created by neural network (Yang et al., 2008).
- Noise Tolerance: Neural networks are flexible in terms of incomplete, missing and unimportant data (Singh and Chauhan, 2009)
- Application to parallel data. If the element of the neural network cannot achieve results, the process can still be carried on with no issues due to their parallel nature. At the earlier level, ‘blame’ is nominated for the local error on the neurones, which will then provide greater responsibility to neurones that are linked by stronger weights. The steps above are to be repeated at the next level, using each one’s ‘blame’ as the basis of its error (Dumitru and Maria, 2013).

On the other hand, the disadvantages of using neural networks are as follows:

- There are no basic methods for establishing the optimal number of neurones that are needed to solve a problem (Tu, 1996).

- It is quite challenging to decide on a training data set that comprises all details pertaining to the problem that needs to be resolved (Dumitru and Maria, 2013).

The neural network technique has many advantages when analysing the text in a short document; however, it is still challenging and needs to be explored in depth due to the huge amount of contextual data in social networks (dos Santos and Gatti, 2014).

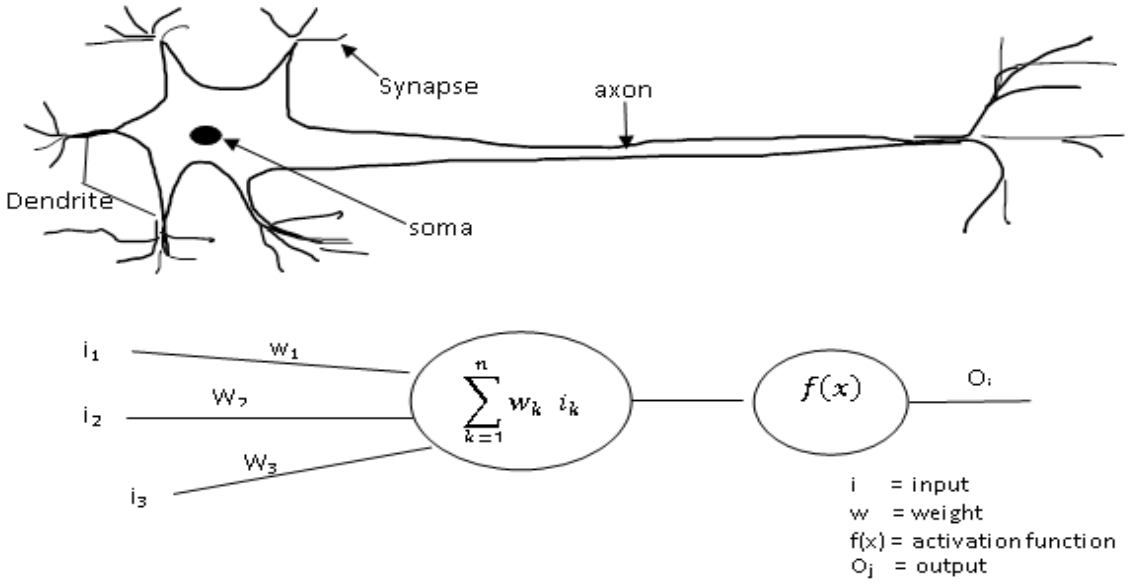


Figure 3 Simulation of a Neural Network in the Human Brain Adapted from Gaur (2013)

The basic architecture of a neural network is comprised of elements based on the interconnections and interactions in the human brain that are commonly known as neurones (Figure 3) (Robert, 1999).

The essential neuron employed in neural network techniques consists of nodes classified into three layers: input layer, hidden layer, and output layer (Figure 4). The input layer contains the origin of the information, the hidden layer contains the activation function that can be either single or multiple layers, and the output layer indicates the overall calculation results from the algorithm (Yang et al., 2008).

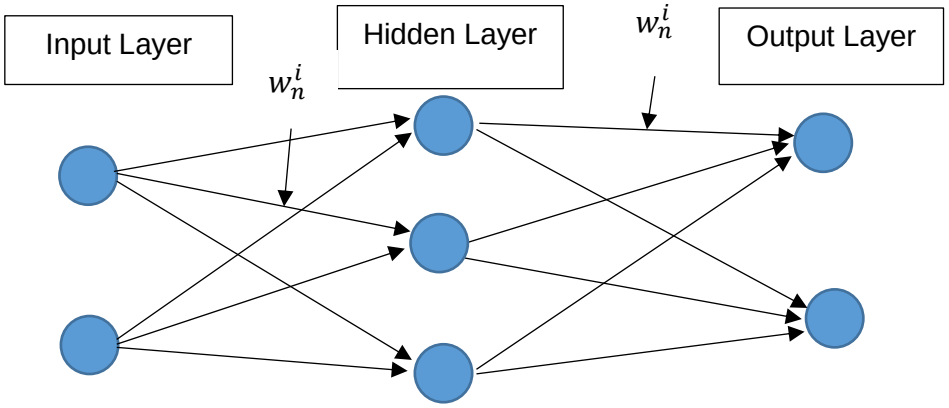


Figure 4 Simple Diagram of Neural Networks adapted from Gaur (2013)

The individual nodes in one neural network mirror the biological neurones as they acquire input data which then undergo basic processes and the selected results pass on to the other neurones (Zhang, 2009). In order to reach the hidden layer, a necessary weight value has to be supplied to discover the iterative data flow via the network. The required weight value has often been correlated to the individual nodes in the networks and constricts how the input data are connected to output data (Müller et al., 2012).

To begin the algorithm of a neural network, a node associated with the weights in the network has to be created. In this case, the weight is represented as w_n^i which is the input to the output via the l^{th} layer (Robert, 1999). An example is given in the diagram below, illustrating the weight on a connection, starting from the input neuron located in the first layer to the output neuron placed in the third layer of a network.

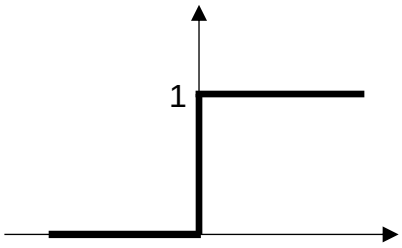
The hidden layer section of the neuron comprises an activation function (*f*), in which the total weight of the inputs and *f* is a threshold or a bias value. The weights are loaded to parts of small random values, meanwhile receiving updates in the training period (Robert, 1999). Below shows the weighted sum.

$$W_{sum} = \sum_{k=1}^{n} weight_k \; x \; input_k$$

There are various activation functions which can be selected and employed such as the Unit step (Weisstein, 2002b), Gaussian (Weisstein, 2002a), and Sigmoid (Weisstein, 2002c). The following depictions show the different characteristics of these three functions:

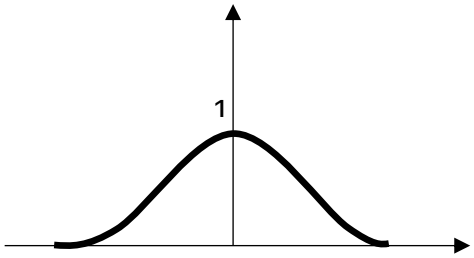
- Unit step. Depending on whether the sum of the input is greater or less than the sum of the threshold values, the output is decided at one of two levels (Weisstein, 2002b).

$$f(x) = \begin{cases} 0 \; if \; 0 > x \\ 1 \; if \; x \geq 0 \end{cases}$$



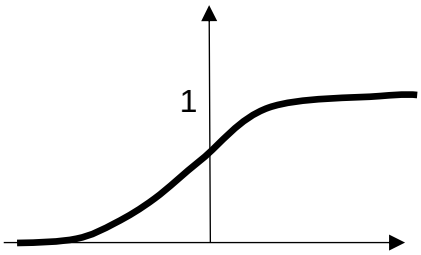
- Gaussian. The Gaussian functions are depicted as bell-shaped, continuous curves. The node output (high/low) is perceived as class membership (1/10); however, counting the closest distance between the net input and a chosen average value (Weisstein, 2002a).

$$f(x) = e^{-x^2}$$



- Sigmoid. The sigmoid function will solely generate positive numbers ranging from 0 and 1. In this case, the sigmoid activation function is very valuable for training data, and is also placed between 0 and 1. The sigmoid activation is one of the activation functions that is often applied (Weisstein, 2002c).

$$f(x) = \frac{1}{1 + e^{-x}}$$



The sigmoid function is used in our work because its derivatives are simple enough to calculate (Basterretxea et al., 2004) . The derivatives are required for the gradient calculation process, as this is necessary for training the neural network. Furthermore, the output is constantly between 0 and 1, which can be considered as a probability estimation. In addition, the derivatives do not overstate the activation as the activation of the training data in the network will cease to increase if they are confined (Kros et al., 2006).

2.12 Conclusion

This chapter has discussed issues related to the process of analysing Indonesian cyberbullying on social networks. In order to identify and understand the messages in the social networks that have the potential to cause harm, association rules of FP-growth and cosine similarity and classification techniques of naïve Bayes, decision tree and neural networks are applied as tools. The term ‘cyberbullying’ has been defined and the characteristics of cyberbullying have also been explained as a reference for analysing messages from the social networks examined in this study. Association rules techniques of FP-growth, cosine similarity, and classification techniques of naïve Bayes, decision tree and neural networks were discussed separately to provide an insight into their differences. The discussion above was

intended to provide an understanding of the cyberbullying issue, especially in Indonesia with respect to this study, and the techniques that this study utilizes.

In the next chapter (Chapter 3), association rules techniques of FP-growth and cosine similarity, which are applied in this study, will be explained in more detail. The techniques are based on the data mining framework from Han et al. (2012), and Larose (2014). Examples include the Indonesian stemming dictionary for the identification of social networks (Asian et al., 2005; Adriani et al., 2007), finding the relationship between words in data-driven text (Li and Han, 2013), and detecting sentiment messages on Twitter (Severyn and Moschitti, 2015). The discussion will include an explanation framework of the techniques; the main section in Chapter 3 discusses the study's application of the techniques.

Chapter 3
Analysing Indonesian Cyberbullying on Social Networks using Association rules

3.1 Introduction

In Chapter 2, a literature review was carried out to build a theoretical framework for this study. The theoretical framework of data mining in this study is used as a method for analysing data text. This chapter explains the methodology and the application of association rules techniques in such processes as the analysis and discovery of hidden interesting relationships among the data in a database. Generally, association rules techniques of FP-growth and cosine similarity are considered algorithms to be used in discovering hidden and interesting data relationships in relation to cyberbullying messages from social networks.

The main objective of this chapter is to discover patterns of insulting words that often appear in cyberbullying messages. These patterns can be found from the analysis of messages derived from Twitter. An analysis of the patterns of insulting words will enable the author of this research to interpret results and draw conclusions about the ways in which cyberbullying takes place in the Indonesian community via social networks.

This chapter presents a design of the analysis model of social issues, particularly in finding cyberbullying patterns in Indonesia, and an application of the advanced association rules mining techniques. The analysis results can serve as a resource for future research that specifies the development of data mining techniques and social analysis.

Chapter 3 has been organised into several sections. The first section will justify the operations of data collection, data cleansing, and data processing. The second section consists of the association rules techniques basic concept as well as the algorithms such as FP-growth and cosine similarity. The final section will discuss the chapter's conclusion.

3.2 Data Collection

Research involves the collection of data. Much of the data collection work for this research involved obtaining data from Twitter and preparing it for analysis. This section explains in more detail the data collection and data cleaning process used to prepare the data for later analysis using data mining techniques.

Research by Dinakar et al. (2011) used 50,000 posts for cyberbullying analysis. Similar work by Kasture (2015) implemented 1,313 posts collected from Twitter for cyberbullying detection processes. Moreover, research by Squicciarini et al. (2015) crawled a total of 16,000 posts for both cyberbully detection and identification of cyberbullying interactions. Therefore, to obtain the data in this research, the author crawled 152,843 messages and extracted information from Twitter.

In downloading data from Twitter, this study used an open source which is Rapid Miner software[1]. This software was able to assist the author to extract text from Twitter. In order to capture messages from Twitter, an authentication number was necessary by requesting a number through Twitter API. This Twitter API offers a streaming API and two distinct representational state transfer (REST) APIs. Users are now able to acquire real-time access to sampled and cleaned tweets through the

[1] Rapid miner software version 5 is available at http://rapid-i.com/content/view/26/201/.

streaming API. Moreover, the API is based on HTTP, and requests to GET, POST and DELETE can be applied to obtain data (Kalucki, 2010). According to Twitter terminology, every message depicts the status of a user. With streaming API, users are allowed to obtain subsets of public status descriptions at approximate real time, along with replies and tags generated by public accounts. However, status descriptions that are generated by secured accounts and the entire direct messages cannot be acquired. One interesting feature of the streaming API is that it can refine status descriptions while employing quality metrics shaped by common and constant status updates, etc. Additionally, as the API applies basic HTTP authentication, a valid Twitter account is compulsory. Furthermore, the data acquired can be in XML or the more concise JavaScript object notation (JSON) format that can be easily understood and broken down since each line finishes with a carriage return that is demarcated (Bifet and Frank, 2010).

The following were the stages of downloading data from Twitter using Rapid Miner connected to Twitter API:

1. The first stage was to link the author's Twitter account
    a. After setting up a new process in Rapid Miner studio, the author of this research dragged the search Twitter operator in the process view format, and selected the operator. By selecting the Twitter icon in the parameters view resulted in the manage connections menu dialog being opened.
    b. The add connection button located in the lower left of the window was selected, then the author established a name for the new connection. Next, the author clicked on the connection type to set into the Twitter connection. By clicking on create, a new Twitter connection was established.

c. Placed on the right side of access token field, the request access token button was selected.

d. After selecting the request access token, the Twitter website was opened in the author's browser. By this stage, the author had logged on into the Twitter account. This can be done by manually copying the uniform resource locator (URL) through selecting the show URL button instead. A pop-up message was opened that requested Rapid Miner to access the author's Twitter account and the author clicked on Authorise app.

e. Then, the author copied the access token shown on the following page. With this, the author returned to Rapid Miner studio to enter the access token number and selected the complete button.

f. To test if the new Twitter connection had been established, the author clicked on the test button located at the bottom of the manage connections panel. If the test failed, this meant that the copied access token was not completed, meanwhile if the test passed, the access token process had been completed.

g. The author then closed the manage connections dialog after clicking on save all changes.

2. The next stage was to search for tweets containing phrases associated with cyberbullying.

a. Placed in the process view, the author clicked on the search Twitter operator to find all tweets consisting of specific phrases as well as the tweets' meta data.

b. On the connection drop down menu in the operator parameters, the author clicked on the established Twitter connection. To find phrases associated

with cyberbullying, the author filled in the query field to search specific Twitter phrases.

c. The author ran the process and observed the results.

Every message that the author collected contains the following meta data: row number, date of when the tweets were sent, users' names, language, text (tweets), location and retweet count. The example of data can be seen in Table 3. More clear examples of data sets can be seen in the appendix.

Table 3 Example Data Set

No.	Time	User	Language	Text	location	re-tweet count
1	Fri Sep 02 15:53:06 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT, BAJINGAN Ahok.!!!! (Ahok, you are a bastard, rascal)	Jakarta	5
2	Fri Sep 02 15:51:49 AEST 2016	hidden	in	@syofaasavr bangsat anjing babi ni anak gasopan amat sama adenya lupa trus.musnah kek lu (What a bastard, dog, pig, very impolite with his brother and even forgot about him. You should kill yourself)	Bandung	0
3	Fri Sep 02 15:50:58 AEST 2016	hidden	in	ini juga Bangsat babi anjing sih, pegawai RS yang nolak pasien BPJS (The hospital officers are bastard pigs, dogs, they rejected a patient)	Jakarta	0
4	Fri Sep 02 15:50:29 AEST 2016	hidden	in	RT @AHMADDHANIPRAST: Cuma si Bangsat babi yg gak mau cuti masa kampanye...pendukungnya idiot anjing ...tragis #ADP (Only this bastard pig did not want to take on leave during his campaign, and all of his supporters are idiot dogs. What a tragic)	Tangerang	85
5	Fri Sep 02 15:50:10 AEST 2016	hidden	in	RT @OKWSsu: Rupamu Lugu Lugu ,Bangsat kaya anjing jga -'- (Your face looks like a bastard dog)	Bekasi	0

For further explanation, the row number represents the order number of tweets saved in a database. While, the date of when the tweets were sent provides

information of day, month and year. Furthermore, the users' names consist of the names provided by the users. The retweet count is the number of the particular tweets being resent. In addition, the language column is the type of language used in the tweets. Moreover, the tweets comprised of the messages sent. Last but not least, the location displays the regional area where the tweets were sent on social networks. These data were downloaded from August to November 2016 and recorded in a database.

The data collected by the author were transformed into tables in spreadsheets of Microsoft Excel and stored by comma separated values (CSV) format. The purpose was to allow the data to be stored in a table structured format. Therefore, this provides easy access for the data to be processed by any applicable software and data can be extracted anytime with the assistance of a basic and simple program. Following this stage, files of data can be exported into repositories in Rapid Miner.

As part of the data collection process, this research also involved ordering the Indonesian insulting terms. The various Indonesian insulting terms were retrieved from the crawled data. From the total data that were downloaded, twenty-three terms were considered to be interesting in context-based mining because they often appeared in messages with cyberbullying content, thus falling into the category of insulting terms. These insulting terms ultimately led to cyberbullying detection from the Indonesian cyberbullying messages for the next analysis process. This similar approach of cyberbullying detection using bad words to generate queries in the English context was implemented by Reynolds et al. (2011) and Kontostathis et al. (2013). The number and the density of these twenty-three terms are illustrated in Figure 5. Next, the twenty-three insulting terms were translated for meaning with the assistance of Indonesian

dictionaries. The total occurrence of insulting terms was 10,780,743. This indicates the density of twenty-three insulting terms were scattered throughout the data.

Data obtained by the author were retrieved using the *processing document* operator to display the density of each term occurring throughout the total document. Then, the results were displayed in the "word list" data as depicted in Figure 5. The order of the terms is not based on the largest number of frequency or the meaning, rather, the terms are arranged in alphabetical order.

Word	Attribute Name	Total Occurences	Document Occurences
		10780743	150299
anjing	anjing	56262	49007
babi	babi	13876	13256
bajingan	bajingan	7955	7525
bangsat	bangsat	56241	49988
bejad	bejad	260	260
brengsek	brengsek	4247	4187
budek	budek	450	410
buta	buta	261	251
geblek	geblek	600	570
gembel	gembel	251	241
gila	gila	2816	2545
goblok	goblok	10857	10147
iblis	iblis	9223	8513
idiot	idiot	8070	7780
jelek	jelek	5733	5293
kampungan	kampungan	60	60
keparat	keparat	36595	35885
kunyuk	kunyuk	1087	1057
monyet	monyet	7383	7201
sarap	sarap	1032	992
setan	setan	21150	18989
sompret	sompret	30	30
tolol	tolol	7101	6741

Figure 5 Word List of Total Occurrence of Terms in Data

For an extensive understanding in respect to the English meaning of the insulting words some dictionaries are available, such as the Indonesian dictionary (Setiawan, 2016) and the online Indonesian dictionary (Indahnesia, 2016). *Kamus Besar Bahasa Indonesia* (A Great Dictionary of the Indonesian Language) (Setiawan, 2016) and *Kamus online* (The Online Indonesian Dictionary*)* (Indahnesia, 2016) were

used to translate the words into English. The main reason for using two dictionaries was to confirm the exact meaning of each term. The online Indonesian dictionary is a free online dictionary that translates from English to Indonesian as well as from Indonesian to English. The translation process involved:

1. Typing up the words that are to be translated on the search query menu;
2. Clicking the search button; and
3. The translated terms are displayed in the results menu next to the search query menu along with the context explanation of the use of the words.

For example, the term *anjing* was translated as a mammalian animal kept for housekeeping, hunting and so on, and also interpreted as insulting victims since their attitudes are represented as a dog. The table below contains more translation of the various insulting terms.

Table 4 Translation of Indonesian Insulting Words from Indonesia to English Language

Insulting words	Indonesia	English
Insulting words related to animals	- Bangsat - Anjing - Babi - Monyet - Kunyuk	- Bastard - Dog - Pig - Monkey - Monkey
Insulting words related to stupidity and Psychology	- Goblok - Idiot - Geblek - Gila - Tolol - Sarap - Kampungan	- Stupid - Idiot - Fool - Mental disorder - Stupid - Crazy - Hick
Insulting words related to disabled persons	- Buta - Budek - Jelek	- Blind - Deaf - Ugly
General Insulting words	- Setan - Iblis - Keparat - Gembel - Brengsek - Sompret - Bajingan	- Satan - Devil - Cursed - Poor - Bastard - Jeepers - Scoundrel
Insulting words related to attitude	- Bejad	- Depraved action

In this scenario, the author did not directly translate each term; instead, both dictionaries were used for this task. If the translated terms from both online Indonesian dictionaries produced similar definitions, this indicates that both dictionaries were consulted. However, when there is disparity in the translated terms, this indicates that the author of this research had used only the Indonesian dictionary (Echols et al., 1994).

3.3 Ethical Considerations

This research was approved by the Victoria University Human Research Ethics Committee. As the nature of this thesis involves cyberbullying messages, this research design has engaged in an extraction process solely of the messages from social networks and without any attributes that connected messages to users' identities. Therefore, this thesis posed a minimal risk to an identity exposure of participants, as only the messages were being collected as data. Each message can be traced to its initiation point, which means the users can be identified. However, in order to protect their private information, the risks to participants were minimised through ensuring the confidentiality of the data and anonymity of the authors of the messages. Only the author of this research is able to trace the identities associated with the data collected.

To some extent, research that involves data collection associated with human or animal populations must encompass ethical considerations (David, 2009). The research in this thesis was undertaken according to the ethical guidelines for secondary data that are downloaded from social networks without users' consent form. Because Twitter permits an open access, data can be collected freely and anonymously. Moreover, the users' tweets had been published on social networks and were collected as the data for this research, hence their tweets were no longer

confidential. Therefore, the users' informed consent was not required in conducting research for this thesis.

While this thesis is exploring cyberbullying messages that have been successfully collected, attributes of the messages in the process of data collection also naturally appear. Therefore, the attributes from the messages were kept as confidential as they contain personal information about the senders of the messages. Accordingly, the author has retained a professional, objective posture through all the research activities including extracting tweets as data collection from Twitter, processing and cleaning data, and ensuring confidentiality of the users' identities. Objectively, the author has maintained accuracy in crawling and processing the collected data by extracting messages associated with cyberbullying, while unrelated messages were ignored and deleted, such as topics about entertainment, sports, and news.

For security purposes, apart from the author of this research electronic data was only shown to the supervisors. Data is securely maintained in the author's drives protected by a strong password and placed on the author's desk top in the author's locked office. The identities of users were saved in a coded file to ensure security of the users' identities.

3.4 Data Processing

3.4.1 Data Cleaning

The data obtained by the author contained many unimportant and inconsistent terms, given that the data collected can have abbreviations, punctuation marks and emoticons because users are able to freely and openly express their opinions in sending tweets on Twitter. However, tweets have a limit of 140 characters that can be

sent for each post on Twitter. Nevertheless, users can repetitively retweet unlimited posts daily. Therefore, the data collected must be cleaned from abbreviations, punctuation marks and emoticons because data analysis process typically requires a specific subset of data that meets certain conditions.

The process of cleaning data contributes to the flexibility of constructing data source design and mining structure to ensure that each mining structure is built based on a data source. In the other processes, such as training and testing different models, cleaners can be applied to only a portion of the data, so that it is not necessary to create a different structure for every data subset.

The files of data that have been stored in CSV format were exported into data repositories. Then, the retrieve operator in Rapid Miner was used to retrieve files of data from repositories. The aim was to fully access the meta data stored in repositories.

By means of the retrieve operator function in the Rapid Miner menu, the local database can be accessed and read from the repository. An easier way to load an object from the repository is to drag and drop the required object from the repositories view. This will automatically insert a retrieve operator with the correct path to the desired object. This operator has no input port. All it requires is a valid value in the repository entry parameter. Hence, the parameter provides an entry guide to the repository that will be restored as this operator's output.

The objective of the retrieve operator is to re-establish every file access as this operator offers a complete meta data processing function, facilitating the use of Rapid Miner. Unlike accessing a raw file, the retrieve operator presents all the meta data, therefore all meta data transformation processes are achievable (Akthar and Hahne, 2012).

Every reliable value in terms of the retrieve operator has to be clarified for the loading process of an object to continue. This parameter serves as reference for an entry in the repository that will be restored as this operator's output. Moreover, repository locations are worked out close to the repository folder that consists of the current process. In other words, repository folders are separated by a forward slash, while a ".." references the parent folder. A leading forward slash represents the base folder of the repository comprising the current process (Klinkenberg, 2013). Meanwhile, a leading double forward slash is a symbol for an absolute direction initiated with a repository name (Akthar and Hahne, 2012). In the process of retrieving files of data, the data's type of attribute in Rapid Miner were converted into text form.

The following was the process of loading data using retrieve operator from the repository in Rapid Miner:

- The author of this research set onto the repositories view and selected the samples in the repository.
- After clicking on the small plus sign in front of the repository, data and process folders were opened.
- Then, the author of this research chose the data folder containing collection of datasets with a file name of cyberbullying.
- The cyberbullying file was dragged to the white centre frame named as process to release the data set where it was transformed into a blue-coloured retrieve operator located on the right.
- By this step, Rapid Miner automatically transformed the dataset in the repository.
- The operator button was selected to open the parameters view. This pointed to the parameter that indicates the data location.

- The data were retrieved from the repository and became available for the next process.

The following is Figure 6 depicting data retrieval from the repository.

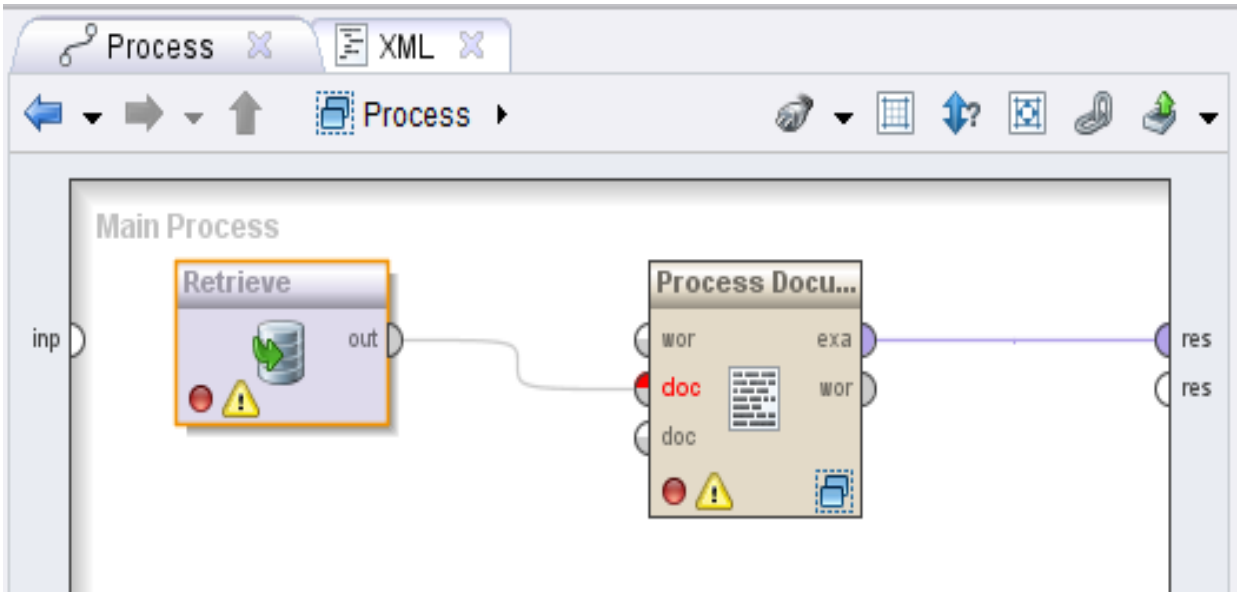


Figure 6 Retrieve Data from Repository

After the data were retrieved, the next step was designing the cleaning process. This process involves data operators and is a stage in the data cleaning process. Designing the document process is one part of the data cleaning process. The aim of designing the document process is to navigate the data processes by linking one process to the other processes.

The next stage is setting up the document processing of the operator by creating word vectors and selecting attributes that can be completed by designing some programs to clean the text first. In the document process operator, some cleaning processes can involve tokenize, transform case, stop words, stem dictionary, and n-grams. Figure 7 show the process of cleaning data in this research. Tokenize, transform case, stop words, stem dictionary, and n-grams are used to remove abbreviations, punctuation marks and emoticons.

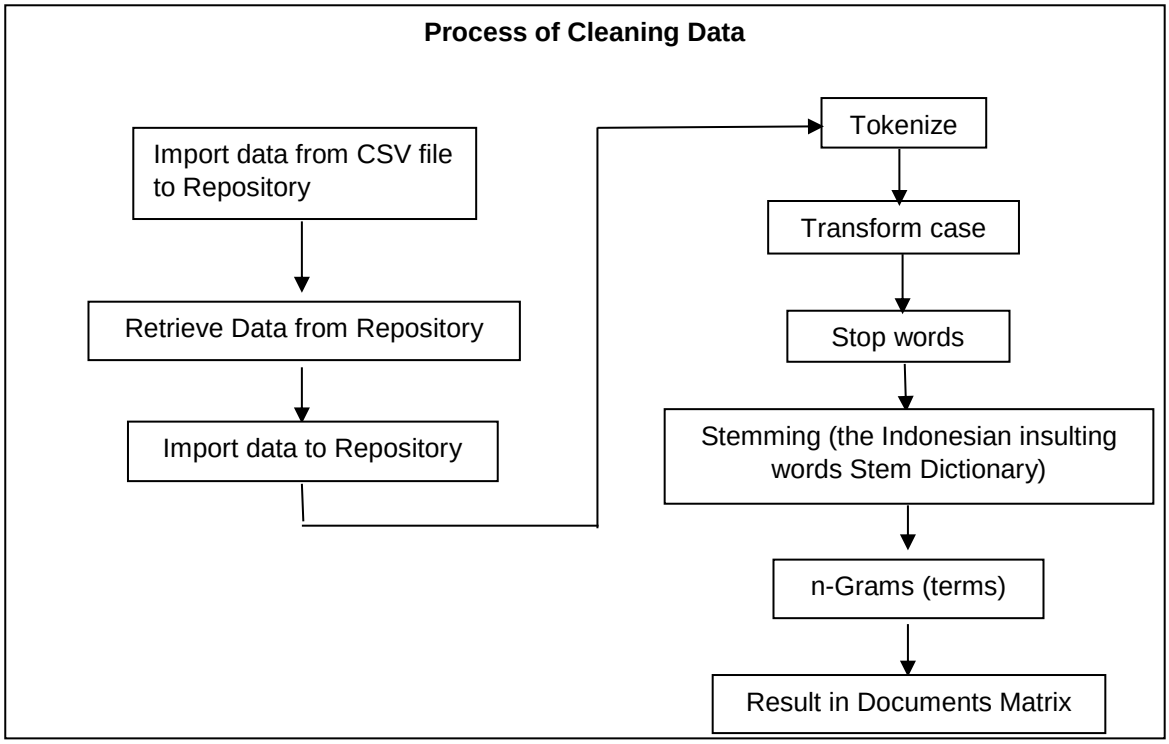


Figure 7 Process of cleaning Data

- Tokenise is the procedure of dividing a text stream into words, phrases, symbols, and other relevant elements. Using whitespace, punctuation marks or line breaks, each sentence is broken up into words that ultimately show single words or token. However, the language grammar recognises the mathematical operator, commonly referred to as a token, to be a separator. Hence, even when various tokens are grouped together, they are nevertheless still being separated by the mathematical operator. The main aim of this tokenisation process is to further explore and analyse words in each sentence. Textual data is merely an understanding of text or a chunk of characters from the starting point. Moreover, the recalling of information needs the words from the dataset (Verma and Renu, 2014).

- Transform case is the process of converting all the characters correspondingly into either lower or upper cases (Akthar and Hahne, 2012).
- The stop word eliminates every token that is equal to a stop word, but only if the file contains at least one stop word in each line. General words that restrict the process of mining text are prepositions, articles and pronouns, which are referred to as stop words. Moreover, unnecessary words are removed prior to text mining (Verma and Renu, 2014).
- Stemming, commonly referred to as lemmatisation, is an approach whereby words are compressed into their stems, base or root format. There is an abundance of English words that can be condensed into their stems; for instance: likely, liking, dislike, unlikely, have like as the stem word. Furthermore, names can be transformed to their root format by eliminating the 's', for instance, stemming the variation "stem's" in a sentence is condensed into "stem"; but this elimination can produce inaccurate stem or root words. However, the stems will not produce any issues during the stemming process if the words are not related to human interactions. Also, the stems are still useful since the whole inflections of the root are transformed into identical roots (Verma and Renu, 2014).
- N-Grams is responsible for creating term n-Grams of tokens in a file. Term n-Gram is a series of consecutive tokens with n length. N-grams creates term n-Grams that contains every consecutive token series with *n* length (McGuigan, 2013).

After the data has been cleaned, every word in every message is replaced in the word base form. The following is an example of cleaning a message using tokenize, transform case, stop words, stem dictionary, and n-grams.

Table 5 Example Cleaning Data using Tokenize, Transforms case, Stop words, Stemming, and n-Grams in Rapid Miner

Operator	Before	After
Tokenize	Avanya andaiy, avanyaâ™¥RT mohamadDFDM: "sarap!! addnan_ch: Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn aaaaahhhh an. (What are you ™¥RT mohamaddDFDM crazy. Oh, He is dog, stupid, f***, monkey, satan, pig as rascal, sh**)	Avanya andaiy RT mohamad DFDM sarap addnan ch Oh anjing, goblog fakyu siah monyet, setan, babi alas bangsat shit damn (What are you RT mohamaddDFDM addnan ch crazy. Oh, He is dog, stupid, f***, monkey satan, pig as rascal, s***)
Transforming Case	Avanya andaiy RT mohamad DFDM sarap addnan ch Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn. (What are you RT mohamaddDFDM addnan ch crazy. Oh, He is dog, stupid, f***, monkey, satan, pig as rascal, s***)	Avanya andaiy RT mohamad DFDM sarap addnan ch Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn (What are you rt mohamadddfdm addnan ch crazy. oh, he is a dog, stupid, f***, monkey, satan, pig as rascal, s***)
Stop Word	avanya andaiy rt mohamaddfdm sarap addnan ch oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn (What are you rt mohamadddfdm addnan ch crazy. Oh, he is dog, stupid, f***, monkey, satan, pig as rascal, s***)	avanya andaiy mohamadd sarap addnan anjing goblog fakyu siah monyet setan babi alas bangsat shit damn (What are you mohamadd addnan crazy. he is dog, stupid, f***, monkey, satan, pig as rascal, s***)
Stem Dictionary	avanya andaiy mohamadd sarap addnan anjing goblog fakyu siah monyet setan babi alas bangsat shit damn (What are you mohamadd addnan crazy. he is dog, stupid, f***, monkey, satan, pig as rascal, s***)	apa kamu apa mohamaddfdm sarap ada anjing goblok jancuk siah monyet setan babi ala bangsat shit damn (What are you mohamadd crazy. he is dog, stupid, f***, monkey, satan, pig as rascal, s***)
n-Grams (terms)	apa kamu apa mohamaddfdm sarap ada anjing goblok fakyu siah monyet setan babi ala bangsat shit damn (What are you mohamadd crazy. he is dog, stupid, f***, monkey, satan, pig as rascal, s***)	apa apa_kamu kamu kamu_apa apa apa_mohamaddfdm mohamaddfdm mohamaddfdm_sarap sarap sarap_ada ada ada_anjing anjing anjing_goblok goblok goblok_jancuk jancuk jancuk_siah siah siah_monyet monyet monyet_setan setan setan_babi babi babi_ala ala ala_bangsat bangsat bangsat_shit shit shit_damn damn (What are you mohamadd crazy. he is dog, stupid, f***, monkey, satan, pig as rascal, s***)

3.4.2 Stem the Indonesian Insulting Words (Dictionary)

This research created a stem (dictionary) of Indonesian insulting words based on the Indonesian stem dictionary technique that was created by Nazief and Adriani (1996). The Nazief and Adriani technique used comprehensive morphological rules;

this technique eliminates the prefix, suffix, infix, and combination of prefixes and suffixes into root words (Asian et al., 2005; Adriani et al., 2007).

Based on Nazief and Adriani's technique, the stem of Indonesian insulting words was created by eliminating the prefix, suffix, infix, and combination of prefixes and suffixes into root words. For example, the suffix "an" was removed from the word *bangsatan* to create *bangsat*. Here, the word *bangsatan* does not have any meaning; however, by removing the suffix "an" then the word becomes *bangsat* which means bastard. Another example, *aaanjinglah* is to express *anjing* (dog). By removing prefix "aa-" and suffix "-lah", then the word *aanjinglah* becomes *anjing*.

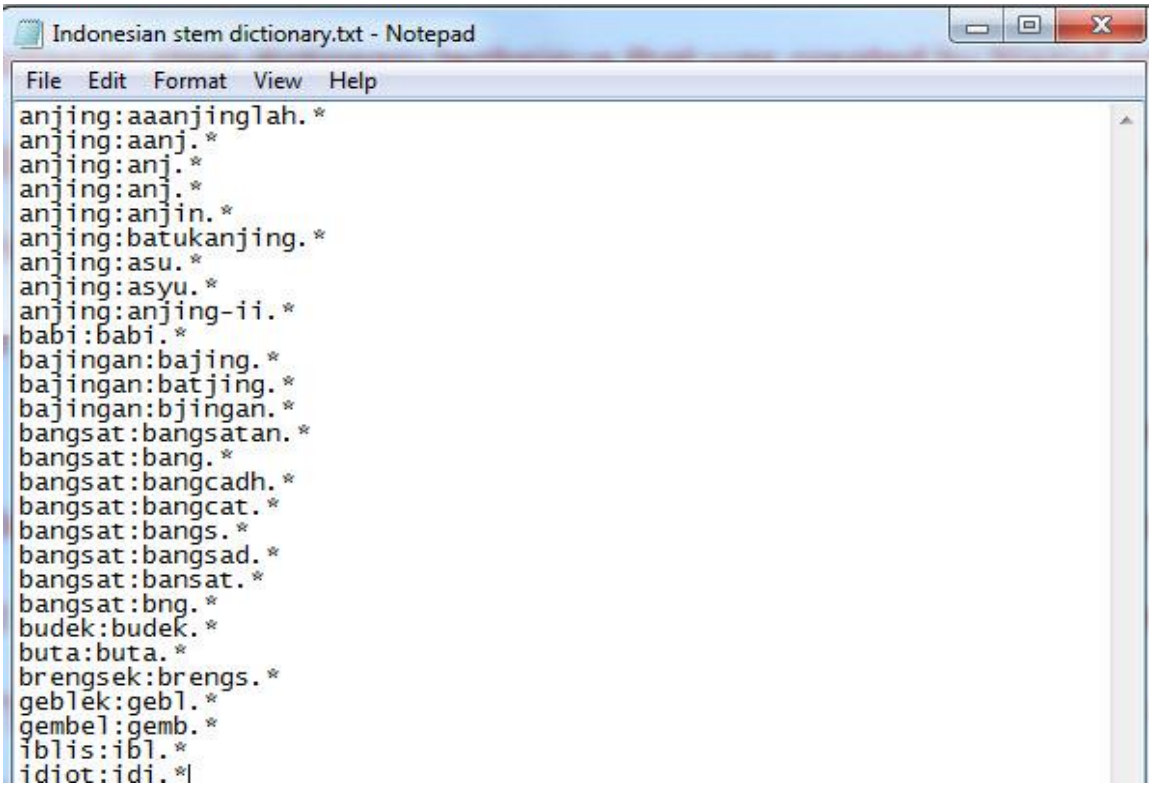


Figure 8 Stem Indonesian Insulting Words Dictionary

The stem of Indonesian insulting words is presented in Figure 8. As the focus in this research is to analyse the relationship between insulting words, the author created the Indonesian stem dictionary that exclusively contains the Indonesian

insulting words. This stem helps with the cleaning of abbreviations and punctuation marks every word in the data that is an Indonesian insulting word. The stem Indonesian insulting words is stored in txt format, since the Rapid Miner software requires documents to be in txt format.

The Rapid Miner stem dictionary operator compresses each term to its base form by applying an external file with replacement rules. In other words, the file must comprise one rule in each line, or can be expressed as: “targetExpression”: pattern1….pattern2, in which targetExpression is where its input terms are decreased, assuming that any of the terms match the patterns. In other words, pattern X is a basic string or regular expression. An example of this mapping can be: weekday: .*day.

An example of the result of stemming the message is given in Table 3 above. The stemming process can also assist with sorting the insulting words into word vectors that represent the document numerically.

3.4.3 Data Transformation

The processing document operator, and all the cleaning operators are connected simultaneously and transform the text document into word vectors where the attributes are weighted by the occurrence of binary terms normalised during the frequency calculation. The word vectors output represent each vector with its original document. Each document is represented as a vector $d_i = (w_{1i}, w_{2i}, \dots, w_{ni})$.

Every dimension represents an independent term. If a term appears in a document, it has a non-zero value in the vector. However, the value of the term will be zero in the vector if it does not appear in the document. A term can be either a single word or a phrase. Each word in the corpus that is determined to be a term is represented as a size in vector by means of weighting.

In this case, the binary term occurrence is chosen as the weight. The value 1 indicates that the term is present in the document and 0 indicates that the term is not present in the document. After the generation process of the document operator, the table of word vectors appear as depicted in Table 6.

Table 6 Example Result after Generating Processing Document Operators in Document Matrix

Row No	Text	anjing	Babi	Bajingan	bangsat	Brengsek	buta	gembel	gila
1	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	1	1	0	1	0	0	0	0
2	gila anjing bangsat (mental disorder, dog, bastard)	1	0	0	1	0	0	0	1
3	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	1	1	0	1	0	0	0	0
4	anjing goblok monyet setan babi bangsat (dog, stupid, monkey, satan, pig, bastard)	1	1	0	1	0	0	0	0
5	bangsat tolol anjing setan (bastard, stupid, dog, satan)	1	0	0	1	0	0	0	0

Table 6 shows the example data matrix produced after processing the document that the operator generated, in which every insulting word is displayed based on its appearance in every message. When the insulting word existed in a message, it is represented by 1; however, if the insulting word does not exist in the message, then it is represented by 0. For example, 1 represents the word of *anjing* (dog) in row number 1 and the word of *bajingan* (scoundrel) is represented by 0,

meaning that *bajingan* does not exist in the first record. This process helped the author of this research to numerically represent data using the association rules techniques.

3.4.4 Data Analysis

Data analysis in this research was used to evaluate data using various approaches in order to determine the relationship between the research questions and findings (Neuman and Robson, 2012). However, the author of this research was aware of the possibility that new issues could emerge during the data analysis process. Since this research focused on Indonesian cyberbullying messages, the data was collected from Twitter. Hence, this was an important source of data that required careful analysis in order to discover additional, hidden information. It was anticipated that this analysis would provide a better understanding of the social phenomenon that has come to be known as cyberbullying via social networks.

For this research, the data analysis process was a necessary stage to uncover hidden patterns and interesting information as a part of knowledge discovery process from the database. Specifically, the significance of this essential stage was related to the complexity of the data which comprised of the content of messages associated with cyberbullying, the insulting terms used in the messages and the various Indonesian local languages used. Hence, the systematic organization of data in preparation for the analysis required significant effort. Data analysis arranged from the beginning of the research and continued until the last analysis process (Neuman and Robson, 2012). This research was able to obtain an understanding of the data analysis at the beginning of the step and subsequently construct a conceptual framework.

The following is a simple example of the format of the results in the data matrix. This vector format that represents the insulting terms is useful to be analysed at the

next stage of the analysis process. This can be referred to Table 6 on the first row containing some of the insulting terms: *sarap anjing goblok monyet setan babi bangsat* (crazy, dog, stupid, monkey, satan, devil, bastard) that were transformed into word vectors and break them into a data table.

$$S_1 = (crazy, dog, stupid, monkey, satan, devil, bastard)$$

$$V_1 = (1,1,0,1,0,0,0,0,1,0,0,1,0,1,1,0,0,0,0,0,0,0,0)$$

S_1 represents the data from the first row of Table 6, while *V_1* is the term in word vectors. The purpose of this approach is to remove the grammatical structures in the Indonesian context and word order, thus simplifying the results for them to be tractable in further analysis process. For example, the grammatical structures that were removed may contain *loh, gue, bapaklu, dasar anak general, makan, jalan ke mall* (you, I am, your father, the son of a General, eat, going to the mall) as these common words do not signify essential meanings in the insulting terms context. This format of word vectors was generated by process documents of data operator in Rapid Miner.

For the data analysis, this research used association rules techniques of FP-growth and cosine similarity in order to discover any interesting relationship between the Indonesian insulting words as represented by the pattern of words indicative of a cyberbullying message. The data analysis in this research was conducted by implementing data mining techniques:

- Firstly, the FP-growth was applied to find the frequent itemsets without candidate generation. This technique derives the frequent itemsets straight from the Frequent Pattern Tree. By means of this technique, frequent itemsets are found in the data before proceeding to the next constraint.

- Second, the association rules were applied to find the relationship between frequent itemsets. This technique shows the patterns of data obtained from the FP-growth technique.
- Third, the cosine similarity measure was applied to determine the strength of the relationship between itemsets by calculating the degree of similarity between words (itemsets) in the vector. The results of this calculation indicate the itemsets that have the strongest relationship.

The use of all techniques mentioned above for the data analysis is intended to produce results that are more detailed and in-depth, enabling the author to find meaningful patterns in the data and gain more insight into the phenomenon being studied. The process of data analysis in this thesis is illustrated in Figure 9.

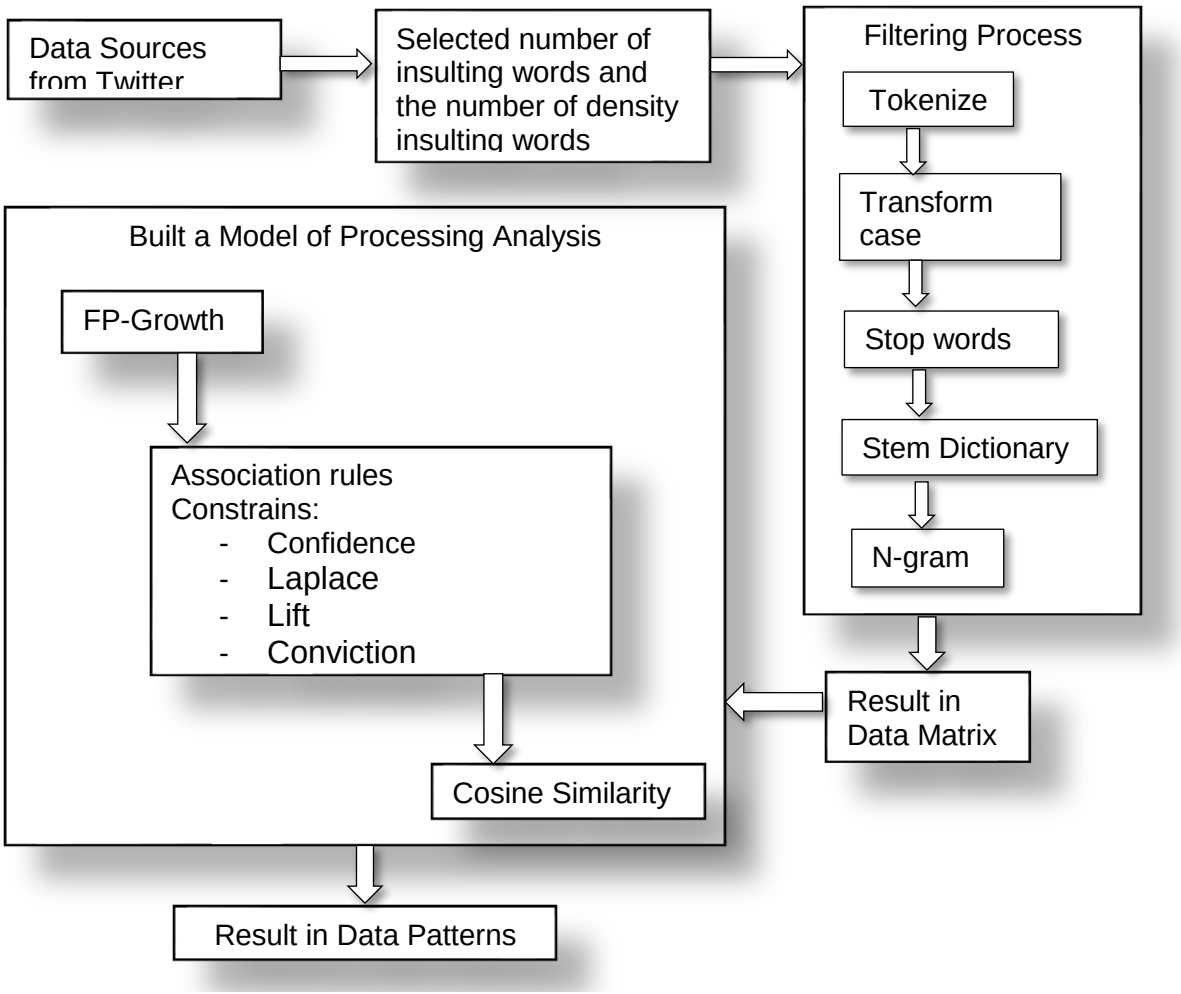


Figure 9 The Process of Data Analysis using Association Rules, FP-Growth, and Cosine Similarity

In this research, the data conceptualisation strategy was applied for the purpose of acquiring a comprehensive understanding of the phenomenon. The conceptualisation process of data analysis began with the cleaning of data which gave the author an early understanding of the data cleansing concept. In this phase of the research, the author developed a model analysis using association rules of FP-growth and cosine similarity techniques in a sequence.

3.5 Mining Patterns in Indonesian Cyberbullying Messages

One of the purposes in this research is to mine data in order to discover whether there are meaningful and interesting patterns in Indonesian cyberbullying messages. Hence, this research employed association rules techniques: FP-growth and cosine similarity to mine the interesting information about cyberbullying from social network messages.

According to Wu et al. (2008), the association rules technique introduced by Agrawal et al. (1996) is one of the most popular data mining approaches used to find frequent itemsets from a transaction dataset. Maheshwari et al. (2014) stated that the association rules technique can help to discover hidden patterns among apparently unrelated data, thereby offering a means of identifying predators in a crime that is increasing alarmingly: cyberbullying.

The strategy of applying association rules techniques to discover patterns in Indonesian insulting words is appropriate as a means of determining the relationship between the insulting words. This can be indicated by the posted cyberbullying messages. The implementation of the association rules techniques in this research are explained below.

3.5.1 The Application of Association Rules to Indonesian Cyberbullying Data

In association rules techniques, the data extracted from tweets is represented in alphabetical order in the data set. The purpose of representing Indonesian insulting words in alphabetical order is to calculate the frequency of the words that occur in the database to determine the strength of the relationship among the words. Representing data in a data set of association rules will be explained below.

All of the Indonesian insulting words should be represented in an alphabetical order. The Indonesian bullying words are listed in alphabetical order to discover the item set patterns in the Indonesian bullying word database. The list of Indonesian bullying words in alphabetical order is shown in Table 7; the letter *a* represents *bangsat*, the letter *b* represents *anjing*, etc.

Table 7 Illustration Example Data set

Item	Represent
a	Bangsat (rascal)
b	Anjing (dog)
c	Babi (pig)
d	Monyet (monkey)
e	Kunyuk (monkey)
f	Bajingan (scoundrel)

An example of the data recorded in the database is shown in Table 8. This indicates that the itemsets that occurred in the data transactions are the tweets containing *bangsat* (a), *anjing* (b), *babi* (c), *monyet* (d), *goblok* (g), and *sarap* (n).

Table 8 Illustration Example Set of Item in Data Set

Tweet ID	Last Tweet
1	a, b, c, d, g, n
2	a, b, d
3	a, b, k
4	a, b, c, d
5	a, b, m, t

Tweet ID (TID) represents a set of all tweets in the database of Indonesian cyberbullying words. I = (a, b, c, d, …, z) is a set of items which contain Indonesian bullying words. Let A be a set of items of Indonesian bullying words. In this scenario, $A \subseteq T$, an implication from $A \Longrightarrow B$ can be referred to an association between items, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \emptyset$. In order to determine whether there is a strong relationship between items, the association rules require that all itemsets should reach the minimum support threshold and minimum confidence threshold.

In this scenario, in order to manage the memory space during the process of discovering frequent itemsets, the Apriori property is applied. The purpose of using the Apriori property is for mining frequent itemsets where any subsets of a frequent itemset must also be frequent, however, when subsets are not frequent, they will be ignored (Han et al., 2012). Assuming that an itemset I does not satisfy the minimum support threshold, *min_sup*, then I is determined to be not frequent, or in another form, *P(I)* < *min_sup*. When an item A is included in the itemset I, this means that the end result of the itemset, (i.e., $I \cup A$) cannot occur more often than I. Thus, this $I \cup A$ is also not frequent, or in another form, *P(*I*)* $\cup A$ < *min_sup*.

3.5.2 Mining the Indonesian Cyberbullying Messages Using the FP-Growth

FP-Growth (frequent pattern growth) is one of the approaches used by the author of this research to discover any frequent patterns in Indonesian cyberbullying messages without candidate generation, which is different from the Apriori algorithm (Han et al., 2012). The constraints on the basis of the efficient techniques which have been specifically developed for Apriori-based mining are applicable, using frequent pattern growth (Han and Pei, 2000). However, more advanced types of constraints cannot be applied to the mining process by Apriori. An example of this type of constraint is convertible constraints, which are constraints that are able to be converted into monotonic or anti-monotonic form. Hence, constraints developed for Apriori-based mining can only be accomplished with frequent patterns growth (Pei et al., 2001).

FP-Growth executes a frequent item-based database projection when the database is large. In this circumstance, FP-growth shifts into main-memory-based mining by building a compact data structure commonly known as the FP-tree. Hence, this compact tree is mined rather than the database (Han et al., 2012). The following is a description of FP-growth construction from the database.

Table 9 Illustration Example Set of Item in Data Set FP-growth

TID	Last Tweet	Frequent items (ordered)
1	a, b, c, i, k, d, g, n	a, b, c, d, g, n
2	e, b, f, a, d	b, a, d
3	h, a, b, g	b, a, g
4	l, a, m, b, c, d	a, b, c, d
5	r, a, w, b, c, n	a, b, c, n

Let the transaction database, abbreviated to database (DB), in Table 9 have a minimum support threshold of 2. The first process involves scanning the DB to collect a list of frequently occurring items, ((a:5), (b:5), (c:3), (d:3), (g:2), (n:2)), in a frequency format of descending order. This type of ordering was essential as every path of the tree followed in the same order. The next step is creating the root of a tree mapped with “null”, including scanning the DB for the second time. The first database scan will produce the initial branch of the tree. As stated previously, the branch is listed in the order from the frequent items in transaction Table 9. In the next transaction, the (ordered) frequent item list (b, a, d) has the same common prefix (a) along the actual path (a, b, c, d, g, n); therefore, each node with the prefix is calculated in increments of 1. Moreover, one fresh node (b:1) is constructed and connected to the child of (a:2) and a different new node (d:1) is constructed and connected to the child of (b:1). To facilitate the process of tree reversal, an item header table is every item point from the head of node-links to its first occurrence in the tree. Nodes that share identical item-names are connected in an ordered sequence by node-links. Figure 10 below shows the tree with the correlating node-links after the entire DB has been scanned.

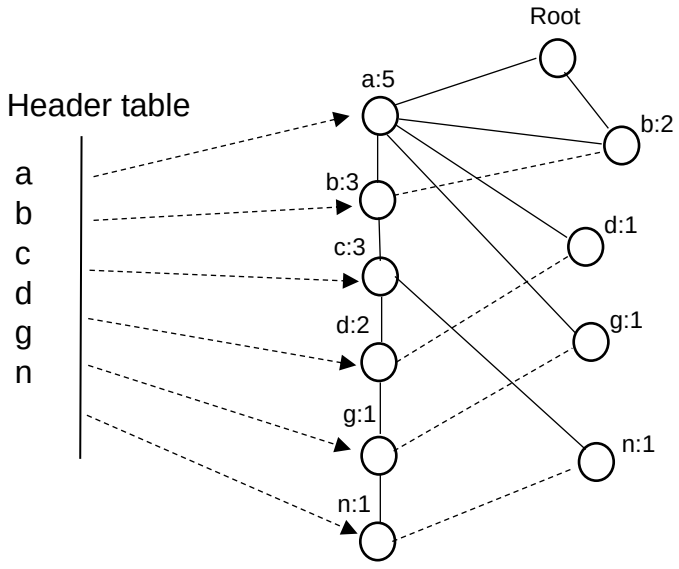


Figure 10 Illustration FP-tree for transaction database from Table 9

To calculate the FP-growth algorithm, this research has used Rapid Miner software to assist with data calculation when using the FP-algorithm. This software allows the user to drag and drop the operator including in assisting data analysis such as association rules techniques of FP-growth and cosine similarity that are used in this research.

After cleaning the data and obtaining the data matrix as described previously in this chapter, the data was converted from numerical to binomial form. The reason for this is that in the software, the FP-growth operator requires the value in binominal form. By using the nominal to binominal operator, the value changes the type of numeric attributes to a binomial form. This operator not only alters the form of selected attributes, but also outlines the entire attribute values to the correlated binominal values. In this situation, binomial attributes exclusively possess two possible values: either ‘true’ or ‘false’. Assuming that an attribute value is between the described minimal and maximal value, the attribute value is labelled as ‘false’; however, in a different circumstance, it will be ‘true’. Minimal and maximal values can be specified by the minimal and maximal parameters. When the value is missing, the new value will be missing. The boundaries are both fixed to a default of 0.0; therefore 0.0 is established to be ‘false’, and every other value is established to be ‘true’ by default. The following is an example of value conversion from nominal to binominal form by employing the nominal to binominal operator.

The result generated by the nominal to binominal operators in Table 10 shows that the value 1 in the data matrix has been replaced by true value; meanwhile, the value 0 has also been replaced by false. This means that the entire range of insulting words that appeared in every record have a true value; conversely, all of the insulting words that did not appear in every record have been given a false value.

Table 10 Example Result after Generating Nominal to Binominal Operators

Row No	Text	anjing	babi	bajingan	bangsat	brengsek	buta	gembel	gila
1	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	true	true	false	true	false	false	false	false
2	gila anjing bangsat (mental disorder, dog, bastard)	true	false	false	true	false	false	false	true
3	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	true	true	false	true	false	false	false	false
4	anjing goblok monyet setan babi bangsat (dog, stupid, monkey, satan, pig, bastard)	true	true	false	true	false	false	false	false
5	bangsat tolol anjing setan (bastard, stupid, dog, satan)	true	false	false	true	false	false	false	false

To continue the analysis process, the nominal to binominal operator was connected to the FP-growth operator. The FP-growth operator computes the entire frequent itemsets from a dataset by creating a FP-Tree data structure from the transactional database. In other words, this is a condensed form of the data, which is commonly consumed to the main memory, especially for larger databases. All of the frequent itemsets are derived from this FP-Tree.

Basically, the FP-growth operator has two simple working modes:

1. Searching for and identifying those itemsets that have the highest support rather than minimum support. Furthermore, this mode is appropriate if the minimum number of item set parameters is acceptable. After this stage, this operator locates the specified number of itemsets within the parameter of a minimum number of itemsets. The minimum support parameter is ignored in this situation

2. Searching for and identifying all itemsets which have greater support rather than the specified minimum support. The minimum support is predetermined by the minimum support parameter. Furthermore, this mode is appropriate when the minimum number of itemsets parameter is set to false.

In FP-growth operator, there is a minimum *support_count* parameter that should be filled. The author of this research used 20% of the value in minimum *support_count* as the minimum standard. According to Hu and Chen (2006), when the threshold is 20% the threshold generates interesting rules. This is supported by Leers (2011) who found that 20% of minimum support resulted in many meaningful rules. Sarma and Mahanta (2012) research has adopted 20% for the minimum support threshold and discovered many meaningful frequent itemsets. This type of approach purposes to acquire more of the rules among items in the database since a higher value in minimum support obviously leads to less frequent itemsets being identified (Leers, 2011). Hu et al. (2016) stated that when the minimum support is too high, frequent itemsets are not found. However, if the minimum support threshold value is at its lowest, then many unimportant frequent patterns are discovered (Zhong et al., 2012). Also, Hu and Chen (2006) and Hu et al. (2016) have noted that when the minimum support is too low, many meaningless rules are generated. Therefore, 20% of minimum support threshold is appropriate in considering various frequent itemsets that do not consist of too many unimportant patterns and also generate many meaningful and interesting rules.

After generating FP-growth using Rapid Miner, the frequent itemsets were identified and counted. Every insulting word (item) at this stage was calculated using the FP-growth algorithm in order to find all frequent itemsets by building a FP-tree data structure from the database. All frequent itemsets were derived from this FP-tree. In

this stage, the *support_count* of itemsets was calculated, and the result obtained was a list of itemsets with a high *support_count*. As mentioned before, this stage discovered all frequent itemsets with a support greater than the specified minimum support. The results after generating FP-growth are given in Table 11.

The important information revealed in Table 11 is that of all the insulting words, there were ten itemsets with a greater *support_count* than other items (insulting words). The ten itemsets were *bangsat* (bastard), *anjing* (dog), *keparat* (cursed), *setan* (satan), *babi* (pig), *goblok* (stupid), *iblis* (devil), *idiot* (idiot), *bajingan* (scoundrel), and *monyet* (monkey). Moreover, the frequent patterns of combinations of two items *bangsat* and *anjing* also had a higher *support_count* compared to others. In addition, there were frequent patterns with three-item combinations *bangsat*, *anjing* and *babi*.

Table 11 Frequent Pattern after Generated FP-Growth in Rapid Miner

support count	**item1**	**item 2**	**item 3**
0.333	bangsat	-	-
0.326	anjing	-	-
0.239	keparat	-	-
0.126	setan	-	-
0.088	babi	-	-
0.068	goblok	-	-
0.057	iblis	-	-
0.052	idiot	-	-
0.05	bajingan	-	-
0.048	monyet	-	-
0.27	bangsat	anjing	-
0.054	bangsat	babi	-
0.081	anjing	babi	-
0.056	setan	iblis	-
0.052	bangsat	anjing	babi

In order to compare the *support_count* between frequent itemsets which emerged after generating the FP-growth, the results of frequent itemsets were exported to CSV format. This was done in order to create a chart using a spreadsheet in Microsoft Excel. The result in chart form is depicted in Figure 11.

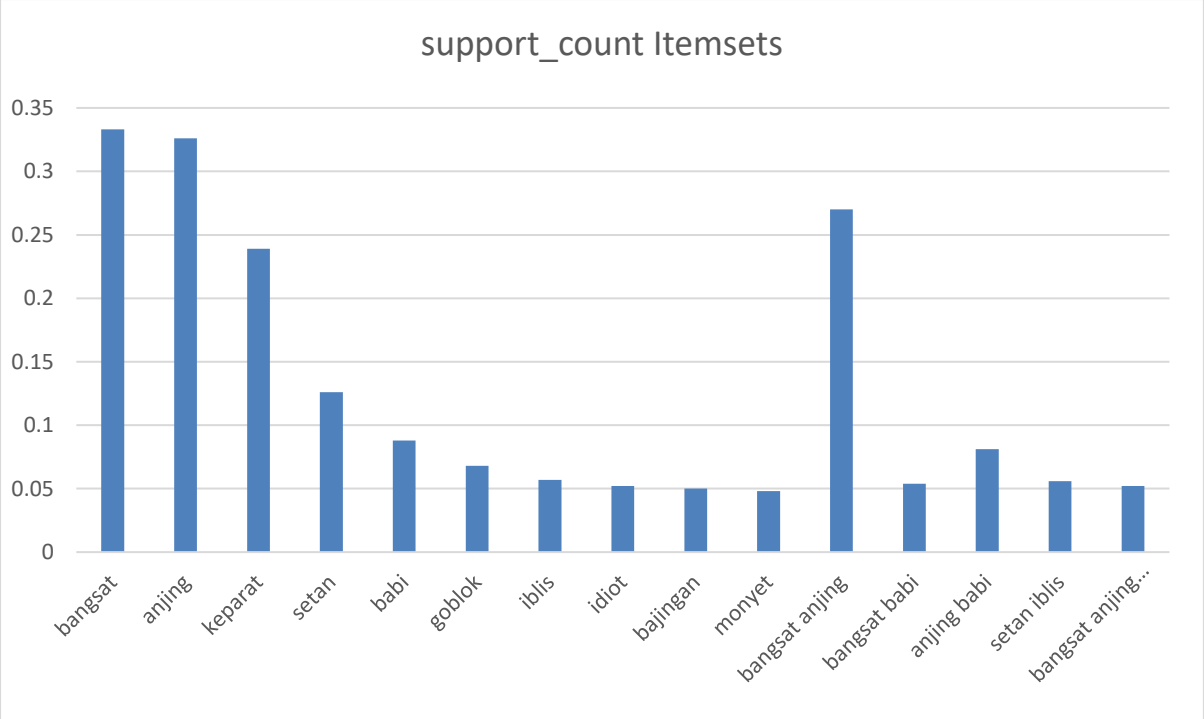


Figure 11 Frequent Pattern after Generated FP-growth in Chart

Figure 11 depicts the *support_count* of the insulting words after the FP-growth generation process. As observed, *bangsat* and *anjing* terms had a high number of *support_count* as compared to others. The combination of two terms *bangsat* and *anjing* also had large number *support_count*. This means that the terms *bangsat* and *anjing* were the two most popular terms that users often included in their messages. In comparison, *idiot* (idiot), *bajingan* (scoundrel), and *monyet* (monkey) had a smaller *support_count*.

The association rules algorithm was applied to the FP-growth results. As stated previously, this research used association rules techniques: FP-growth and cosine similarity; therefore the results produced by the FP-growth generation process were mined in depth using the association rules algorithm. In this case, the association rules were used to find the correlation between frequent itemsets produced by the FP-growth.

3.5.3 Mining the Indonesian Cyberbullying Patterns Using the Association Rules

Association rules techniques have been developed based on the market basket analysis (Agrawal et al., 1996; Han et al., 2012). Association rules are often applied in supermarkets, grocery stores, or book stores as a means of discovering the sales trend of items which are often purchased together. This enables stores to organise the layout of store items more effectively in order to increase sales by cross-selling items.

The association rules technique mainly involves 'if and then' statements, which facilitate the discovery of relationships among apparently disparate or unrelated data. An example of an association rules is: if a buyer purchases fruit, there is an 80% likelihood that the buyer will also purchase vegetables (Abaya, 2012). The association rules technique has two units: an antecedent (if) and a consequent (then). By definition, an antecedent is an identified item (or item set) in the data, while a consequent is an item (or item set) appearing together with the antecedent (Chapman and Feit, 2015).

The association rules technique involves the data analysis of frequent 'if and then' patterns and uses criteria support and confidence thresholds to discover the most significant relationships. In other words, the criteria support indicates the frequency with which items occur in the database. Whereas, confidence indicates the number of times that the 'if and then' statements are considered as true (Chapman and Feit, 2015).

This research used the association rules techniques (Han et al., 2012) to find the probability of co-occurrence of insulting words which are often used in cyberbullying messages. The discovered relationship between insulting words can be used to identify the message patterns that are indicative of cyberbullying. For example,

the user may be recognized from the recurrence of certain words in cyberbullying messages.

The association rules operator in Rapid Miner has four parameters that have to be considered. The explanations of the four parameters are as follows:

1. *Confidence* is an association rules measurement that is used to determine the strength of the relationship between itemsets (Han et al., 2012). The formula for confidence was introduced by Agrawal et al. (1996) as:

$$confidence\ (A\ \Rightarrow\ B) = \frac{support\ count\ (A \cup B)}{support\ (A)}.$$

In this case, reading expression support count $(A\ \cup\ B)$ means that support for the occurrence of both A and B items appeared in the transaction database, but there is no support for the occurrence of either *A* or *B* items in the transaction database. The value confidence threshold in Rapid Miner ranges from 0 to 1. The calculation of confidence is used to estimate $P(B|A)$, the probability of observing *B* given *A*. The support *(A)* of an itemset *A* is interpreted as the number of transactions in the dataset comprising *A* itemsets.

2. *Lift*, often referred to as Interest, was first introduced by Brin et al. (1997) who proposed the lift parameter as a correlation measure to improve the efficiency of data mining. The lift parameter measures the frequency with which A and B occur together rather than singly.

Brin et al. (1997) defined lift as

$$Lift\ (A\ \Rightarrow\ B) = \frac{support\ (A \cup B)}{((support\ (B) x\ (support\ (A))}.$$

Erlandsson et al. (2016) developed the lift measures for the interdependence ratio of the observed values.

The formula for lift introduced by Erlandsson et al. (2016) is

$$Lift\ \{(A,B) \Longrightarrow C\} = \frac{support\ (\{A,B,C\})}{support\ (\{A,B\})\ x\ support\ (\{C\})}.$$

Lift measures the extent to which A and B are independent from one another. The measurable range is from 0 to positive infinity. When lift is 1, this indicates that both the rule and items are individual and therefore independent. However, if the lift is greater than 1, then this signifies the dependency of itemsets.

3. *Conviction* is a different measurement that addresses the shortcomings of confidence and lift. However, conviction depends on the direction of the rule. For instance, conviction $(A \implies B)$ is not identical to conviction $(B \implies A)$. As defined by Erlandsson et al. (2016) the conviction rule is:

$$conviction\ (A, B \Longrightarrow C) = \frac{(1 - support\ (A,B)}{(1 - confidence\ (A, B \Longrightarrow C)}$$

4. *Laplace*, named after Pierre-Simon Laplace who developed this parameter, is an integral transformation (Azevedo and Jorge, 2007). Laplace is responsible for determining the confidence in terms of support, and becomes negative when the support of A declines. This parameter ranges within (0,1). Below is the laplace formula introduced by Azevedo and Jorge (2007):

$$Laplace\ (A \Rightarrow B) = \frac{support\ (A \cup B) + 1}{support\ (A) + 2}$$

When generating the association rules in Rapid Miner, the minimum confidence value has to be met. In this thesis, the author set the minimum confidence threshold at 60% in order to generate many meaningful and interesting rules, and sort them by confidence. As reported by Zhou and Yau (2007), using 60% as a minimum confidence generates a valid rule. This type of approach is used to search for interesting rules among items in the database (Azevedo and Jorge, 2007).

The association rules generated by Rapid Miner show the measured frequent itemset. Every frequent itemset that has been computed in FP-growth was generated using the association rules algorithm to find any interesting correlation between itemsets. This process also used the four parameters of confidence, conviction, laplace and lift. As stated previously, each of the four constraints has its own capability in terms of measuring the strength of the relationship between items. The results after generating the association rules are shown in Table 12.

Table 12 shows that five insulting words appear to have a relationship with each other. The five insulting words are *anjing* (dog), *babi* (pig), *bangsat* (bastard), *iblis* (devil), and *setan* (satan). *Bangsat* and *anjing* appeared to be the most frequent itemsets in confidence, laplace, lift and conviction.

Table 12 Frequent Patterns after Computed using Association rules in Rapid Miner

No.	Premises	Support	Confidence	Laplace	Lift	Conviction
1	bangsat babi anjing	0.052	0.969	0.998	2.972	21.906
2	iblis setan	0.055	0.985	0.999	7.804	61.979
3	bangsat	0.332	1	1	1	Infinity
4	anjing	0.326	1	1	1	Infinity
5	keparat	0.238	1	1	1	Infinity
6	setan	0.126	1	1	1	Infinity
7	babi	0.088	1	1	1	Infinity
8	goblok	0.067	1	1	1	Infinity
9	iblis	0.056	1	1	1	Infinity
10	idiot	0.051	1	1	1	Infinity
11	bajingan	0.050	1	1	1	Infinity
12	monyet	0.047	1	1	1	Infinity
13	bangsat anjing	0.269	1	1	1	Infinity
14	bangsat babi	0.054	1	1	1	Infinity
15	anjing babi	0.081	1	1	1	Infinity

Table 12 shows that both of these insulting words *bangsat* and *anjing* (*bangsat* $\Longrightarrow$ *anjing*) are within 0.27 of support_count, 1 of confidence, 1 of laplace, 1 of lift and

infinity of conviction, which indicates a strong relationship between the words, followed by *anjing* ⟹ *babi* with a support_count of 0.08, confidence 1, laplace 1, lift 1, and conviction is infinity. The third position is *bangsat* ⟹ *babi* with a 0.05 support_count, confidence 1, laplace 1, lift 1, and conviction is infinity. In the fourth position is *iblis* ⟹ *setan* with a support_count of 0.05, confidence 0.98, laplace 0.99, lift 7.8, and conviction 61.97. In the last position is *bangsat*, *babi* ⟹ *anjing* with support_count 0.05, confidence 0.97, laplace 0.99, lift 2.97, and conviction 21.9. Meanwhile, the independent terms that have a high number of support_count, confidence, laplace, lift, and conviction are *bangsat*, *anjing* and *keparat* (cursed).

According to the association rules concept, frequent itemsets indicate frequent patterns ($a \Rightarrow b$), which means that items are not independent but are related to others. Hence, the frequent patterns in Table 12 are *bangsat* ⟹ *anjing*, *anjing* ⟹ *babi*, *bangsat* ⟹ *babi*, *iblis* ⟹ *setan* and *bangsat*, *babi* ⟹ *anjing*.

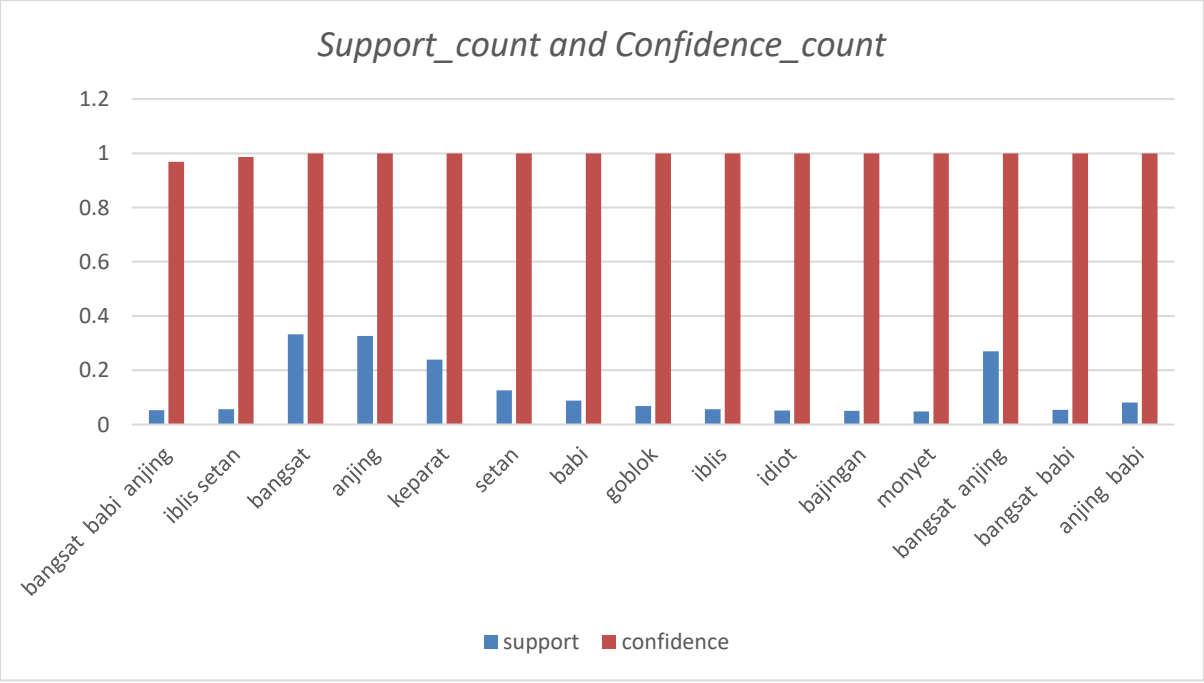


Figure 12 Result Calculation of Support_count and Confidence_count in Association rules

Figure 12 shows the association rules chart for *support_count* and *confidence_count*. Most itemsets have a high confidence value, although their *support_count* is lower. As illustrated in the figure above, $bangsat \Longrightarrow anjing$ have high *support_count* and confidence values. Therefore $bangsat \Longrightarrow anjing$ is considered as a popular frequent pattern based on the association rules result. However, in order to determine the extent to which the patterns are related to each other, this research has used cosine similarity to measure the relationship between itemsets that occurred as frequent patterns.

3.5.4 Mining the Indonesian Cyberbullying Patterns Using Cosine Similarity

The measurement of similarity between documents that contain sentences is one of the important tasks in this thesis. Since the research focuses on finding the relationship between documents, the measurement of similarity between them is an appropriate data mining technique for measuring the strength of similarity between documents (Achananuparp et al., 2008). The calculation of similarity between texts facilitates the assessment of text integrity (Lapata and Barzilay, 2005).

Every document that was recorded in the database contained the frequency of a particular word or phrase. Therefore, every document can be considered as an object that contains frequently occurring words (Han et al., 2012). The frequency with which words appear in a document can be represented as a vector. This vector contains integer values from 0 to positive value (depending on how many times the word appeared in each document). For example, the first document in the database recorded seven insulting words: *sarap* (crazy), *anjing* (dog), *goblok* (stupid), *monyet* (monkey), *setan* (satan), *babi* (pig), *bangsat* (bastard) that occurred once, while the terms *bajingan* (scoundrel), *gembel* (poor), *buta* (blind), *gila* (crazy), *idiot* (idiot), and

geblek (fool) did not appear at all. The words that did not occur frequently in the document of the database were weighted as 0, while 1,2,3,4…n represented the frequency of words occurring in the entire document.

One of the similarity measurements in data mining techniques is cosine similarity. In this research, the cosine similarity technique is used to measure the similarity between two documents or provide a ranking of the documents according to the word vectors (Han et al., 2012). This research used the cosine similarity technique to measure the similarity between the frequent patterns which contain some of the Indonesian insulting words and to determine the strength of the relationship between the Indonesian insulting words that occurred in the frequent patterns.

This research used Rapid Miner software to calculate the similarity between Indonesian insulting word patterns produced by analysing the Indonesian cyberbullying messages using association rules techniques. The Rapid Miner software provides data to a similarity operator as the technique to analyse document similarity with many different measures for similarity computation such as *Euclidean distance, nominal distance, dice similarity, Jaccard similarity, correlation similarity and cosine similarity.* The data provided to the similarity operator does not return the same similarity comparisons of data twice. For example, when a sample of dataset A is compared to a sample of dataset B to determine their similarity, the operator will not compute the similarity again because the result will be the same.

This research employed the cosine similarity technique using the formula for cosine similarity that was popularised by Han et al. (2012).

$$similarity = \cos\theta = \frac{a.b}{||a||||b||} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\ \sqrt{\sum_{i=1}^{n} b_i^2}}$$

Let a and b be the two vectors that will be measured for the purpose of comparison. $||a||$ and $||b||$ are the Euclidean norm of vectors $a=(a1,a2,a3,\ldots an)$ and $b=(b1,b2,b3,\ldots bn)$, outlined as $\sqrt{a12+a22+a32,\ldots+an2}$ and $\sqrt{b12+b22+b32,\ldots+bn2}$. Conceptually, it is defined as the length of the vector. This measurement calculates the cosine of the angle between vectors a and b. By definition, a cosine value of 0 implies that two vectors are at a 90-degree angle to each other, or orthogonal, including a no-match. In other words, when the cosine value is closer to 1, the angle is smaller and there is a better match between two vectors. An example of data vector is shown in Table 13, where the calculation has been done using the cosine similarity technique.

Table 13 Example Data Vector

no	text	anjing	babi	bajingan	bangsat	brengsek	budek	buta	gebelk	gila	goblok	iblis	idiot	jelek	monyet	sarap	setan
1	sarap anjing goblok monyet setan babi bangsat	1	1	0	1	0	0	0	0	0	1	0	0	0	1	1	1
2	gila anjing bangsat babi monyet sarap	1	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0

Suppose that a and b are the first two document frequency vectors in Table 11. The vector $a=(1,1,0,1,0,0,0,0,0,1,0,0,0,1,1,1)$ *and* $b=(1,1,0,1,0,0,0,0,0,1,0,0,0,0,1,1,0)$. Then, the computation of cosine similarity is as follows:

$a^t.b= (1x1+1x1+0x0+1x1+0x0+0x0+0x0+0x0+0x0+0x1+1x0+0x0+0x0+0x0+0x0+1x1+1x1+0x0)$
$= 5$

$$||a|| = \sqrt{1^2+1^2+0^2+1^2+0^2+0^2+0^2+0^2+0^2+0^2+1^2+0^2+0^2+0^2+0^2+1^2+1^2+1^2} = 2.64$$

$$||b|| = \sqrt{1^2+1^2+0^2+1^2+0^2+0^2+0^2+0^2+0^2+0^2+1^2+0^2+0^2+0^2+0^2+1^2+1^2+0^2} = 2.44$$

$cos(a,b) =0.77$

In this stage, the insulting word patterns result which came from the analysis in the association rules techniques was analysed using cosine similarity. Every pattern that occurs is recorded in a database. Next, the data was connected to the processing document operator in Rapid Miner which cleaned the data to produce the results for the document vector. In this cleaning process, the author of this research has duplicated the previous processing document operator. The document vector result from this stage is shown in Table 14. There are fifteen frequent patterns which can be either independent or dependent items. Nevertheless, there are ten items that occurred as frequent patterns: *bangsat* (bastard), *anjing* (dog), *keparat* (cursed), *setan* (satan), *babi* (pig), *gobl*ok (stupid), *iblis* (devil), *idiot* (idiot), *bajingan* (scoundrel), and *monyet* (monkey).

Table 14 Document Vector or Term-Frequency Vector in Cosine Similarity

No	Pattern	anjing	babi	bajingan	bangsat	Goblok	iblis	idiot	keparat	monyet	setan
1	bangsat babi anjing	1	1	0	1	0	0	0	0	0	0
2	iblis setan	0	0	0	0	0	1	0	0	0	1
3	bangsat	0	0	0	1	0	0	0	0	0	0
4	anjing	1	0	0	0	0	0	0	0	0	0
5	keparat	0	0	0	0	0	0	0	1	0	0
6	setan	0	0	0	0	0	0	0	0	0	1
7	babi	0	1	0	0	0	0	0	0	0	0
8	goblok	0	0	0	0	1	0	0	0	0	0
9	iblis	0	0	0	0	0	1	0	0	0	0
10	idiot	0	0	0	0	0	0	1	0	0	0
11	bajingan	0	0	1	0	0	0	0	0	0	0
12	monyet	0	0	0	0	0	0	0	0	1	0
13	bangsat anjing	1	0	0	1	0	0	0	0	0	0
14	bangsat babi	0	1	0	1	0	0	0	0	0	0
15	anjing babi	1	1	0	0	0	0	0	0	0	0

Table 14 shows the frequency of terms in the document vector or *Term-Frequency Vector*. 0 indicates that the term is not present in the document, while $1,2,3\ldots n$ indicates the frequency of terms that occur in the document. When measuring the similarity between documents using the cosine technique, every document must be compared to another. On the other hand, in terms of similarity, the comparison does not return the same similarity data twice, since both results will be identical. For example, the sample data for number 1 will be compared to sample data for number 2, but it will not compute the similarity between both data samples.

In terms of understanding the calculation process using the cosine similarity technique to find the correlation between documents in this stage, the sample calculation for data number 1 and 14 in Table 14 is shown as:

Suppose letter a represents data number 1, and b represents number 14.

$$a^t.b = (1x0+1x1+0x0+1x1+0x0+0x0+0x0+0x0+0x0+0x0) = 2$$

$$||a|| = \sqrt{1^2 + 1^2 + 0^2} + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1.73$$

$$||b|| = \sqrt{0^2 + 1^2 + 0^2} + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1.41$$

$$cos(a,b) = 0.823$$

The result of manually calculating the similarity between the data in row 1 and 14 is 0.823. However, the calculation process done through the Rapid Miner software produced a similarity of 0.866. Nevertheless, the manual results closely approximate the computational result produced by Rapid Miner software as shown in Table 15. The reason for this small discrepancy in the results is caused by the decimal point.

Table 15 shows the outcome of data similarity calculation using the cosine similarity technique. From the results depicted in Table 15, the large number of results within 0.866 are for the comparisons between data in rows 1 and 13, rows 1 and 14,

rows 1 and 15. This implies that *bangsat* (bastard), *anjing* (dog), and *babi* (pig) have a strong relationship with one another.

Table 15 Similarity Data

Doc 1	Doc 2	Similarity	Doc 1	Doc 2	Similarity	Doc 1	Doc 2	Similarity
1	2	0.288	3	12	0.5	7	9	0.5
1	3	0.707	3	13	0.816	7	10	0.5
1	4	0.707	3	14	0.816	7	11	0.5
1	5	0.353	3	15	0.408	7	12	0.5
1	6	0.353	4	5	0.5	7	13	0.408
1	7	0.707	4	6	0.5	7	14	0.816
1	8	0.353	4	7	0.5	7	15	0.816
1	9	0.353	4	8	0.5	8	9	0.5
1	10	0.353	4	9	0.5	8	10	0.5
1	11	0.353	4	10	0.5	8	11	0.5
1	12	0.353	4	11	0.5	8	12	0.5
1	13	0.866	4	12	0.5	8	13	0.408
1	14	0.866	4	13	0.816	8	14	0.408
1	15	0.866	4	14	0.408	8	15	0.408
2	3	0.408	4	15	0.816	9	10	0.5
2	4	0.408	5	6	0.5	9	11	0.5
2	5	0.408	5	7	0.5	9	12	0.5
2	6	0.816	5	8	0.5	9	13	0.408
2	7	0.408	5	9	0.5	9	14	0.408
2	8	0.408	5	10	0.5	9	15	0.408
2	9	0.816	5	11	0.5	10	11	0.5
2	10	0.408	5	12	0.5	10	12	0.5
2	11	0.408	5	13	0.408	10	13	0.408
2	12	0.408	5	14	0.408	10	14	0.408
2	13	0.333	5	15	0.408	10	15	0.408
2	14	0.333	6	7	0.5	11	12	0.5
2	15	0.333	6	8	0.5	11	13	0.408
3	4	0.5	6	9	0.5	11	14	0.408
3	5	0.5	6	10	0.5	11	15	0.408
3	6	0.5	6	11	0.5	12	13	0.408
3	7	0.5	6	12	0.5	12	14	0.408
3	8	0.5	6	13	0.408	12	15	0.408
3	9	0.5	6	14	0.408	13	14	0.666
3	10	0.5	6	15	0.408	13	15	0.666
3	11	0.5	7	8	0.5	14	15	0.666

The similarity data can also be seen from the calculated results in the comparisons of data in rows 13, 14, and 15 that displays the value of 0.666 in Table 15. Therefore, the combination of the terms *bangsat*, *anjing*, and *babi* are the most common insulting words used when sending cyberbullying messages. Another common pattern of insulting words in cyberbullying messages is *iblis* (devil) and *setan* (satan).

3.6 Labelling the Data Results

This section describes the process of labelling data results following the generation of association rules techniques. The process involved the application of *Kamus Besar Bahasa Indonesia (Great Dictionary of the Indonesian Language)* (Setiawan, 2016), and *Kamus online (Indonesian online dictionary)* (Indahnesia, 2016). The purpose of using two sources - *Kamus Besar Bahasa Indonesia and Kamus online* - was to confirm and to reduce the bias of the meaning of the words.

Kamus Besar Bahasa Indonesia which is available online is the appropriate and well-translated Indonesian dictionary created to facilitate the search, use and reading of the word definitions; i.e., entries or sub-entries. This dictionary was published by the language centre of the Ministry of Indonesia Education and Culture. The database *Kamus Besar Bahasa Indonesia* refers to *Kamus Besar Bahasa Indonesia* edition III, so that the content of the meaning of words was developed and copyrighted from the language centre, and the Ministry of Indonesia Education & Culture (Setiawan, 2016). Meanwhile, the *Kamus online* is one of the well-translated Indonesian dictionaries that is free online and available for translation from English to Indonesian or Indonesian to English.

From generating the results of association rules techniques, *bangsat* (bastard), *anjing* (dog), *babi* (pig), *iblis* (devil) and *setan* (satan) were the five popular insulting words that occurred. In a later stage, all five insulting words were confirmed by using the *Kamus Besar Bahasa Indonesia* and the *Kamus online*. The result was that the five insulting words comprising of *bangsat*, *anjing*, *babi*, *iblis* and *setan* were the Indonesian swear words which related to animals and Satan.

As these five Indonesian insulting words are also referred to as swear words, it can be concluded that these words are cyberbullying words which often occurred in cyberbullying messages. Thus, the five insulting words were labelled as the Indonesian cyberbullying words. The labels of the Indonesian cyberbullying words were used for further analysis in this research.

3.7 Conclusion

This chapter provided an analysis of Indonesian cyberbullying messages using association rules techniques: FP-growth and cosine similarity in sequence to extract the patterns of Indonesian insulting words that can help to identify cyberbullying messages from Twitter. The discovery of data patterns in a text document can be done in several ways. In this research, in order to facilitate the search process, association rules techniques were used.

Association rules techniques of FP-growth and cosine similarity techniques were applicable for the analysis of Indonesian messages from data corpus obtained from Twitter. The analysis showed that the most common Indonesian insulting word patterns were *bangsat*, *anjing*, and *babi*, *iblis* and *setan*. Although this research used association rules techniques of FP-growth and cosine similarity, several components of constraints including confidence, lift, conviction, and laplace emerged in the analysis

process. The association rules techniques were implemented using the Rapid Miner software which enables FP-growth and cosine similarity to be applied.

This research contributes to the development of an analysis model and the implementation of data mining techniques for the analysis of cyberbullying messages through the implementation of association rules techniques in sequence which can be used in future studies. Furthermore, these results provided important data that can be used as a resource in further research in respect to the combination of data mining techniques with the analysis model for social issues.

In the next chapter, the insulting word patterns that were explored in this chapter will be used as the basis for determining whether or not messages are cyberbullying. The chapter also discusses the application of several types of classification techniques as a means of determining whether or not messages can be considered as cyberbullying.

Chapter 4
Identifying Indonesian Cyberbullying Messages

4.1 Introduction

The data mining techniques used to predict class label data are known as classification techniques. The approach used in classification techniques involved assigning an object to a certain class of labelled data based on the previous data of other objects. Classification techniques could indicate the similarity of objects that are definitely members of a given class (Han et al., 2012).

The prediction of class label data can be illustrated by a simple example. Let us assume that the author of this research is collecting data of messages from social networks. Every message is recorded in a database and comprises words or terms that are used in some sentences. Every record in a database represents one user and contains some terms posted by the user on social networks, for instance, “you are such a big boy with many talents”. Based on previous data that have been labelled according to a class, when the message contains terms such as *you, big, boy,* and *talent*, then the message will be labelled as a positive message and allocated to the appropriate class. However, if the message contains terms such as, *you, big, boy,* and *incompetent,* displayed in full form as “You are such a big boy but you are so incompetent”, then the message will be labelled as a negative message and placed in the negative class.

This chapter discusses how the classification techniques function in predicting interesting data classes in a database. The data referred to in this chapter are the

messages from Twitter. Therefore, this work has implemented the naïve Bayes, decision tree, and neural network as classification techniques to predict the unlabelled data, specifically, whether or not the messages from social networks are instances of cyberbullying.

The aim in this chapter is to find hidden interesting class labels of messages based on the Indonesian social network of Twitter. These classifications can be discovered by analysing the messages that may be considered to be either instances of cyberbullying or non-cyberbullying. Therefore, data classification techniques are considered as really useful methods in facilitating proper data categorisation from the content of messages.

This chapter consists of several sections. The first section explains the processing of the data to be analysed using classification techniques. The basic concept of the classification techniques and the description of the illustration and function of naïve Bayes, decision tree, and neural network techniques are explained in the second section. The third section describes the process of mining the Indonesian cyberbullying messages by creating and implementing a design analysis model using classification techniques, particularly naïve Bayes, decision tree, and neural network.

4.2 Data Preparation in Classification

This section explains the processing of data collected from Twitter. The data processing in this stage was similar to the previous stage. The amount of data used for analysis consisted of 152,843 records taken from the messages sent via Twitter for data testing. The data in the form of messages mostly contained Indonesian insulting words that tended to be cyberbullying messages. Therefore, processing the

data from Twitter had to be accomplished prior to the analysis of the Indonesian cyberbullying messages. Chapter 4 will describe the prediction stage of these messages.

In analysing the Indonesian cyberbullying messages from Twitter using classification techniques, a local database of Indonesian insulting words was required. The data that was crawled from Twitter was recorded in a local database created by means of Rapid Miner. Using the menu option of importing data from CSV format into a database in Rapid Miner, all data was transferred to the database. The purpose of recording data in a database was to facilitate data retrieval for the data processing. The process of data retrieval was the same procedure previously completed in Chapter 3 and explained in section 3.4.

Once the data had been retrieved from the repository through the retrieve operator, the next stage involved cleaning data using the processing document operator in Rapid Miner. Just like the previous process used to find the relationship between insulting words, the purpose of document processing was to clean abbreviations and punctuation from the messages, so that the data contained only the various words that made up a sentence.

In the next stage of data preparation, a data cleaning process was necessary. Cleaning aims to obtain a data matrix so that the data being generated is a specifically-built word vectors. This process was similar to the previous stage of cleaning data described in Chapter 3 in section 3.3. Using the process document operator in Rapid Miner, the object of textual data as input was generated to acquire term vectors. The purpose of term vectors was to map the terms that appeared in every recorded document for the further process of classification data prediction.

In the process of document operator, similar to the previous stage in association rules mining, cleaning processes involve tokenize, transform case, stop words, and stem dictionary. These cleaning processes are used to eliminate any abbreviations, punctuation marks and emoticons.

The result after generating the document processing operator are presented as vectors $d_i = (w_{1i}, w_{2i}, w_{3i}, \dots, w_{ni})$. Each dimension, perse, w_{1i}, w_{2i} is adjusted to an independent term. For example, if a term occurs in a document, the term has a non-zero value in the vector. On the other hand, if a term does not occur in the document, then the value of the term is zero in the vector. A term is defined as an individual word and keyword. If a word is a term, the number of words that appear in the corpus is indicated by size of the vector.

In the document operator process, the appearance of binary terms is considered as the weight. The number 1 indicates the presence of a term in a document, while 0 indicates the absence of a term in a document. The occurrence of terms will produce binary numbers in the documents, particularly when textual documents are transformed into vectors. Following the generation process of the document operator, the table of word vectors will be displayed as shown in Table 16.

Table 16 shows that each word is displayed according to its appearance in every message. The number 1 indicates the presence of a word in a message; a zero (0) indicates that the word is absent in a message. For instance, the word *anjing* (dog) located in row number 1 is represented by the number 1, and the word *bajingan* (scoundrel) is symbolised by 0, which means that *bajingan* is not in the first record. This process assisted the author of this research with the next process, which was predicting data using the classification techniques.

Table 16 Example of Results after Generating Processing Document Operators in Words Vector

Row No	Text	anjing	babi	bajingan	bangsat	brengsek	buta	gembel	gila
1	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	1	1	0	1	0	0	0	0
2	gila anjing bangsat (mental disorder, dog, bastard)	1	0	0	1	0	0	0	1
3	sarap anjing goblok monyet setan babi bangsat (crazy, dog, stupid, monkey, satan, devil, bastard)	1	1	0	1	0	0	0	0
4	anjing goblok monyet setan babi bangsat (dog, stupid, monkey, satan, pig, bastard)	1	1	0	1	0	0	0	0
5	bangsat tolol anjing setan (bastard, stupid, dog, satan)	1	0	0	1	0	0	0	0
6	anjing bangsat monyet (dog, bastard, monkey)	1	0	0	1	0	0	0	0
7	bangsat anjing (bastard, dog)	1	0	0	1	0	0	0	0
8	anjing bangsat (dog, bastard)	1	0	0	1	0	0	0	0
9	anjing bangsat (dog, bastard)	1	0	0	1	0	0	0	0
10	bangsat anjing (bastard, dog)	1	0	0	1	0	0	0	0

4.3 Data Training

One of the basic concepts underlying the data analysis process using classification techniques introduced by Han et al. (2012) is learning data in which a classification algorithm creates a classifier by analysing or reviewing a training set

comprised of tuples in a database, along with their correlated class labels. Say, a tuple *Z*, is expressed by a n-dimensional attribute vector, $Z= (z_1, z_2, z_3, \ldots z_n)$ that illustrates the n measurements, built on the tuple from n database attributes, in respect to $A_1, A_2, A_3, \ldots, A_n$. Each tuple that can be referred as X is expected to reside in a predefined class determined by a different database attribute known as the class label attribute. The class label attribute is in discrete values and non-sequential; instead, the attributes are nominal where every value acts as a category or class. In other words, a single tuple of a training set is referred to as training tuples randomly sampled from the database through an analysis. In respect of classification, data tuples can be pointed out as samples, examples, instances, data points, or objects.

Dinakar et al. (2011) downloaded 50,000 comments from social networks to be analysed in order to identify textual cyberbullying. They clustered the comments into three groups of gender, race and culture, and intelligence, which were believed to be personal and sensitive topics, thus establishing candidates for cyberbullying. They assigned 1,500 comments into every group to be annotated manually to ensure the accuracy of assigning labels to them. They concluded that 627 comments were positive for gender, 841 comments for race & culture, and 809 for intelligence. Additionally, Dinakar et al. (2011) divided the data set for each group so that 50% was training data, and every data set was subjected to pre-processing data in order to clean the data.

Kasture (2015) extracted 1313 unique tweets to be analysed for cyberbullying messages. With the assistance of psychological studies in relation to an individual's behaviour and the implementation of dialects referring to verbal aggressiveness, using a manual process, 376 tweets were labelled as cyberbullying tweets. Kasture (2015) used the resulting data as a training data set in order to determine the data used for

testing. This data set was later input into the train machine learning classifiers in Weka to construct a predictive model for cyberbullying detection. Randomly, data was divided into 66% for training the predictive model, and 34% for the testing procedure (Kasture, 2015).

Referring to research by Dinakar et al. (2011) and Kasture (2015), the author of this research collected 80,006 messages from Twitter for the training dataset. This training data was examined repeatedly to confirm that there was no overlap between the two sets of files. Each piece of recorded data was selected to be different with each other. The purpose was to acquire an enriching data training set that could be a reference data model to predict data testing. However, the training data set contained both cyberbullying and non-cyberbullying messages.

This thesis has employed clustering techniques to construct a data model for the training data set. Clustering was used to group the data into two categories: cyberbullying and non-cyberbullying. In this case, the natural messages were included in the non-cyberbullying class. The training data set was taken from the random data that had not yet been classified. Thus, the messages were categorised using the clustering technique to identify the characteristics and the features of the document (Jain, 2010; Shiells and Pham, 2010). By means of the clustering technique, the messages in the database were distributed into groups of data according to the similarity of objects. These distributed groups comprise messages with similar details, meaning that an individual group can serve as a representation for the similar characteristics and features of the messages (Rosa et al., 2011; Kim et al., 2012).

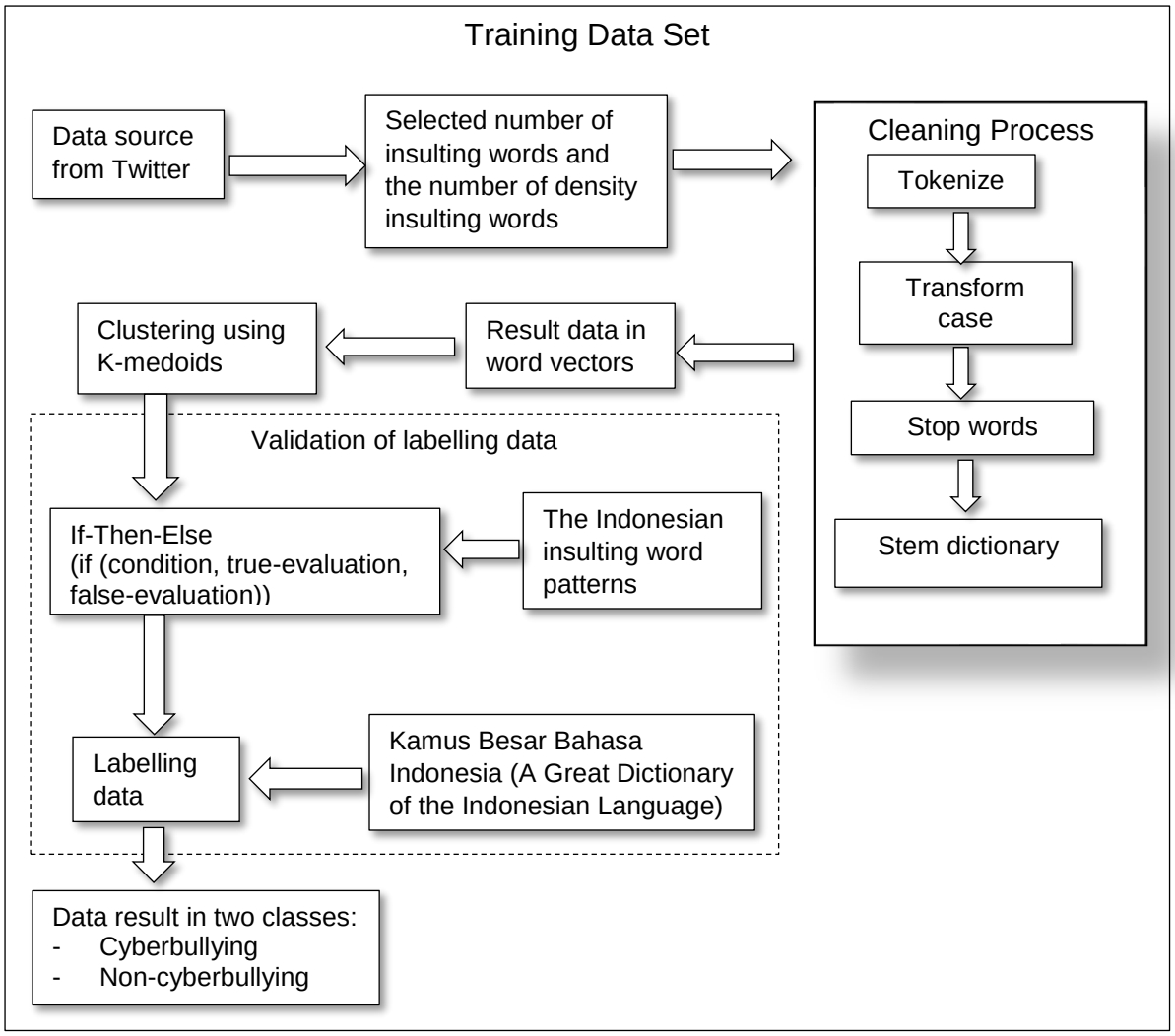


Figure 13 Development Process of Analysis Model Training Set

Figure 13 depicts the process of labelling data classes in a training data set. Within this labelling process, there are several steps involved that initially include retrieving documents from the repository in Rapid Miner, then selecting the number and density of insulting words. The results were moved to the cleaning stage in processing document operator using filters of tokenise, transform case, stop words and Indonesian insulting stemming. After this, the cleaned data were converted into word vectors. The word vectors data were clustered through the use of *k*-medoids. The prior stage produced outcomes of two clusters that were then labelled manually by the author. From this labelling process, the author implemented If-Then-Else

mathematical expression to determine data into cyberbullying class or non-cyberbullying class based on the insulting term patterns. The insulting term patterns were discovered from the previous process of data analysis using association rules, which was explained in Chapter 3. Also, the data labelling process involved the application of *Kamus Besar Bahasa Indonesia (Great Dictionary of the Indonesian Language)* (Setiawan, 2016) and *Kamus online (Indonesian online dictionary)* (Indahnesia, 2016) to validate the exact meanings of the insulting terms. The final results created two data labels of cyberbullying and non-cyberbullying.

4.3.1 K-Medoids Algorithm

The clustering approach is a type of unsupervised learning due to the unlabelled nature of a data class. Hence, clustering is a type of learning through observation, instead of learning through example. Dinakar et al. (2011) demonstrated how clustering techniques can be used to separate the training dataset. They employed the clustering technique to divide their data training into four clusters. This research adopted the same clustering technique to divide the training dataset.

The implementation of *k*-medoids purposes to enhance the quality of the clustering outcomes by minimising the absolute error while grouping objects into clusters in large data sets. Also, *k*-medoids functions independently from data order, and thus produces high computational efficiency (Hämäläinen, 2006). As compared to other clustering techniques, the *k*-medoids approach is more robust and effective in terms of noise and outliers, since *k*-medoids is not too heavily affected by the outliers or other extreme values (Han et al., 2012).

The author used the *k*-medoids algorithm to divide 80,006 recorded data into two classes: cyberbullying and non-cyberbullying. *K*-medoids is a clustering technique

that clusters the data set of *n* objects into *k* clusters known a priori (Han et al., 2012). *K*-medoids that are relevant to *k-means* clustering is a part of the segmenting approach in clustering techniques intended to reduce the distance between points that are meant to be tagged in a cluster, with an assigned point as the centre of that particular cluster. In contrast to the *k*-means clustering, *k*-medoids selects data points as the centre and functions under an arbitrary metrics of distances between data points, instead of l_2. Proposed in 1987, this method is capable of functioning with other distances of data points (Kaufman and Rousseeuw, 1987). Thus, this research applied the *k*-medoids algorithm in order to separate data by decreasing the number of objects in a cluster within an average dissimilarity to every other object in the cluster.

The formula for *k*-medoids introduced by Han et al. (2012) is:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} dist(p, o_i)$$

The total number of complete errors for the entire objects *p* is in the data set, while oi serves as a representation of object C_i. In other words, this is the foundation of the *k*-medoids technique where *n* groups of objects are placed into *k* clusters through the process of reducing absolute error. Basically, this concept considers decreasing the total number of dissimilarities between each object *p* along with its parallel representative objects, which can be known as the absolute error (Han et al., 2012).

Specifically, $o_1,\ldots,o_k$ are the current group of representative objects, or medoids. For determining whether a non-representative object, denoted by o_{random}, is a suitable substitute for the current medoids o_j. $(1 \leq j \leq k)$, a distance calculation process should be applied from each object *p* to the nearest object in the set $(o_1,\ldots,o_{j-1},o_{random},o_{j+1},\ldots,o_k)$, including the implementation of that distance to renew the function of the cost. Let us assume that object *p* has been currently appointed to a

cluster with a representation of medoids o_j. In other words, object p is required to be retransferred to either o_{random} or another cluster under a symbol of o_i $(i \neq j)$, whichever is the nearest. In this case, object p is the nearest to o_i and hence, is shifted to o_i. Nonetheless, the nearest object to o_{random} is p, and therefore is reallocated to o_{random}. Furthermore, object o continues to be placed into the cluster with a representation of o_i for as much as object o is neighbouring to o_i than to o_{random}. If not, o is relocated to o_{random}.

In order to divide the training dataset using the *k*-medoids algorithm, this research used Rapid Miner to obtain the data for the clusters. The process of running the *k*-medoids algorithm in Rapid Miner involves several procedures as follows:

1) The retrieve data operator is used to access data from repositories. The explanation of this process has been mentioned previously in Chapter 3 in section 3.4.
2) The processing document operator is selected to generate word vectors from a text object. This operator uses only an individual text object as the input for the process of generating a term vector. In this research, the processing of documents at this stage is the same as that in previous works involving association rules techniques. To remove abbreviations, punctuation marks and emoticons from data, when processing the document, several parameters have been applied such as tokenize, transform case, stop words, and stem dictionary. The words derived from this stage are displayed on a data matrix of word vectors, similar to the previous works in association rules techniques.
3) The *k*-medoids clustering operator is used to perform clustering using the *k*-medoids algorithm. *K*-medoids clustering is an absolute clustering algorithm

(Akthar and Hahne, 2012). For instance, every object is allocated to a particular cluster. Furthermore, all of the objects in a cluster are similar to one another. Thus, the similarity between objects is indicated by the distance between the words. The *k*-medoids clustering operator offers several parameters that can be employed to establish the measurement of data similarity, such as measuring the menu type applied in choosing the measurement types which will be further employed in calculating the distance between points. The available preferences are: *mixed measures, nominal measures, numerical measures and Bregman divergences.* In this case, numerical measures have been chosen, since this parameter offers a cosine similarity algorithm-based parameter. Hence, cosine similarity was implemented to measure the similarity between data in this thesis.

4) The three operators -- retrieve data, processing document and clustering *k*-medoids -- are connected in systematic order.
5) All processes are implemented to obtain the overall result of clustering data using *k*-medoids techniques.
6) The *k*-medoids operator was applied on this data set with $k = 2$ signifying that the number of clusters should be detected; the type of measure used was the numerical measure; and the numerical measure was cosine similarity. The reason for selecting $k = 2$ was to divide the data set into two clusters of data labels: cyberbullying and non-cyberbullying. As discussed above, when the data are still in the form of normal messages, they are labelled as non-cyberbullying, therefore the required number of clusters in this case was two.

The process of modelling analysis using *k*-medoids is shown in Figure 14.

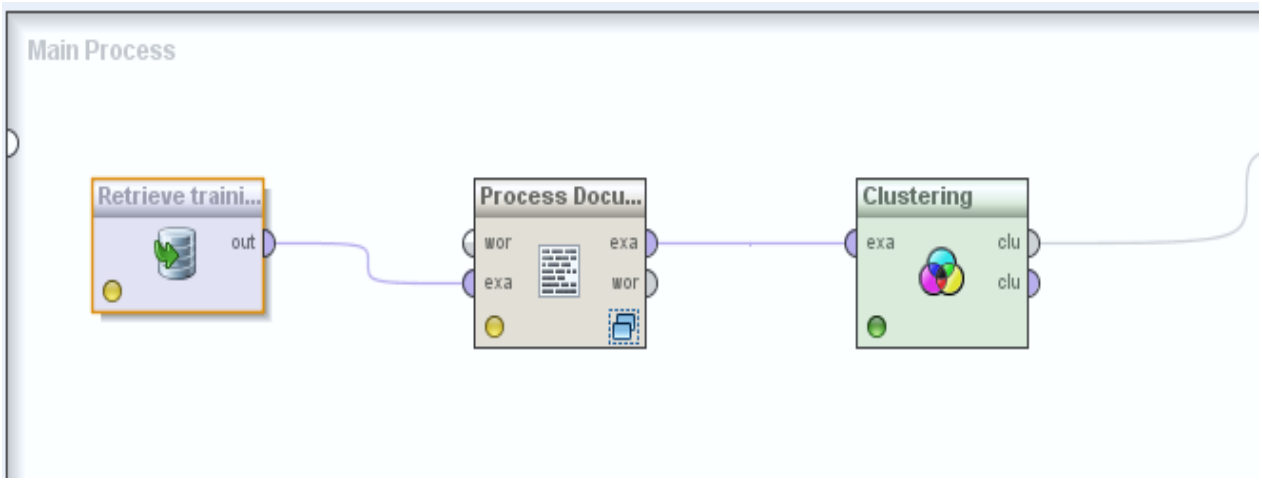


Figure 14 The Main Process of Analysing Data using K-Medoids Algorithm in Rapid Miner

The result also displays the performance of the *k*-medoids algorithm which divided the data set into two clusters as a cluster model (Akthar and Hahne, 2012). The members of each cluster can be seen in the folder view in folder format. The results after generating the *k*-medoids are shown in Figure 15.

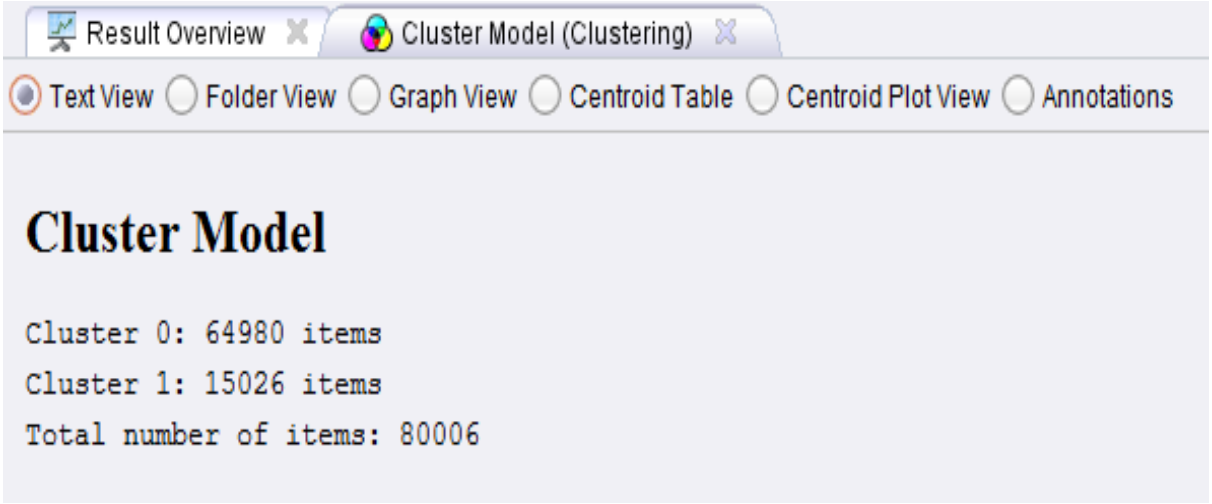


Figure 15 Results after Generated K-Medoids Algorithm as Cluster Model

Figure 15 presents a cluster model obtained by using the *k*-medoids algorithm in Rapid Miner. As observed, the data set was divided into two clusters as expected from setting $k = 2$ in the earlier stage. Based on the result, the 80,006 items in the data

set were divided into cluster 0 and cluster 1. Cluster 0 is predominant with 64,980 items in comparison to cluster 1 with 15,026 items. Cluster 0 and cluster 1 are two candidates of groups of cyberbullying and non-cyberbullying messages. At this stage, cluster 0 and cluster 1 had not yet been labelled as cyberbullying or non-cyberbullying. The process of labelling data will be explained in the next section. However, the data result from each cluster was transformed into a spreadsheet. The purpose of this transformation was to group the items in the data set into cluster 0 and cluster 1.

To evaluate the performance of centroid in clustering techniques, especially *k*-medoids, this research used the cluster distance performance operator in Rapid Miner. The operator used here shows a record of performance criteria values in accordance with the cluster centroids. The centroid-based clustering operators, such as *k*-medoids, develop a centroid cluster model and a clustered group. Moreover, the centroid cluster model carries information about the performance of the clustering approach. The operator of centroid cluster performance accepts this centroid cluster model and clustered group as the input and judges the model's performance according to the cluster centroids. The two performance measures are: *Average within cluster distance* and *Davies-Bouldin index.*

- *Average within cluster distance*: The average within cluster distance is obtained by calculating the average of distances between the centroid and all other items in the cluster.
- *Davies-Bouldin index*: This algorithm generates clusters of low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity), which produces a low Davies-Bouldin Index. In other words, the Davies-Bouldin index is the appropriate clustering algorithm that constructs an assembly of clusters.

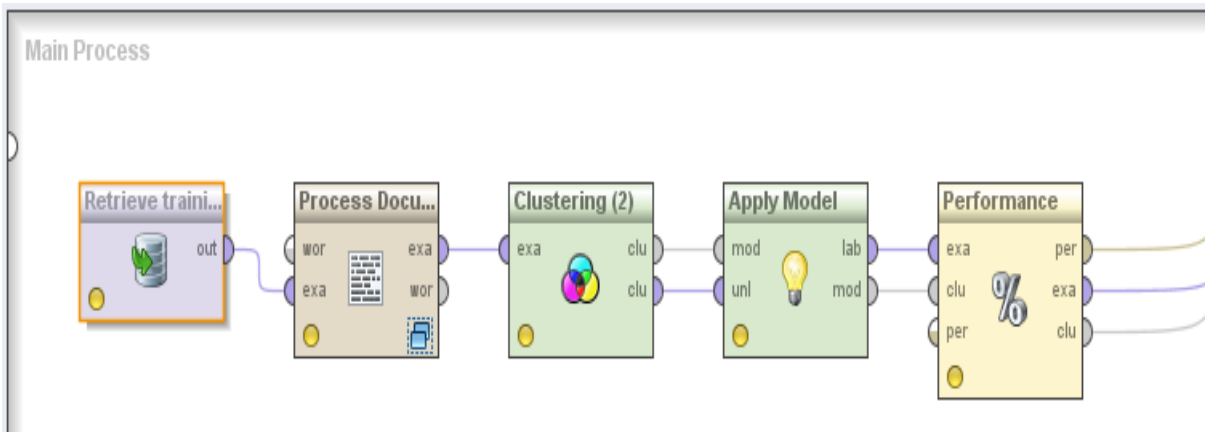


Figure 16 the Main Process of Measurement of the Performance of K-Medoids Algorithm in Rapid Miner

Figure 16 demonstrates that the process of analysing data sets using *k*-medoids is a sequence from the retrieving data process to the measurement of the performance of *k*-medoids. This means that the analysis of the data set was conducted in this order:

1. Retrieving data from the repository
2. Cleaning data in processing document operator
3. Applying the *k*-medoids algorithm as machine learning for data
4. Evaluating the performance of *k*-medoids which is named as the centroid cluster model (Akthar and Hahne, 2012).

The products by *k*-medoids operator, in this case the cluster model and the clustered set, serve as an input to the Cluster Distance Performance operator. In return, this operator evaluates this performance model and produces a performance vector with a list of performance criteria values. The overall results from the performance vector are displayed in the results workspace. The measurement result of the evaluation performance *k*-medoids can be seen in Figure 17.

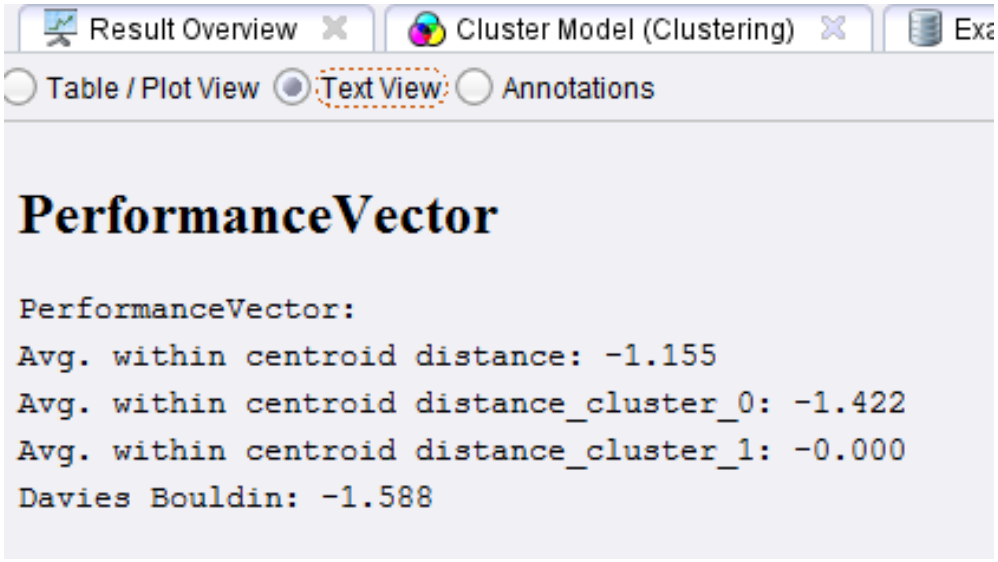


Figure 17 Result of Evaluation Performance K-medoids

As observed in Figure 17, the average of the centroid distance for both cluster 0 and cluster 1 is -1.155. The negative value is returned by Rapid Miner and indicates that the best density is the smallest absolute value, so negating this means the best density would be the maximum enabling it to be used as a stopping criterion during optimisation (Akthar and Hahne, 2012). Every performance of clusters is measured by totalling all cluster performances against the weight of the number of points in each and dividing the result by the number of examples. Thus, the result of averaging the distance between the centroid and all examples of two clusters shown in Figure 17 is the maximum performance in *k*-medoids. The values illustrated in Figure 17 confirm the success of *k*-medoids in dividing the data set into two clusters because of the high performance shown by calculating the average distance between centroids in a cluster.

4.3.2 Labelling the Training Data

This section explains the process of labelling the data set for data training after the division stage, using the *K*-mediods clustering technique. The process of labelling the data set for data training involved data labels from the results of association rules in Chapter 3. The patterns of bullying words obtained from association rules techniques have been used to identify the bullying words that occurred in the data set.

Using data patterns obtained from the association rules techniques can determine the class labels for data training. The outcomes of clustering as explained above was the data set that was divided into two clusters where cluster 0 is predominant with 64,980 items and cluster 1 with 15,026 items. All items in cluster 0 and cluster 1 were presented in the spreadsheet. The next step was to retrieve data stored in the CSV format of the spreadsheet using the menu read excel operator in Rapid Miner. All data were directly imported to the repository, where the spreadsheets in the workbook were used as the data table. It is strongly recommended that the table have a format, so that every row and column represent an attribute. Moreover, the data table can be situated anywhere on the sheet that includes arbitrary formatting instructions, unoccupied rows and columns. In addition, data values that are missing in table should be indicated by empty cells.

The generate attribute operator is used to produce new attributes from the input data set attributes and arbitrary constants through the application of mathematical expressions. The names of the input data set attributes can be applied as variables in relation to mathematical expressions to produce the new attributes. Furthermore, in the course of managing this operator, the mathematical expressions will evaluate each example, and then the attribute values of the examples will be placed in the variables. Therefore, apart from creating new columns for new attributes, this operator also

populates the columns with the corresponding values of the attributes. In other words, if a variable is undefined in one expression, the whole expression will also be undefined, and a question mark, '?' will show its location. In this case, the mathematical expression that has been used is *If-Then-Else.* The following is the expression of the *If-Then-Else* statement:

If condition [Then]
[statements]
[Else If else if condition [Then]
[else if statements]]
[Else
[else statements]]
End If

In the first argument, the condition is detailed. When the condition is determined as true, the result from the second argument will be delivered; alternatively, the third argument is then conveyed. The regular expressions that have been written in this stage are:

If((anjing+bangsat+babi+bajingan+goblok+iblis+idiot+monyet+setan)>1,"cyberbullying",

if((anjing+bangsat+babi+bajingan+goblok+iblis+idiot+monyet+setan)==1,"noncyberbullying","cyberbullying"))

In this case, the whole-word patterns produced generated by association rules techniques were used in *If-Then-Else* regular expressions in order to identify the bullying words in each item in the data training set. In the next stage, the results were divided into two categories labelled: cyberbullying and non-cyberbullying. The results were confirmed by the outcomes of the two clusters obtained by clustering techniques. All items in cluster 0 were considered as cyberbullying, while cluster 1 comprised non-cyberbullying items. Hence, cluster 0 was labelled as the cyberbullying class and

cluster 1 as the non-cyberbullying class. To confirm the validity of the data class labelling, Indonesian dictionaries were used for the next stage.

Another way to label the data training set is to use Kamus Besar Bahasa Indonesia (Comprehensive Dictionary of the Indonesian Language) (Setiawan, 2016) and Kamus online (the online Indonesian Dictionary) (Indahnesia, 2016). These were used to translate the words into English. The main reason for using two dictionaries was to confirm the exact meaning of each term and to reduce the ambiguity in the meaning of the words. The online Indonesian dictionary is a free online dictionary that translates from English to Indonesian as well as from Indonesian to English.

Kamus Besar Bahasa Indonesia was the best Indonesian dictionary to facilitate the search, use and reading of word definitions i.e., entries or sub-entries. This dictionary was published officially by the language centre of the Ministry of Indonesia Education and Culture. The database Kamus Besar Bahasa Indonesia refers to Kamus Besar Bahasa Indonesia edition III, so that the content of the meaning of words was developed and copyrighted by the language centre and the Ministry of Indonesia Education & Culture (Setiawan, 2016).

The process of labelling, using both dictionaries was as follows:

- Every word including the insulting words that appeared in every message in the training data set was typed on the search query menu and then the search button was clicked.
- Every translated word from both dictionaries was confirmed in the original sentence, so that the meaning of the sentence corresponded correctly to the original meaning.

- If the accurate meaning of all sentences in every message suggested bullying, then the message could be labelled as cyberbullying. If not, then the message could be labelled as non-cyberbullying.

The result of the labelling of the data training set using both dictionaries was nearly the same as the labelled results using mathematical regular expression and distributing the data into two clusters using *K*-mediods, with a discrepancy of three records. The three discrepancies were confirmed again to determine whether their label should be either cyberbullying or non-cyberbullying. Finally, the final data labels were confirmed.

The method of using both dictionaries potentially had the limitation of being subjective. However, the use of a combination of clustering techniques and mathematical regular expression *If* and *Then* were an important means of determining the labels, thereby reducing any bias in labelling the data set items correctly, apart from being an accurate and speedy process.

4.4 Machine Learning

The type of artificial intelligence that provides a computer with the ability to learn automatically without human intervention, recognising complex patterns and creating intelligent decisions from the data itself is called *machine learning* (Witten and Frank, 2011; Han et al., 2012). Machine learning is a part of data mining techniques which examines data in order to detect patterns and adjusts the automatic decision based on the observation of the data itself. Machine learning is an automatic method used to observe a data set by employing a learning algorithm in order to learn automatically without human intervention.

The machine learning that is related to data mining techniques consists of four types of learning processes introduced by Han et al. (2012). These are as follows:

- Supervised learning. The process of learning comes from supervising examples of data that have been labelled in the training data set. This process is usually known as a *classification model* since it uses training examples to supervise the process of deciding to which class a data group belongs.
- Unsupervised learning. The process of learning comes from the iteration of the algorithm when grouping the data itself. In this case, the data input items have not been labelled according to their classes. Typically, this method, called the *clustering model*, is applied to discover classes in data.
- Semi-supervised learning. This learning process is another type of machine learning technique which adopts both labelled and unlabelled examples during the learning process of a model. Using one approach, labelled examples are employed to learn class models, while the unlabelled examples are applied in refining the boundaries between classes more clearly.
- Active learning. This type of machine learning approach allows users to actively participate in the learning process. Active learning requires users to label an example that can originate from a group of unlabelled examples or, on the other hand, those incorporated by the learning program.

The development of a messages analysis model from social networks using the supervised learning method is one of the main tasks in this research. As the concept of supervised learning entails learning from supervising the labelling of items from the training data set, this research adopted several supervised learning techniques to predict and supervise the data features to be labelled according to their respective classes.

To identify cyberbullying messages, three techniques of classification which are naïve Bayes, decision tree, and neural network were chosen in this research. The purpose was to obtain high accuracy in data prediction. These three techniques were implemented to find the class of data labels. The basic concept of these three techniques is based on supervised learning techniques that often use a data training set as a data model to supervise other data. The analysis model that was developed in this research can be seen in Figure 18.

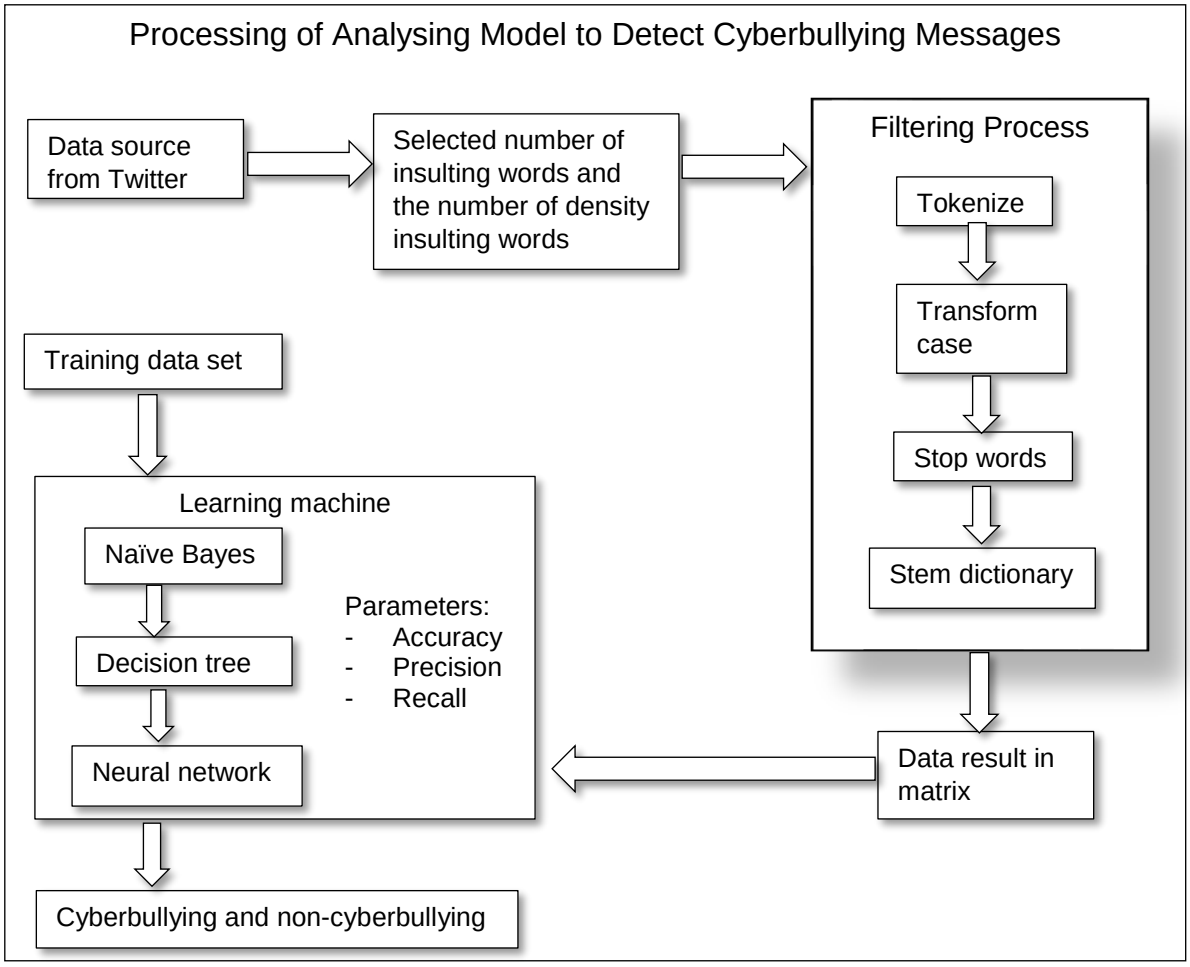


Figure 18 Processing of Analysing Model to Detect Cyberbullying Messages

Figure 18 depicts the analysis model that was developed using naïve Bayes, decision tree, and neural network. The performance assessment in terms of the model created by the author of this research can be calculated depending on the number of

accurate and inaccurate test records indicated by the model's prediction. These numbers are presented in table format, commonly known as the confusion matrix which is a useful mechanism for the accurate recognition of tuples from different classes by the classifier. The table below displays the confusion matrix.

Table 17 Confusion Matrix

		Predicted		
		yes	*no*	*Total*
Actual	*yes*	*tp*	*fn*	*tp+fn*
	no	*fp*	*tn*	*fp+tn*
	total	*tp+fp*	*fn+tn*	

tp = true positive
tn = true negative
fp = false positive
fn = false negative

The confusion matrix provides necessary information to determine the performance adequacy of the classification model. Nevertheless, if the information is represented by an individual number, this would be more convenient for comparing the performance of the different models. Thus, it can assist the accuracy, precision and recall. The accuracy of a classifier in a particular test set is the correct percentage of test set tuples classified by the classifier. The accuracy measurement involves taking the correct predictions percentage from the overall sum of examples (Han et al., 2012). Precise predictions are defined as those examples where the prediction attribute value is equivalent to the label attribute value. The following is the accuracy equation proposed by Han et al. (2012):

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} = \frac{tp + tn}{tp + tn + fp + fn}$$

Precision and recall are the other two parameters employed in the classification techniques. These two parameters are often applied to indicate an information retrieval system's condition. Moreover, the precision parameter is accepted to be the percentage of the restored and later classified documents in relation to the query. In terms of the classification process, the precision of a class withholds the number of true positives divided by the entire number of labelled components under an affiliation of the positive class (Han et al., 2012). For instance, in a text inquiry of a document set, precision is the number of precise outcomes divided by the entire number of results that have been restored. On the other hand, recall is the percentage of every related documents fetched from the data set. In this situation, recall is understood to be the number of true positives divided by the overall number of components residing in the positive class (Han et al., 2012). As an illustration, in a text inquiry for a group of documents, recall is the number of accurate results divided by the number of outcomes that should have been restored. The formula for precision and recall introduced by Han et al. (2012) is as follows:

$$Precision = \frac{(relevant\ document) \cap (retrieved\ documents)}{(retrived\ documents} = \frac{tp}{tp + fp}$$

$$Recall = \frac{(relevant\ documents) \cap (retrieved\ documents)}{(relevant\ documents)} = \frac{tp}{tp + fn}$$

In this case, the perfect score of accuracy, precision and recall are expected to obtain a great predictive label class data set and can help to decide the perfect prediction of the data class and evaluate the performance of the classifier model in terms of speed, robustness, and scalability.

4.4.1 Naïve Bayes

This section describes the basic concept of the naïve Bayes technique relevant to the functionality of the technique. Naïve Bayes is one of the statistical methods proposed by Thomas Bayes, a scientist. These methods predict future probabilities based on past experiences. The Bayesian classifier is a statistical classifier that can estimate the probability of class membership, such as the likelihood of a given tuple being owned by a certain class. The Bayes theorem explains the relationship between the probabilities of the occurrence of event A conditioned with event B, along with the occurrence of event B conditioned with event A. This theorem is based on the principle of improving probabilities with additional information.

The Bayes theorem is effective in altering and updating the calculated probability from the availability of the data and additional information. In accordance with a subjective probability, if one looks at event B and believes that there is a possibility of B appearing, the probability of B can be referred to as the prior probability. Once there is extra information that such events of A have emerged, perhaps there will be changes to the original estimation of the likelihood of B appearing. The probability of B is now the conditional probability of A and is referred to as the posterior probability. The concept of the Bayes' rule that was formulated by Thomas Bayes (Han et al., 2012; Raschka, 2014) is simply expressed as follows:

$$Posterior\ probability = \frac{conditional\ probability\ x\ prior\ probability}{evidence}$$

In naïve Bayes theory, the probability calculation of data categories depends on the previous data where the label has been known. The naïve Bayes technique performs the maximum result calculation to determine the category of a document.

The theorem is in conjunction with naïve Bayes theory whereby the conditions between attributes are assumed to be independent. Naïve Bayes classification assumes that the availability of characteristics in a class has no relevance to the characteristics in the other classes. Let $P(A|B)$ be the estimation of probability of a word occurrence in a document with class B category and $P(A)$ is the probability estimation prior to the appearance of category A.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

- $P(A|B)$ is the posterior probability of class (target) given predictor (attribute).
- $P(A)$ is the prior probability of class.
- $P(B|A)$ is the likelihood of predictor in a given class.
- $P(B)$ is the prior probability of predictor.

In the process of classifying the data, the maximum results are applied to decide the classes for every data item, so the formula given above can be:

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(a|b)$$

Map is the maximum a posteriori or most likely class

Bayes Rule $$C_{map} = \underset{c \epsilon C}{\operatorname{argmax}} \frac{P(b|a)P(a)}{P(b)}$$

Dropping the denominator

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(b|a)P(a)$$

Document b represented as feature $x_1, x_2, x_3, \ldots x_n$

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(x_1, x_2, x_3, \ldots, x_n|a)P(a)$$

Assume the feature probabilities $P(X_i|A_j)$ are independent given the class a, so

$$P(x_1, x_2, x_3, \ldots, x_n|a) = P(x_1|a) * P(x_2|a) * P(x_3|a) * \ldots * P(x_n|a)$$

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(x_1, x_2, x_3, \ldots, x_n|a)P(a)$$

$$C = \underset{c \in C}{\operatorname{argmax}} P(a_j) \prod_{x \in X} P(x|a)$$

Whereas: $$P(a_j) = \frac{doccount\ (a_j)}{N_{doc}}$$

$$P(w_i|a_j) = \frac{count(w_i, a_j)}{\sum_{w \in V} count\ (w_i, a_j)}$$

- $P\ (a_j)$: The probability for each data
- $P\ (w_i|a_j)$: The probability of the occurrence of the word w_i in class a_j.
- *doccount* (a_j) : The frequency of documents in each category
- N_{doc} : The total number of documents.

For example: $$P(bangsat|bullying) = \frac{count("bangsat", bullying)}{\sum_{w \in V} count\ (w, bullying)} = 0$$

Zero probabilities cannot be conditioned away, so add 1 when there is a word in testing data that is not available in the data training set. This condition is called Laplace (Han et al., 2012) which is used to avoid zero probability. Pierre Laplace states that all real numbers should be more than 0, a function f (t) for $t > 0$. The formula is as follows:

$$P(W|A) = \frac{count\ (w, a) + 1}{count\ (a) + |Total\ Vocabulary|}$$

- *count (w,a)* : The frequency of the word w to a in each category
- *count (a)* : The total amount of words that appear in each category
- *total vocabulary* : Total number of words in test document.

This naïve Bayes technique is typically known as a supervised learning method, requiring previous knowledge in order to be able to make decisions. The steps in the text classification process are:

1. Create class labels for each data in data training
2. Calculate the probability in each class
3. Determine the frequency for every Indonesian bullying word listed in classes
4. Calculate the probability of classes for each word in data testing
5. Calculate the maximum values probability
6. Determine the classes in each data testing document based on the maximum values.

4.4.2 Decision Tree

The decision tree is the process of learning in terms of predicting labelled data with classes through a study of a flowchart-shaped tree structure in which every internal node is a test of an attribute, every branch acts as a symbol for a test result, and every leaf node (or terminal node) carries a class label (Quinlan, 1996). The node located on the very top of a tree is known as the root node. A test node calculates several results depending on the attribute values of an instance in which every probable result is linked with one of the subtrees. An instance is classified by initiating at the tree's root node. Assuming that the node is a test, then the outcome for the instance is decided and the process proceeds by applying the suitable subtree. If a leaf is ultimately detected, the leaf's label provides an estimated class of the instance (Han et al., 2012).

Furthermore, a decision tree is able to be built based on a group of instances through the strategy of divide-and-conquer. When the entire instances have been

allocated to the same class, the tree's leaf has a label for that particular class. Alternatively, if a test has different outcomes, at least two of the instances are selected and partitioned depending on the outcomes. Moreover, the tree has a node on its roots indicating the test, including every outcome in sequence, and the equivalent subtree is obtained through the application of an identical operation to the instances' subset with that outcome.

Decision tree learning is an approach generally applied in data mining. The aim is to construct a model that predicts the target's value relying on several input variables. In other words, a tree is able to be "learned" by dividing the source set into subsets according to its attribute value test. The process involves the repetition of every acquired subset recursively, which can be known as recursive partitioning. The recursion process is accomplished if the subset at a node has the identical value of the target variable, or if the division stage is no longer producing a value and adding it to the predictions. This process is known as top-down induction of the decision tree (Quinlan, 1996).

In data mining, a decision tree can be described as a mixture of mathematical and computational approaches that assist with the description, categorisation and generalisation of a particular group of data. Data is recorded as follows:

$(x, Y) = (x_1, x_2, x_3, \ldots x_n, Y)$

The dependent variable, Y, is the target variable to be classified. The vector x is comprised of the input variables, which are x_1, x_2, x_3, …x_n, used for that particular task. For example, to label class data *Y* (cyberbullying or non-cyberbullying), suppose *x* is a data message recorded in the database, then the vector *x* consists of the input variables, $x_1, x_2, x_3 \ldots x_n$. = (*bastard, scoundrel, idiot, stupid, crazy, ugly*….x_n), meaning that the class of data is cyberbullying. Therefore, class label *Y* depends on the content

of the variable *x*. If the content of variable *x* is positive, then the class label of *Y* is in the positive category; conversely, if the content variable *x* is negative, then the class label of *Y* is in the negative category.

There are three popular decision tree algorithms introduced by Han et al. (2012), which are listed as follows:

1) Iterative Dichotomiser 3 (ID3) is an algorithm established by Quinlan (1996) and applied to produce a decision tree derived from a data set. Typically, ID3's implementation is primarily in the machine learning and natural language processing domains. The initiation of the ID3 algorithm involves the original set S serving as the root node. Furthermore, in each algorithm's iteration, every untouched attribute of set S is iterated. Apart from that, the entropy (E) and information gain (IG) of that particular attribute are calculated, which then chooses the attribute with smallest entropy and largest information gain value. The next stage involves the splitting of set S according to the chosen attributes in order to generate data subsets. The algorithm progresses to each subset recursively as some of the attributes have never previously been chosen. However, the recursion process on a subset may be halted for one of the following reasons:
    - Each element within the subset resides in the same class (whether positive or negative). At that point, the node is converted into a leaf and labelled under a class of items.
    - The selected attributes are non-existent; however, the items are still not affiliated with the same class, (some are positive and some are negative). Thereupon, the node is remodelled into a leaf under the label of the most general class in terms of the items in the subset.

- No items exist in the subset. This scenario occurs when items in the parent set are not to be found as a matching and specific value of the chosen attribute. For instance, assuming that no item with >= 100 is to be found. In this case, a leaf is generated and labelled under the most typical class in terms of the items residing in the parent set.

Entropy is a measurement used for calculating the amount of uncertainty in data set *S*, or strictly speaking, entropy represents data set *S:*

$$E(S) = \sum_{i=I}^{c} -p_i \log_2 p_i$$

where,

- S : The current data set where the entropy is being calculated. This means that every iteration of the ID3 algorithm is converted.
- I : Set of classes in *S*
- $p(i)$: The ratio between the number of elements residing in class *i* as well as the number of elements in set *S*.

Information gain *IG(A)* calculates the differences between the entropy from the previous procedure and the subsequent procedure that includes dividing set *S* on an attribute *A*. The declining number of uncertainty in *S* after the splitting stage of set *S* on attribute *A*.

$$IG(A, S) = H(S) - \sum_{c \in C} P(c)E(c)$$

where,

- *H(S)* : Entropy that belongs to set S
- *C* : The subsets constructed upon the split of set S via attribute A, which is displayed as: $S = \cup_{c \in C} C$

- *p(c)* : T The ratio between the number of elements residing in class c and the number of elements located in set *S*
- *H(c)* : Entropy that belongs to subset *c.*

Furthermore, the information gain can be calculated for each attribute in *ID3*. If an attribute has the greater information gain, it will be applied to divide set *S* on this iteration.

2) C4.5, known as the successor of ID3, is an algorithm which was proposed and developed by Quinlan (1996) to construct a decision tree. In other words, C4.5 is the continuation of Quinlan's previous ID3 algorithm. The generated decision tree through C4.5 is applicable for further classification processes; hence, it is commonly known as a statistical classifier. Likewise ID3, C4.5 creates a decision tree based on a set of training data by applying the notion of information entropy. In other words, data training is a set of $S = s_1,s_2,s_3\ldots, s_n.$, or data that have been classified. Every sample s_i comprises a vector with a p-dimensional nature; that is, $(x_{1,i}, x_{2,i}, x_{3,i},\ldots x_{n.i},)$ whereby x_i symbolises attribute values or sample's features including which class that s_i resides in. In every node of the tree, C4.5 selects attributes of data that often divide its group of samples into subsets enriched in a class. Hence, the splitting principle is the information gain that has been normalised. Therefore, the attribute with the maximum normalised information gain is selected in order to create a decision. Following this step, the C4.5 algorithm is repeated for the small sub-lists. Moreover, the C4.5 algorithm has fewer base cases.

- The entire samples mentioned in the list exist in the identical class. As this occurs, a leaf node is openly generated for that decision tree to select that class.

- Not all features offer information gain. Tackling this case, C4.5 generates a decision node located in a higher position on the tree by applying the predicted value of the class.
- Class instances emerge that had not been seen earlier. As stated previously, C4.5 builds a higher positioned decision node on the tree via the expected value.

3) Classification and Regression Tree, abbreviated to CART, was introduced by Breiman et al. (1984). The analysis by classification tree is assumed to be the predicted class in which the data resides. Meanwhile, the analysis by regression tree occurs when the predicted result is considered to be a real number. The common ground between regression and classification is the similar trees used. However, the difference between the two approaches is the method applied to select the area to be split. CART is a non-parametric decision tree learning approach that generates a classification or regression tree, contingent upon whether the dependent variable is in categorical form or numeric form.

ID3, C4.5, and CART endorse a method where the decision tree is built on a top-down style in a recurrent divide-and-conquer way. In addition, almost all of the algorithms for inaugurating a decision tree follow the same top-down method, initiating from a training group of tuples and its correlated class labels. The training set is partitioned repeatedly into lesser subsets while the tree is being constructed.

4.4.3 Neural Network

A neural network is a mathematical or computational model influenced by the structural and functional elements similar to a biological neural network (Han et al.,

2012). It consists of a set of interconnected artificial neurons, dealing with information through the use of a connectionist approach extending to computation (the central connectionist technique involves describing a mental experience by interconnected networks consisting of basic, mostly consistent units). Often, the ANN system is flexible, altering its structure depending on the external or internal information, which proceeds straight to the network during the learning stage. Furthermore, modern neural networks are typically implemented in modelling sophisticated relationships, comprising inputs and outputs, simultaneously identifying patterns in data (Akthar and Hahne, 2012).

As proposed by Nettleton (2014), an ANN is usually defined by three parameters, which are:

1) The interconnected patterns in between different layers of neurons, as every layer is comprised of units. The inputs to the network correlate to the attributes that are calculated for every training tuple. The inputs are simultaneously nourished with data, which forms layers of input. Next, these inputs flow through every input layer, which is then loaded and fuelled synchronously to the second layer with “neuron-like” units, otherwise known as a hidden layer. In other words, the outputs from the hidden layer can serve as the inputs to the other hidden layer, and this process continues. The interconnected patterns in between the input layer and the hidden layer are more condensed than those between the hidden layer and the output layer
2) The total weight of the interconnections that are regularly updated while in the learning process. A weight is appointed to every connection that is applied as an influence for the data flow via the connection. Typically, the weight has values between 0 and 1. The higher the value, the easier is the access for the

data to flow; meanwhile, the lower the weight, the more restricted is the data connection. The weight formula proposed by (Han et al., 2012) is:

$$W_{sum} = \sum_{k=1}^{n} w_k \, x \, i_k$$

Where w_k is the weight of the connection from the previous layer of the unit k layer, while i_k is the unit k 's input located from the past layer

3) The functionality of the activation function is altering the input of a neuron's weight into its activation of the output. During the training stage of neural network processes, the locations of a few neurones are activated. Meanwhile, over in the other sections, neurons continue to be inactive due to the result of applying data input. There are several activation functions that can be implemented in classification models; however, this research employs the sigmoid activation function in order to provide an easy and simple calculation by means of the derivatives of sigmoid (Basterretxea et al., 2004). In the gradient calculation process, the derivatives are necessary for training the neural network. Moreover, the output has constant values of 0 and 1 that can also be expressed as a probability evaluation. Additionally, the derivatives do not overwork the activation process since, when they are restrained, the data training activation within the network cannot advance further (Kros et al., 2006). Below is the formula of the sigmoid activation function introduced by Weisstein (2002c):

$$f(x) = \frac{1}{1 + \ e^{-x}}$$

Where f is the sigmoid curve, x is the values ranging from positive to negative most often going either from 0 to 1 or from −1 to 1, and e is the natural logarithm base (Euler's number = constant value = **2.72**).

Generally, the three parameters above are used in two popular types of neural network: feed-forward and back propagation (Han et al., 2012).

a) A feed-forward neural network is an artificial neural network in which no form of a directed cycle is built upon connections between units. In this kind of network, the information flows in one direction straight to the output nodes, if the situation allows it, via any hidden nodes and to the final destination which is the output nodes. In other words, no cycles or loops are formed within the network.

b) The back propagation algorithm is a supervised learning approach that has two stages: propagation and weight update. These two stages are repeated until the network's performance satisfies the standard. In terms of the back propagation algorithms, the output values are examined against the correct answer in order to calculate some predefined error-function values. Through numerous methods, the error is delivered back via the network. With the acquired information, the algorithm adapts to the weights from every connection for the purpose of decreasing the error function values by a small amount. Following the repetition of this process until there is a sufficient number of training cycles, the network typically assembles into few conditions with small calculations of error. Having reached this state, the network can be considered to have learned a specific target function.

4.5 Implementation of Machine Learning for Data Analysis

As explained in the previous chapter, this research has developed an analysis model using several machine learning techniques for data mining to find features in text patterns, especially the patterns of Indonesian insulting words categorised as bullying words. In this case, machine learning tools are used to analyse data and to evaluate the development of the analysis model since they are able to indicate and assess the success of learning experiments.

This section explains the development of analysis models for classifying messages from Twitter as cyberbullying or non-cyberbullying messages. In order to establish groups of data classes, this research used three machine learning classification techniques: naïve Bayes, decision tree and neural network. By using three techniques simultaneously, a high accuracy of prediction class data can be achieved and high precision and recall data can be attained.

4.5.1 Learning the Model

Rapid Miner assists the machine learning process that builds analysis models through the application of an extensive range of leading algorithms (Akthar and Hahne, 2012). Furthermore, Rapid Miner offers a user-friendly integration of recent and well-established data mining techniques. The Rapid Miner Studio analysis procedure involves dragging and dropping the operators, assembling parameters, and linking the operators. Rapid Miner Studio comprises over 1,500 operations covering all aspects of professional data analysis ranging from data partitioning, to market-based analysis, and to attribute generation (Akthar and Hahne, 2012). These processes offer all the tools required for subsequent stages of data analysis. However, there are other available approaches such as text mining methods, web-mining, the automatic

sentiment analysis derived from Internet discussion forums (sentiment analysis, opinion mining), timer series analysis and prediction (Klinkenberg, 2013).

This research has identified the following algorithms as most useful for analysing cyberbullying messages. By utilising the functions of Rapid Miner, this research applied naïve Bayes, decision tree and neural network techniques to analyse the Indonesian cyberbullying messages. This learning model required several processes explained below:

1. Retrieving the training data set from the repository was the same process as in Chapter 3 where it has been explained in section 3.4.
2. The processing document function is used to clean data containing operators such as tokenize, transform case, stop words, and stem dictionary. By dragging and dropping the process document operator to main the view in Rapid Miner, several parameters were set based on the required measurements such as a word vectors as well as the output of this process, prune method, selecting attribute and weight as a target data that will be cleaned. In this case, the term frequency–inverse document frequency (TF-IDF) was chosen as a parameter of the word vectors because the TF-IDF values increase relative to the number of words appearing in the document. However, it is counterbalanced by the frequency of the word located in the corpus as it helps to identify frequently-occurring words. TF-IDF is part of the well-known term-weighting arrangements. The complete value was determined as a parameter through the pruning approach with a given range value of 2 for pruning below the absolute, meanwhile, 80,006 for pruning above the absolute. For selecting the attribute and weight parameter, text attribute has been determined to be a target data for analysis and was given a value by default of 1.0 as the weight.

3. An analysis model was created by means of three classification techniques, naïve Bayes, decision tree and neural network, by dragging and dropping the these classification techniques into X-validation operator on main process view in Rapid Miner. In this stage, every X-validation operator consisted of one technique: the first X-validation operator consisted of the naïve Bayes technique, the second X-validation operator comprised the decision tree technique and the third X-validation operator applied a neural network technique. Moreover, the X-validation operator involved executing a cross-validation process for the purpose of evaluating the statistical accomplishment of a learning operator, typically on undetected data sets. Mainly, it is applied to assess the accuracy of a model's actual performance (learnt via a specific learning operator). In other words, the X-Validation operator is an installed operator that holds two sub-processes, which are a training sub-process and a testing sub-process. The training sub-process is applied for training a model; meanwhile, the trained model is implemented to evaluate the testing sub-process. The model's performance is calculated during the testing stage. The input from the item set created in the processing document is separated into k subsets in a balanced size. From the k subsets, an individual subset is maintained for a testing data set, which is the input for the testing sub-process, while the leftover $k - 1$ subsets are applied further as a training data set or, in other words, the input of the training sub-process. Then, the cross-validation process is done recurrently k times under a condition where every k subset has been used once as the testing data. Following this stage, the k outcomes derived from the k iterations can be calculated for the average value, or in other cases combined to create a sole prediction. However, the value of k can be

modified by using a number of validation parameters. Moreover, the learning processes in this phase typically advance the model in order to create a more appropriate model for the training data.

The three stages of building an analysis model involve connecting the three stages simultaneously, then running the model. The three models of classifier techniques were set to three X-validation operators to estimate the statistical performance of the learning model in terms of the data prediction. Then, the operator performed an accuracy, precision and recall test of the model. The results of this model's performance are presented in Table 18.

Table 18 shows the model's performance of the training set based on the estimation of accuracy, precision and recall. Based on the result, the three techniques have different performance in terms of predicting cyberbullying and non-cyberbullying classes.

Table 18 Performance Model Training Set using Naive Bayes, Decision Tree (C4.5), and Neural Network

Class labels of data	Parameters	Naïve Bayes	Decision tree (C4.5)	Neural network
Prediction Cyberbullying	% Accuracy	100	99.97	99.99
	% Precision	100	100	100
	% Recall	100	99.96	99.99
Prediction Non-cyberbullying	% Accuracy	100	100	100
	% Precision	100	99.85	99.97
	% Recall	100	100	100

Although, naïve Bayes technique is a basic classifier technique, its performance in this model achieved 100% in accuracy, precision, and recall for both the cyberbullying and non-cyberbullying classes. This indicates that naïve Bayes has high performance in relation to document retrieval and in this case, the parameter in naïve Bayes is set up by default. The decision tree has an accuracy of 99.97% in the

cyberbullying class and 100% in the non-cyberbullying class. The values for precision and recall for the cyberbullying and non-cyberbullying classes were distinctive. The performance of the decision tree in the cyberbullying class was 100% for precision and 99.96% for the recall. In contrast, the performance of the decision tree in the non-cyberbullying class was 99.85% for precision and 100% for recall. This is due to the setup of parameters having different values based on the weighting.

The performance of neural network also has a different value in the cyberbullying and non-cyberbullying classes. In the cyberbullying class, the accuracy of performance of neural network was 99.99%; whereas in the non-cyberbullying class it was 100%. The precision and recall values in both class were also different. This is similar to decision tree where the performance value of precision was 100% for the cyberbullying class, compared to the non-cyberbullying class that had 99.97% precision. The performance of recall for the cyberbullying class was 99.99%, but for the non-cyberbullying class it was 100%.

4.5.2 Class Weighting

In the learning process, selecting weight for the training data set is strongly recommended in order to achieve a high level of accuracy in the prediction of data testing (Akthar and Hahne, 2012). Weight can be assigned to correct classifications (Klinkenberg, 2013). Hence, weight can be applied for further calculations of statistical performance. This parameter has no repercussion when there are no attributes that carry a weight role. To consider the weights of examples, the data set input is recommended to have an attribute with a weight role (Akthar and Hahne, 2012; Klinkenberg, 2013). Apart from this, assigning weights in Rapid Miner is possible with the various available operators.

Therefore, the models of the training data set mentioned each have a different weight, so that naïve Bayes, decision tree, and neural network also have unique performances. Although, the weights for every technique have been assembled by default, the application of the three techniques produced results that had high accuracy, precision and recall. In this case, the establishment of weighting for every technique was not altered, because the outcome had an expected high value.

As discussed earlier, the process of labelling the data training set involved several processes both using machine learning. Furthermore, the performance of the three techniques was adequate and there was no need to change the weighting values from default to other values.

4.5.3 Results

This section describes the process of accurately predicting instances of cyberbullying or non-cyberbullying data from Twitter. The highly accurate result was significant since this research focused on identifying the cyberbullying messages. Although some research involving the analysis of cyberbullying messages has been conducted previously, the result of identifying cyberbullying messages in this research contributes to knowledge, specifically regarding the development of analysis model using three classification techniques.

As mentioned in Table 18, the development of a model training set and its performance had to be considered to obtain an accurate analysis of the training data set. The model performed with a high level of accuracy, precision and recall of data, so the confidence of running the model to analyse the data set was high. After running the machine learning in Rapid Miner, the results were converted to table and figure formats to provide an overview of the data in more detail.

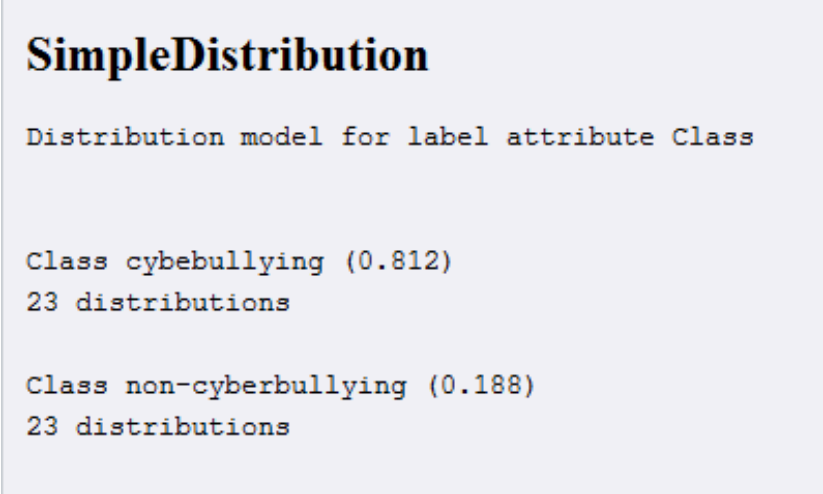


Figure 19 Label Attribute Data Class into Cyberbullying and Non-Cyberbullying

Figure 19 shows simple label attribute data in both the cyberbullying and non-cyberbullying classes. The model analysis that was developed successfully divided the data into two classes (cyberbullying and non-cyberbullying). Figure 19 demonstrated that 23 insulting words were spread across both class. This indicated that although the messages contained insulting words, the messages might be non-cyberbullying ones, although this depends on the content and context of messages. Figure 19 also illustrated that the posterior probability of label cyberbullying was0.812 and the posterior probability of label non-cyberbullying was 0.188. This suggested that the probability of the cyberbullying class in the data set was dominant compared to the non-cyberbullying class. In other words, this also meant that most of training data set items were considered to be instances of cyberbullying.

Based on the experiment results, the density of insulting words was also indicated. There were six insulting words with a high value of density estimation: *babi* (pig), *bajingan* (scoundrel), *bangsat* (bastard), *goblok* (stupid), *iblis* (devil), and *monyet* (monkey). These words were mostly allocated to the cyberbullying class. It can be

seen that the six insulting words had a major influence on the labelling and allocation of data to the cyberbullying class. The graph of the density for the *iblis* term that was one of six insulting words is depicted in Figure 20.

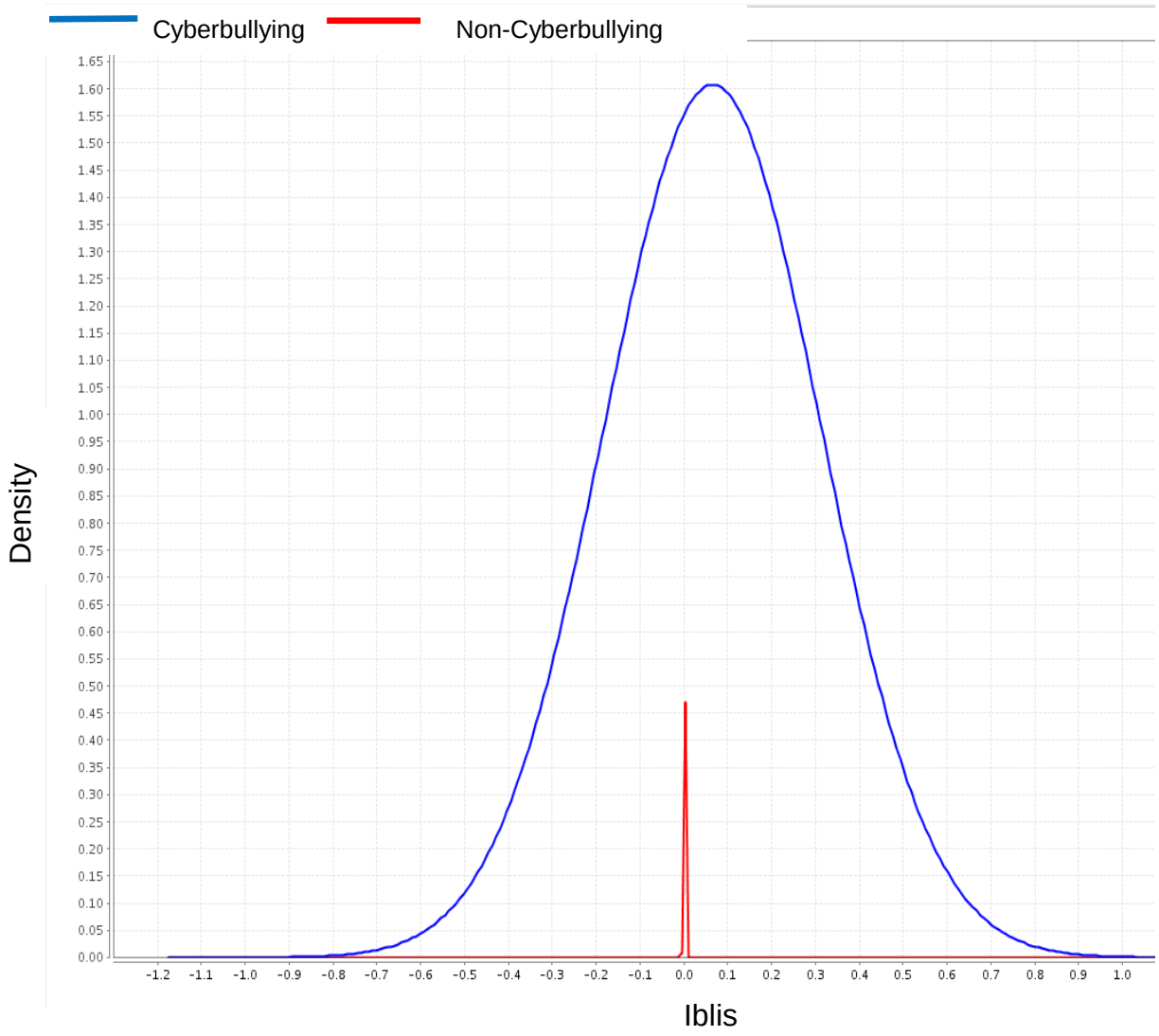


Figure 20 Density of the Iblis term in Cyberbullying and Non-Cyberbullying Class

Figure 20 demonstrates that the graphs of the *iblis* term that occurred holds high density in both the cyberbullying and non-cyberbullying class. It can be assumed that the six insulting words are those that most suggest cyberbullying or reasonably, these words were more commonly used than the other more rarely used insulting words that appeared in the messages.

Another parameter appeared after generating the model had a confidence measurement in both cyberbullying and non-cyberbullying classes. Parker et al. (2005) proposed that an output of a classification system is a confidence evaluation. The classification confidence measurement, in this case, is understood to be the parameter of a correct label (Akthar and Hahne, 2012). A high value indicates that there is a high probability that the model is correct (Gelman et al., 2014). In this scenario, the confidence evaluation result was confirmed by the result from generating the entire model in terms of data set analysis. An example of a confidence measurement is presented in Table 19.

Table 19 Example of Calculated Confidence Measure for Labelling Correctness

No.	Tweet	Confidence of cyberbullying	Confidence of non-cyberbullying	Prediction Class
1	Avanya andaiy, avanyaâ™¥RT mohamadDFDM: "sarap!! addnan_ch: Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn aaaaahhhh an	1	1.23E-15	cyberbullying
2	GILA GILA ANJING GANGERTI SM HENRY BANGSAT GUE GAKUAT WOY MSH GO NIH WOY	1	8.60E-12	cyberbullying
3	sarap!!addnan_ch: Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn aaaaahhhh anjing anjing ava nya si eta ah anjing ah. Cageur?addnan_ch: Oh anjing goblog fakyu siah monyet setan babi alas bangsat shit damn aaaaahhhh anjing anjing ava nya si eta ah anjing ah	1	1.23E-15	cyberbullying
4	Bangsat deuh tolol =)) RT"syemaAT: Anjing lu setan"	1	1.47E-15	cyberbullying

The scope of the confidence measurement for the whole system of classification is simplified as a number between 0 and 1 (Parker et al., 2005). Assuming that the resulting value is closer to 1, then the estimation of true class condition is high. In contrast, when the resulting value is closer to 0, then the estimation of false class condition is high.

Interestingly, the results from the developed model were two classes for the 152,843 data items: 122.842 items (80.37%) in the cyberbullying class; 30,001 items (19.63%) in the non-cyberbullying class. A close analysis of three classification techniques resulted in an outcome of a correct label which showed the larger number of predictive data in cyberbullying compared to the non-cyberbullying class. Figure 21 shows the distribution of the two classes.

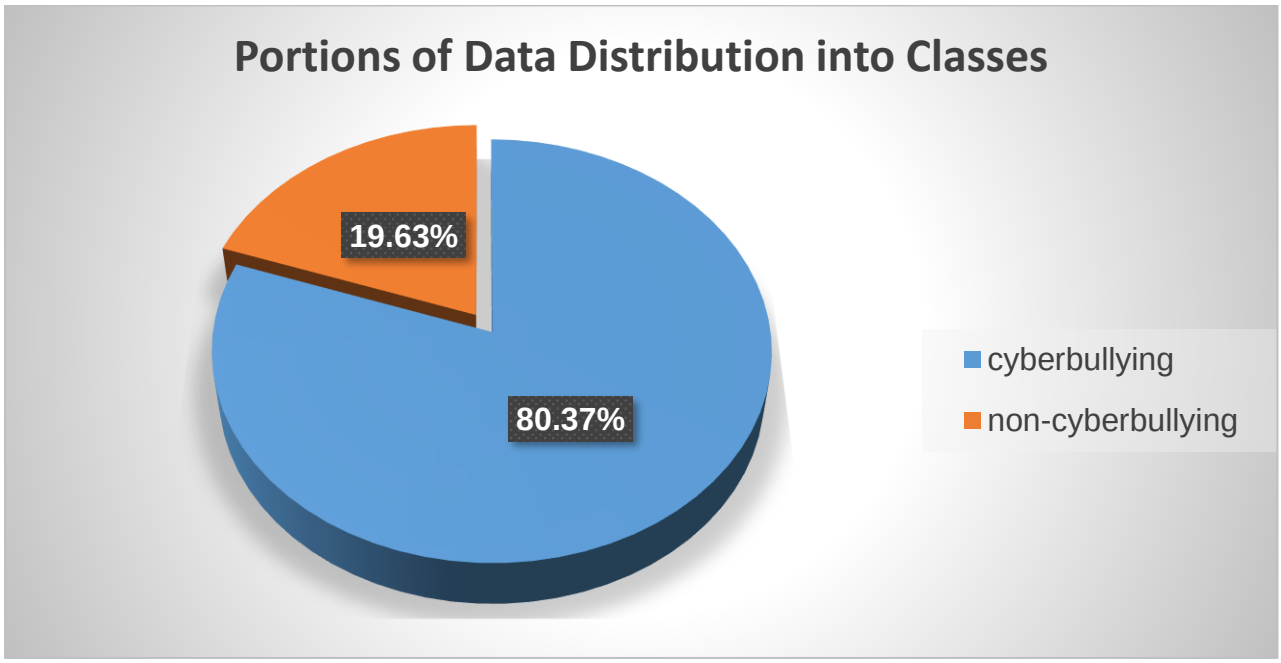


Figure 21 Chart of the Result Prediction Cyberbullying and Non-Cyberbullying Data

Figure 21 shows that most data were detected as containing cyberbullying messages and a small portion of data was identified as non-cyberbullying messages. At first glance, it is clear that most of the data, after generating the machine learning using the classification techniques, was in the cyberbullying category. Since the models used for analysing cyberbullying data discussed above performed with high accuracy, the definite result after testing the data set in the cyberbullying class was 80.37% of the total amount of data. However, only a small proportion of 19.63% was

in the non-cyberbullying class. This small proportion may comprise of exasperation towards self in terms of various context, as instanced in "Bangsat gua terus mendapatkan sial, keparat emang" (translation: Bastard, I always receive bad luck, what an a*s). Other reasons can be annoyed and negative expressions towards their personal pets, such as "Sial anjing gue gigit tangan gue" (translation: sh*t my dog just bit me), or positive reactions to their pets, for instance "Anjing gue ternyata lucu juga" (translation: My dog is actually hilarious too). Another viewpoint that can be taken is the insulting words may be used to convey what people are feeling at the moment, for example, "AC keparat dinginnya" (translation: The air conditioner is f***ing cold). From the 19.63% of non-cyberbullying data, this signifies that the insulting words used in the messages tend to be independent and have no relationships among the words and messages.

4.6 Conclusion

In this chapter, the number of insulting words-based method of detecting cyberbullying used a model of classification. The classification techniques that were employed were naïve Bayes, decision tree and neural network. With the assistance of Rapid Miner it was possible to apply the three classification techniques simultaneously.

From the experiment results obtained, several phases were necessary in order to achieve a high accuracy of data prediction; to be more specific, the simultaneous application of unsupervised and supervise learning techniques. Unsupervised learning such as the *k*-medoids was used to divide the data training set into two clusters. The purpose of this was to group data and label it as either cyberbullying or non-cyberbullying based on the data patterns that emerged, as described in the previous chapter. Moreover, to obtain highly accurate labelling of the data training set, *k*-

medoids and data patterns of relationship between itemsets were applied to validate the results of the labelled classes.

Supervised learning such as naïve Bayes, decision tree and neural network were connected synchronously and adopted to analyse the data. The process of data analysis required an examination of the analysis model performance. The main goal of this process was to achieve a high level of accuracy in data class prediction. Hence, the processes of data precision and recall in the model were also employed. As a result, the model had a high level of accuracy, precision and recall. Therefore, there was great confidence in using the model to analyse messages from Twitter. Moreover, the analysis model was able to correctly identify 80.37% of the data consisting of cyberbullying messages and 19.63% for non-cyberbullying messages.

This research contributes to the development of the analysis model and the implementation of data mining techniques for the study of cyberbullying messages by applying unsupervised and supervised learning techniques. Furthermore, these results also provided important data that can be used in further research for the development and application of data mining techniques by building an analysis model for social issues.

The next chapter presents the analysis and discussion of the results obtained through both association rules and classification techniques. The outcomes from determining whether messages are classified as either cyberbullying or non-cyberbullying explored in this chapter will be used as the basis for constructing a framework and identifying cyberbullying messages from the social network. Furthermore, the experiment used for the development of the analysis model can serve to improve the further use of machine learning for data derived from the social networks.

Chapter 5
Discussion

5.1 Introduction

Cyberbullying has been a prevalent global phenomenon that requires serious attention (Nahar et al., 2012; Kasture, 2015). Therefore, the research in this thesis aimed to detect cyberbullying through the messages collected from the social networks and exploring cyberbullying characteristics through identifying insulting term patterns was the main objective. This was achieved by obtaining the messages from social networks in order to detect cyberbullying messages. The main findings of the research in terms of its contributions are: cyberbullying recognition frameworks, development of analysis models for cyberbullying, expanded measurement of data relationships and the extended measurement of data class prediction. Chapter 5 discusses the further overview of the research of this thesis, including research findings and the contribution, strengths and limitations of this research.

This research was conducted to analyse cyberbullying messages from social networks in the Indonesian context. The research design included the implementation of data mining techniques to examine the harmful messages identified, based on the density of insulting words as an important potential indicator of patterns of cyberbullying messages. These patterns were used to detect harmful messages which may or may not be instances of cyberbullying. Data were obtained by randomly extracting from tweets containing predominantly harmful messages. Although this research was focusing mainly on analysing the harmful messages, an analysis model was developed using data mining techniques for other purposes such as: extending the measurement parameters of relationship between data; assisting the extension of

concepts and instruments: and finally, supporting the interpretation and clarification of this research results.

The construction of the analysis model for social issues, in particular the detection of cyberbullying messages on social networks, was accomplished by applying data mining techniques. The analysis of the cyberbullying messages and the implementation of the data mining techniques were discussed to provide insight regarding how to identify the characteristics of cyberbullying messages using the analysis model. The application of data mining techniques for the analysis of cyberbullying messages improves our understanding of the various lexical features of the cyberbullying words extracted from social networks.

This chapter makes an important contribution to the extant literature by examining how cyberbullying messages conveyed via social networks can be identified. The analysis model proposed in this research was constructed using a combination of naïve Bayes, decision tree, and neural network techniques in order to develop a processing model for detecting potentially harmful cyberbullying messages by means of machine learning. Furthermore, this research outcome can be used as a resource to illustrate instances of cyberbullying using Indonesian terms.

This chapter is divided into several sections. The first section discusses the characteristics of cyberbullying messages which enable these to be identified as such. The second section discusses the development of various analysis models using data mining techniques. The final section discusses an effective way of identifying cyberbullying messages on social networks.

5.2 Review of This Research

Six data mining techniques were employed in this research and were used as tools in analysing Twitter messages within the Indonesian context. In the first analysis, association rules techniques of FP-growth and cosine similarity were implemented to discover the patterns of a data set as a representation for cyberbullying messages. For the next analysis, three additional techniques were used to separate the messages into two classes of data; these techniques were naïve Bayes, decision tree and neural network. A detailed account of how each of these analysis was operationalised and measured can be found in Chapters 3 and 4. Confidence, laplace and conviction were applied to the results presented in Chapters 3 and 4, in order to measure the strength of the relationship between items in the data set. The data set prediction was further measured in terms of accuracy, precision and recall.

In the following section, the main findings are reviewed for each analysis followed by a discussion of the implications of these findings in future studies. Several suggestions are made concerning the relevance of these findings for analysis model development applied in cyberbullying identification through the data mining approach.

5.3 Identifying the Cyberbullying Characteristics

As discussed in Chapter 2, cyberbullying can be defined as the repetitive acts of intimidation and aggressive harassment by an individual or group with the intention of intimidating or harming a targeted person by means of insulting text messages, pictures and videos sent via communication technologies (Smith et al., 2008; UNDP, 2013). This indicates that cyberbullying does not solely involve one or more verbal attacks; it involves a range of electronic media used to send possibly destructive messages to the targeted victims, while the perpetrators whether as an individual or

with the assistance of friends and followers perform the constant cyberbullying action within a certain period of time.

By understanding the characteristics and forms of cyberbullying messages, positive or negative messages can be identified through the characteristics of the contents of messages (Sanders et al., 2009). This means that in order to determine whether or not a message is an instance of cyberbullying, it is important to know, understand and recognize the characteristics of messages because every sentence in a message contains a combination of words that conveys either a positive or negative sentiment.

Willard (2006) created various classifications for cyberbullying messages. These include: flaming, harassment, denigration, impersonation, outing and trickery, and cyberstalking. Cyberbullying is characterised by the anonymity of the perpetrator (O'Brien and Moules, 2010); the timing for sending and re-sending harmful messages (Privitera and Campbell, 2009); the intentionality of re-sending messages (Spears et al., 2009); and the actual content of messages (Mishna et al., 2009).

The characteristics and form of cyberbullying explained above provide a rich opportunity to examine common findings and emerging models concerning the identification of other characteristics of cyberbullying messages. This research suggests that there is another method that can be used to recognize the characteristics of cyberbullying messages - that is, detecting the pattern of insulting words that frequently appear in social network messages. The recognition of patterns of insulting words is another important means of discovering whether messages contain cyberbullying.

This research explains the process of finding the pattern of insulting words using data mining techniques which helps to detect the characteristics of cyberbullying

messages. Figure 1 below shows the cyberbullying message characteristics mind map expanded with the findings of the research in this thesis of insulting word patterns.

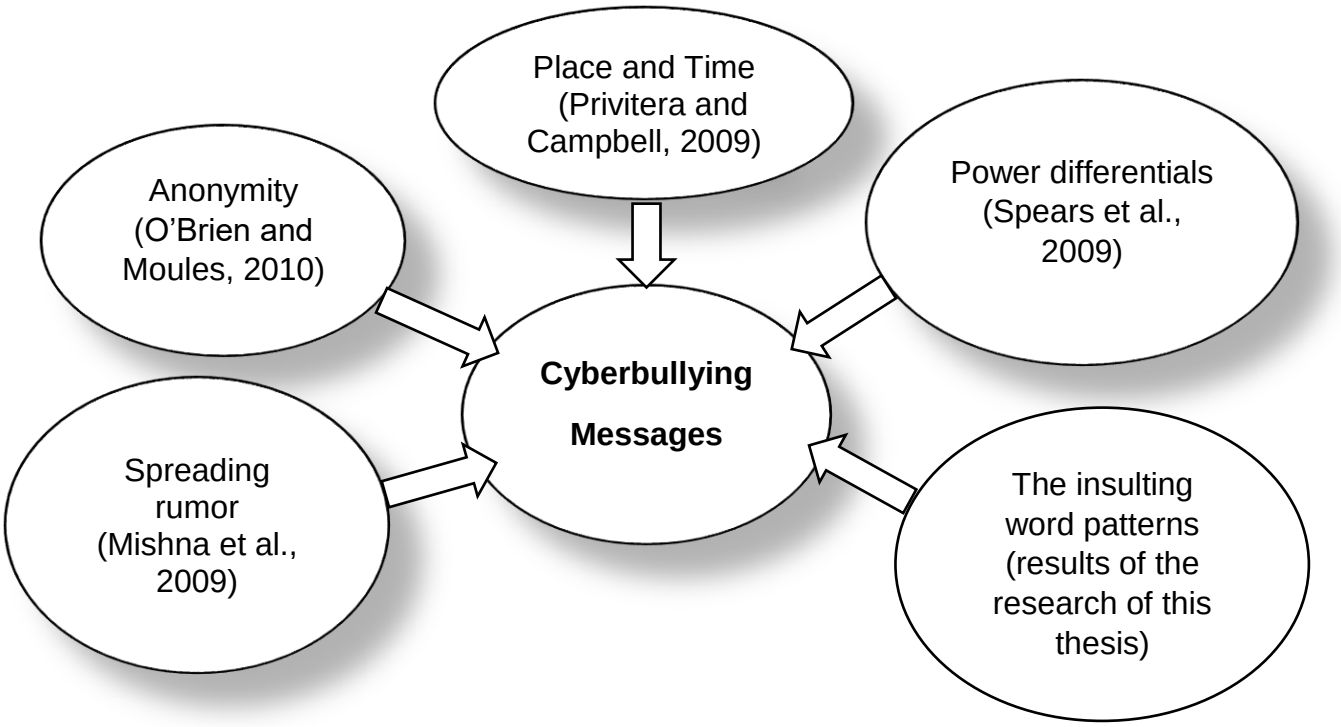


Figure 22 Framework of Cyberbullying Message Characteristics

In this case, the findings in this research show that the patterns of Indonesian insulting words can be used to detect cyberbullying messages from Twitter. Three general principles can be applied in the identification of Indonesian cyberbullying messages. First, the patterns of Indonesian insulting words can be used, but there are constraints. That is, the patterns of Indonesian insulting words can be used to identify cyberbullying in the Indonesian context, but cannot be used to identify cyberbullying messages conveyed in other languages. Second, the patterns of Indonesian insulting words have undergone a measurement process to determine the strength of the relationship between insulting words. Third, the identification of the cyberbullying messages can be investigated empirically by observation and experience in finding the patterns of the insulting words that often occur in cyberbullying messages.

A combination of hierarchical assigning tasks and structured mapping tasks was used by applying association rules techniques of FP-growth and cosine similarity in order to find the data patterns. In addition, a combination of naïve Bayes, decision tree and neural network are used to detect the classification of data labels. These combinations were effective as it produced relevant results and was more time-efficient compared to a manual cyberbullying identification method. The findings presented in Chapters 3 and 4, among other things, illustrate a distinctive difference between those of previous research and this research. First, in previous research the cyberbullying messages were identified through the anonymity of the perpetrator (O'Brien and Moules, 2010); the timing for sending and re-sending harmful messages (Privitera and Campbell, 2009); the intentionality of re-sending messages (Spears et al., 2009); and the actual content of messages (Mishna et al., 2009). However, this research has successfully identified cyberbullying messages by examining the patterns of insulting words. Therefore, the patterns of insulting words can also be another means of identifying cyberbullying messages. Second, the outcomes of experiments have shown that the results obtained from the structural combination of using data mining techniques are highly accurate in identifying cyberbullying messages. The reason for this is that a well-design analysis model developed by the author in this research resulted in high accuracy in data prediction through extended parameter instruments such as confidence, laplace, conviction including accuracy, precision and recall. The result of high accuracy in data prediction was explained in Chapter 4.

The structured mapping task of the analysis model was explicitly designed in this research in such a way that the author can easily recognise and develop the identification process of an analysis model in future work. Therefore, to conclude this

section, the results obtained by the identification process indicate a high level of accuracy in the identification of cyberbullying messages. Also, the patterns of insulting words that have been found in this research contribute to the detection of cyberbullying words in message contents.

5.4 Cyberbullying on Indonesian Twitter

Social networks are a popular media for sending cyberbullying messages (Dredge et al., 2014; Whittaker and Kowalski, 2015) because they provide access to every user, since messages are posted and accepted openly. Social networks offer facilities enabling people to send their ideas as text, pictures, and video without any restrictions (Duggan et al., 2015; Lenhart, 2015). Therefore, they provide abundant opportunities for people to send both positive and negative messages. If the messages received are offensive and harassing, then they can be considered as cyberbullying (Ybarra et al., 2006; Livingstone et al., 2010). Therefore, early prevention and detection of cyberbullying on social networks can lead to a significant improvement in reducing the frequency of spreading cyberbullying messages.

Although several other popular social networks are used for cyberbullying, such as Facebook, YouTube, Ask.FM, Instagram, and Tumblr (Ellis, 2013), Twitter is one of the social networks that is increasingly being used as a means of conveying cyberbullying messages (Fitzgerald, 2012). In Indonesia, Twitter has also become one of the most popular media for cyberbullying (Wijaya et al., 2013). Research conducted by Yulianti (2014) has confirmed a high prevalence of cyberbullying on social networks among Indonesian adolescents, one of these networks being Twitter. This indicates that many teenagers in Indonesia have engaged in cyber victimisation.

The increase of cyberbullying in Indonesia is related to the fact that Indonesian children and teenagers do not have adequate knowledge of the internet, and their parents and Government provide insufficient protection against cyberbullying (Jong, 2016). Another reason for the high number of cyberbullying incidents may be that the Indonesian authorities are not enforcing strict laws (Putra, 2016). Indonesian authorities do not respond to cyberbullying cases unless a formal, official complaint is lodged (Rahmadi, 2014). Furthermore, in order for cyberbullying cases to be pursued, official complaints lodged by the victims must be collected by the law enforcement authorities (Cahyaningtyas, 2014; Police, 2016). Hence, what is required is extensive research to identify whether or not the messages constitute an act of cyberbullying in the Indonesian context via the content of the messages (Safaria et al., 2016) which might consist of name-calling, shaming or embarrassing images of the victim, threats of physical harm, and derogatory comments about ethnicity or religion (Akbar and Utari, 2015).

It should be noted that the results of this research contribute to the enrichment of previous literature in identifying the insulting words which characterise cyberbullying content, particularly in the Indonesian context. Previous literature reported that the characteristics of Indonesian cyberbullying messages consisted of name-calling, shaming or embarrassing images of the victim, threats of physical harm, and demeaning comments related to ethnicity or religion (Akbar and Utari, 2015). However, this research has extended the characteristics of Indonesian cyberbullying messages by recognising the insulting words that were often used to embarrass a person. The characteristics of Indonesian cyberbullying messages are as follows:

1. The use of insulting words was predominant in the cyberbullying messages.

2. Insulting words in cyberbullying messages associated with animal terms were prevalent, followed by negative characters and disabilities.
3. Local terms, or dialect words, were also used in cyberbullying messages, such as *goblok* (meaning stupid – often used within the Java region). Such terms are often used by a local person.

The findings of this research are somewhat similar to those of the study conducted by Kontostathis et al. (2013)where several query terms were used to detect cyberbullying messages from Formspring.me. For example, the word *stupid* had the same meaning as *goblok* or *bodoh* in the Indonesian language. However, the difference between this research and the research of Kontostathis et al. (2013) is that this research has successfully identified other Indonesian insulting terms used for cyberbullying that relate to animals and negative characters, for example, liars, slanderers, emotional persons and traitors.

When examining the detection of cyberbullying messages, there are at least three points relating to Indonesian cyberbullying that are necessary to consider:

1. Recognising the content of the message, whether they contained name-calling, shaming or embarrassing images of the victim, threats of physical harm, and derogatory comments about ethnicity or religion.
2. Recognising every word that was contained in a message. This means that the user should be aware of every message that they have received, either the message did or did not contain insulting words.
3. Recognising the patterns of insulting words in each message assuming that the messages contained some insulting words. By using the results in this research, identifying the patterns of insulting words in each message can be detected.

A final point in regard to finding the patterns of cyberbullying messages in the data set, is that the evidence from the experimental result and the data with the most cyberbullying content indicated that the data extracted from the Indonesian Twitter contained insulting words associated with animals, negative characters and disability such as *bangsat* (bastard), *anjing* (dog), *babi* (pig), *iblis* (devil) and *setan* (satan).

5.5 Practical Uses of Finding

To the best of the author's knowledge, no previous studies have investigated in depth whether cyberbullying messages can be detected through the identification of patterns of insulting words. This study has made an important contribution to the field by providing strong support for such a link. This suggests that future studies should examine not only the relationship between these insulting words, but more specifically, how the recognition of patterns can help to reduce the instances of cyberbullying. It would be of particular interest to establish whether the detection of patterns of insulting words in the messages would be able to prevent cyberbullying. As Parime and Suri (2014) have emphasised, one of the ways to prevent cyberbullying is by a detection process using machine learning although their research involved social networks such as Facebook, Twitter, Ask.FM, Instagram, and LinkedIn. It may therefore be relevant for cyberbullying studies to not only investigate the effect of cyberbullying on the victims' mental health and the prevention of cyberbullying, but also to use machine learning to detect cyberbullying messages by means of data mining techniques.

In order to prevent the spread of cyberbullying messages on social networks, an analysis of cyberbullying messages using data mining techniques is required. As Nahar et al. (2013) have noted, there is a vital need for further studies on cyberbullying in order to detect, prevent and mitigate this behaviour. Therefore, the application of

the machine learning approach to data mining is one possible means of preventing or at least mitigating the spread of cyberbullying messages, since machine learning can help to reduce the incidence of cyberbullying by automatically detecting messages that may be construed as instances of cyberbullying.

The six data mining techniques that have been implemented in this research merit consideration when undertaking the process of recognising patterns of insulting words, particularly in the Indonesian context. This is because the application of the cyberbullying detection and analysis process using the data mining techniques developed by this research resulted in an 80.37% detection rate of cyberbullying messages in the total data sample. Hence, both the analysis model used for detecting cyberbullying and the results of this research confirm the effectiveness of using our proposed model to detect cyberbullying messages in social networks, particularly in the Indonesian context.

From the research literature reviewed in Chapter 2, it appears that the management of cyberbullying cases is ineffectual as Indonesian authorities are not proactive; unless victims formally lodge a complaint with the appropriate authorities, nothing much is or can be done (Rahmadi, 2014). Furthermore, in order for cyberbullying cases to be further processed, official complaints from the victims must be collected and accepted by the law-enforcing authorities (Cahyaningtyas, 2014; Police, 2016). Additionally, even though the Indonesian rule of law number 11, 2008 article 27-29 pertains to cyberbullying issues (Indonesia, 2008), the law is still inadequate in handling cyberbullying cases, since it is only concerned about what cyberbullying is and who is doing the cyberbullying (Putra, 2016). This indicates that the current law does not fully cover all aspects of cyberbullying and is ineffective as a preventative measure.

Hence, with regards to the results of this research, the model process of analysing cyberbullying within the Indonesian context is expected to assist the public, the Indonesian government and the cyberbullying authorities in combating the distribution of cyberbullying messages in future cases, since the cyberbullying analysis model is able to automatically detect cyberbullying and does not have to rely on reports lodged by the victims or the public. The findings from the research in this thesis have the following implications:

1. In terms of the dissemination of the “stop cyberbullying” message, the insulting word patterns discovered in this research can be used as an information source for disseminating the early identification of cyberbullying messages within the Indonesian context and can thus assist in reducing the spread of cyberbullying.
2. The patterns found in the research results can offer users on social networks access to resources and information to gain an awareness when they receive cyberbullying messages containing insulting word patterns in the Indonesian context.
3. The patterns in the research results can be a source of primary data which can be mined in more depth in future research.

5.6 Analysis Data from Social Network Sites

In research literature discussing content analysis, the content of social networks sites such as Facebook, Twitter and YouTube has drawn the attention of researchers in recent years (Kossinets and Watts, 2006). This is due to the effectiveness of social networks as a channel of communication. In addition, social networks enable users to establish a virtual identity and network connection among friends and families,

allowing them to send direct messages including texts, images, and videos (Acemoglu et al., 2010). The facilities offered by social networks have enabled them to become a medium of communication and invective where users can slander and ridicule others, react to provocation, and engage in cyberbullying (Pang and Lee, 2008).

As stated in Chapter 2, content analysis of social networks has attracted the interest of researchers, particularly in analysing the process of mining opinion (Pang and Lee, 2008; Liu, 2012), the extraction of opinion (Lo and Potdar, 2009), the mining of sentiment content, the analysis of subjectivity, the analysis of affect, the study of emotion, and the mining of reviews (Aggarwal and Zhai, 2012). Content analysis is also an important tool used for classifying every form of message content (Riff et al., 2014). Thus, the application of data mining techniques for content analysis of social network data requires further investigation in order to develop this research field (Barbier and Liu, 2011; Aggarwal and Zhai, 2012; Han et al., 2012).

For the purposes of this research, the data used for content analysis was derived from tweets, and the identification and analysis of cyberbullying messages was conducted using data mining techniques. The results indicated that 80.37% of the total data (152,843 posts) were identified as cyberbullying messages, while 19.63% were detected as non-cyberbullying messages. This suggests that Twitter is a medium of social interaction that most users utilise to spread cyberbullying messages in preference to other social networks such as Facebook, Instagram, YouTube, Ask.Fm. and Tumblr.

This research has identified patterns of Indonesian insulting words which assist in determining whether or not Twitter messages, within the Indonesian context, are instances of cyberbullying. The process used to analyse data is important in order for results to be accurate. This research has produced meaningful results that can be

used for future research and that illustrate the offensive words users often used to spread harmful messages. The further development of the analysis process can be carried out by thoroughly exploring a range of data mining techniques

5.7 Extended Analysis Model using data Mining Techniques

Chapter 2 reviewed the literature relating to data analysis in which data mining techniques were applied to analyse cyberbullying data. These data mining techniques were explained in detail to illustrate how they can be used to analyse social issues such as cyberbullying. As Ott and Longnecker (2015) noted, the types of techniques and approaches applied in order to derive essential information from the data analysis process heavily depends on the variables to be analysed.

The data mining techniques that were applied in this research are:

1. Association rules techniques of FP-growth and cosine similarity were employed for pattern recognition of offensive Indonesian words often used to send cyberbullying messages.
2. Naïve Bayes, decision tree, and neural network were applied to detect messages from Twitter that could potentially be cyberbullying messages.

Although these six data mining techniques were the tools used for the analysis of data in this research, various other techniques were considered for the process of data analysis, such as the clustering technique applied for the labelling of the training data set.

5.7.1 The Indonesian Stem Dictionary

In order to find specific terms related to cyberbullying messages, stemming techniques are required to eliminate any suffixes and prefixes from the root words

(Frakes and Fox, 2003); this includes reducing the whole words to the root format of their morpheme words by eliminating all derivational and inflectional suffixes from the selected words (Smirnov, 2008). According to Alfred et al. (2007), the process of stemming involves the decomposition of words from their affixed formats to their root origin, which usually affects the quality of the outcome for further document retrieval purposes and data classification.

Leong et al. (2012) have mentioned that the stemming procedure generally involves the use of a dictionary. The dictionary contains the entire root words which can be referred to in order to determine whether or not the root words exist in the dictionary. A traditional algorithm browses through every recorded word in the dictionary regardless of whether or not the word contains affixes. However, this is a time-consuming process. Apart from this, assuming the root words in the dictionary do not contain suffixes, the stemmer may restore a word to its stemmed word inaccurately. As the application of the stemming procedure is intended to condense the derived word into its root composition, this will also decrease the number of words that are to be refined since the derived words might possibly originate from an identical root word.

Many types of stemming techniques such as those of Snowball, Porter, Lovins, German, Arabic, and Dictionary presented in Chapter 2 were used to explain and give a better understanding of stemming techniques suitable for this research. Therefore, a particular stemmer is essentially needed in order to enable the removal of the suffixes and prefixes and render each word into a root word.

The author of this research created a new stemming process specifically for offensive Indonesian words that are commonly used in cyberbullying messages. The Indonesian stem dictionary was created using the approach of Nazief and Adriani

(1996) which was explained in Chapter 2. Based on broad morphological rules, the Nazief and Adriani technique functions to cluster and encapsulate the procedures which pass and restrict affixes, prefixes, suffixes, infixes otherwise known as insertions and confixes, or a combination of prefixes and suffixes (Asian et al., 2005; Adriani et al., 2007). Adriani et al. (2007) have emphasised that the Nazief and Adriani approach was an effective stemming technique that was very accurate in converting some terms to their morphological roots.

Following the application of the stem dictionary of Indonesian insulting words, some of the offensive terms were transformed into their morphological roots. For example, the term *anjinglo* (you dog) was transformed into its root composition of *anjing* (dog). *Anj*.* was also transformed into its root format *anjing.* Most of the offensive words that occurred in messages were changed into their roots. According to Al-Kabi et al. (2011), stemming is implemented in decreasing the variance of word formats into base roots to enhance the effectiveness of filtering process. The effectiveness of stemming application as a filtering stage is to assist in the performance of data prediction within a high value of precision (Al-Kabi et al., 2011). Therefore, the offensive Indonesian words stemming that has been developed in the research of this thesis can be considered as an approach that produce precise results. This can be referred to the results of this thesis where classification of cyberbullying class achieved 80.37%. The research classification results signifies that the stemming approach developed in this thesis is robust and effective. Other advantages of using this approach can include easy access for further development. In addition, the fact that the algorithm controls the use of an auxiliary dictionary comprising of root words brings an advantage into the applied steps, so whether or not the stemming has attained a root word is determined.

On the other hand, the stem dictionary of Indonesian insulting words needs to be further developed because the stem focuses only on the number of insulting terms and the resulting density of these terms in the messages. Another limitation of the stem strategy is that offensive terms in local languages such as Javanese, Sundanese, and other regional languages are not well-covered in the stem dictionary. Therefore, offensive local terms need to be further covered in order for the dictionary to be more comprehensive and to reduce the potential errors when transforming words into their root composition because of the ambiguity of local languages.

Although, the stemmed offensive Indonesian words had several limitations when transformed into their root composition, following the generating procedure, the experimental results showed that some of the insulting words in their abbreviated format had prefixes and suffixes inserted, resulting in greater accuracy of the root words. The reason for this is that the stemming techniques do not involve removing or ignoring the abbreviation together with its prefixes and suffixes of insulting words. The stemming techniques are used solely to confirm the entire abbreviation, prefixes, and suffixes of insulting words, which are then transformed into their root. Therefore, the stemmed offensive Indonesian words illustrate a special stemming technique that is appropriate for the identification of offensive words associated with cyberbullying messages. Also, this stemming technique can easily be developed in future research, as it does not require advanced programming in order to develop the stem.

5.7.2 Finding Data Patterns

A well-designed analysis model greatly assisted in generating machine learning which yielded accurate results in this research. The research was designed so that an analysis model could be applied to discover patterns in the data set using association

rules techniques of FP-growth and cosine similarity sequentially. This produced significant results in deriving patterns precisely and efficiently (Mahoto et al., 2014). A well-designed parallel application of data techniques, for example, for finding hidden patterns of data in cyberbullying (Maheshwari et al., 2014), was used to apply actual theoretical knowledge to the activities undertaken in this research. The developed scientific analysis model was a significant component of the research process (Chapters 2, 3, and 4).

In the analysis process, using constraints as parameters to measure the significance and accuracy of the result is required. As Srikant et al. (1997) stated, the use of several data analysis strategies can dramatically reduce the execution time of generating an algorithm. In practice, several strategies can be used to identify infrequent items in the data set. Since the purpose of this research was to find patterns of Indonesian offensive words, several strategies were applied such as stem dictionary, tokenize, transforming case, stop words, n-gram, laplace, lift and conviction. By using these techniques, an outcome was achieved that was efficient in terms of time, and effective in terms of accuracy.

The results of the analysis indicated that this model was successful in finding patterns with a strong relationship between items. This means that the model is robust and has tremendous discovery potential in terms of the correlation of data. However, the model requires several processing steps, making the analysis process more time consuming. The analysis model has several capabilities and advantages that merit consideration. These are as follows:

1. The model yields more accurate results in comparison with models that apply a single technique because it has sequentially implemented association rules techniques, each of which has a specific function. FP-Growth is used to find

patterns of itemsets while using constraints of *minim_support* count without candidate generation (Han et al., 2012; Gu et al., 2015). The association rules involve four constraints: confidence, laplace, lift and conviction in finding special rules with a strong relationship between items. The cosine similarity is used to measure similarity between items based on the degree of proximity of the vector between them (Akthar and Hahne, 2012).

2. The accuracy of the results can be attributed to the fact that each process in the model was conducted using several parameters as the constraints. By using this model with the four constraints of confidence, laplace, lift and conviction, the strength of the relationship between items can be measured (Akthar and Hahne, 2012). Therefore, the interesting patterns of itemsets that have a strong relationship between itemsets can be discovered. As a result, from the research in this thesis, patterns of offensive Indonesian words in cyberbullying messages were illustrated. Confidence, laplace, lift and conviction were explained in Chapter 3.
3. The model can be used for more than just the analysis of general terms; it can analyse specific terms such as finding patterns of offensive terms in International languages from social networks.

On the other hand, the model also has several limitations as follows:

1. As the recorded data are numerous and large in size, generating an algorithm requires memory space to be larger in order for the data to fit in the storage. Hence, memory space can be a limitation to recording and storing data processes.

2. Obviously, data patterns of insulting terms discovered from this research are limited to the Indonesian context. Meanwhile, other languages have not been covered yet using the model generated by the author of this research.
3. The analysed data from this research were obtained from Twitter. There are many social networks sites such as YouTube, Facebook, Tumblr and Ask.FM that have not been used to collect data to identify insulting terms patterns like this research.
4. The research only analysed data of messages, in this case, tweets. Variables such as perpetrators, victims, time and location were not analysed to identify cyberbullying acts on social networks.

When examining the comparison between the advantages and disadvantages of the model, the limitations of the model can be noted. However, for discovering the correlation of contextual messages from Twitter, using the association rules techniques of FP-growth and cosine similarity sequentially is necessary. The purpose is to discover hidden associations in the textual contents and user-generated contextual features for understanding how the patterns of data will yield significant information. As Mahoto et al. (2014) noted, the association rules technique is considered to be a robust approach for discovering hidden associations and generating a taxonomy, according to the relevant relationships, for more detailed analysis of the content from the social networks data. Similarly, Mahoto et al. (2014) have also mentioned the benefits of discovering mutual connections in comprehending hidden relationships between the textual content and contextual characteristics in respect of user-generated content. The information collected from Twitter enables analysts to obtain an in-depth understanding of the meaning of user-generated content and the real-life implications. Moreover, Maheshwari et al. (2014) posited that the use

of association rules helps to reveal hidden patterns among data that seem to be unrelated but in fact help to identify cyberbullying victims and predators.

5.7.3 Development Framework of Identifying Cyberbullying Messages

In expanding an analysis model in its relevance to harmful messages identification on social networks, for instance, the detection of cyberbullying messages was accomplished by applying several classification techniques (Sanchez and Kumar, 2011). According to Dinakar et al. (2011), the classification techniques can be employed to discover the class of textual data considered to be cyberbullying or non-cyberbullying. Therefore, in expanding the analysis model of cyberbullying messages prediction from one of the social networks sites (Twitter) in the Indonesian context, classification techniques were implemented to discover hidden interesting class labels of messages of Indonesian context data collected from Twitter.

By implementing these classification techniques, both cyberbullying and non-cyberbullying messages can be discovered with high accuracy, precision and recall. As reported by Kasture (2015) and Nalini and Sheela (2015) the classification techniques were effective in predicting the dataset from Twitter as cyberbullying messages with a high precision value. Therefore, the analysis model developed using the classification techniques can be employed to cyberbullying identification in the Indonesian context on Twitter.

A well-designed analysis model that can achieve highly accurate results can assist in classification techniques for data class prediction. The analysis model using naïve Bayes, decision tree, and neural network techniques can achieve accurate data prediction (Xhemali et al., 2009). Sanchez and Kumar (2011) found that naïve Bayes was an effective technique in predicting cyberbullying messages from Twitter with a

high accuracy value of prediction. Nandhini and Sheeba (2015) research also found that the application of naïve Bayes in identifying cyberbullying messages from social networks had a high interval confidence of validation. Meanwhile, Reynolds et al. (2011) established that the decision tree techniques, specifically, J48 classifier detected cyberbullying or non-cyberbullying messages. Moreover, Dinakar et al. (2011) found that J48 classifier was able to identify cyberbullying content on social networks with a high accuracy value. Of the other classification techniques, neural network can identify sentimental messages from short text through combining character-level, word-level, and sentence-level, according to dos Santos and Gatti (2014). Ghiassi et al. (2013) reported that the neural network technique had a high level of performance in classifying sentimental tweets from Twitter. Therefore, the combination of the three classification techniques for analysing Indonesian cyberbullying messages was appropriate in terms of achieving high accuracy for the prediction results.

Based on the experimental results of the research of this thesis using naïve Bayes, decision tree, and neural network techniques, 152,843 data items were recorded in the database and were successfully allocated to two classes. 122,842 data items were allocated to the cyberbullying class, and 30,001 data items to the non-cyberbullying class. The application of the three classification techniques produced accurate labels, with the greater number of data items being in the cyberbullying class.

Prior to developing an analysis model for identifying cyberbullying messages, an established analysis model for the training set is necessary for classification techniques, in order to accurately label the data classes. The labelling of the training set influences the results of data prediction. When the training set data labelling is accurate, this indicates that the identification of items in the main data set will be

accurate. Therefore, the validation of data labels is necessary in order to achieve accurate data prediction. The process of labelling unsupervised data from the training set was explained in Chapter 4.

The process of labelling unsupervised data using *k*-medoids and IF-THEN-ELSE based on the number of insulting words occurring and the number or density of offensive words can be described as follows:

1. Selecting the number of insulting words and the density (number) of insulting words.
2. The cleaning process that involves tokenizing, transformed case, stop words, and stem dictionary.
3. Grouping data by using *k*-medoids clustering techniques.
4. Labelling through validation using the patterns of offensive Indonesian words and Kamus Besar Bahasa Indonesia (A Great Dictionary of the Indonesian Language).

Although, the model for the labelling process of the training set is more complex and time consuming, the advantages of this model are as follows:

1. The results were more accurate in comparison to the single process of labelling, because result validation processes were involved.
2. The results were confirmed by using the Indonesian insulting word patterns resulting from generating the association rules in order to find frequent patterns as described in Chapter 3 and using Kamus Besar Bahasa Indonesia to confirm the actual meaning of the word.
3. The accuracy of the results was ensured by applying certain parameters as constraints to every process in the model. The model successfully found two classes of data which were categorised as cyberbullying and non-cyberbullying classes.

Han et al. (2012) stated that a training set is needed as a learning model when the process of prediction is initiated. Therefore, in the development of the learning process for the purpose of identifying cyberbullying messages, the model result from the training set was produced.

The analysis model has several advantages, as follows:

1. Accuracy. To gain high accuracy in detecting cyberbullying messages especially in the Indonesian context, the analysis model displayed in Figure 5. which used accuracy, precision and recall as parameters to identify harmful messages which could possibly be cyberbullying messages can be applied.
2. Specific. Based on the results, the analysis model is an effective means of identifying specific terms in textual documents as it was able to discover patterns of offensive words in messages.
3. Early detection. The model can help to detect early harmful messages that potentially are instances of cyberbullying as it focuses on identifying the patterns of swear words.

Although, the analysis model has some advantages, it has the following two disadvantages:

1. As the volume of data increases, a larger memory space is required for better performance in the computational process.
2. If the categorical variable has a classification, however that was not noticed in training data set, then the analysis model will designate the value of 0 for its data probability, thus being incapable of creating a prediction.

Even though the analysis model has the limitations of requiring memory space and the completeness of data classes in the training data set, nevertheless it can be used to accurately identify cyberbullying messages sent via social networks.

5.8 Comparison of Cyberbullying Messages Identification

A previous research by Al-Kabi et al. (2011) have focused on Arabic as the main language to derive the roots of the obtained words from a total of 1100 documents. Al-Kabi et al. (2011) have expressed the uniqueness of the Arabic language that distinguishes itself from other non-Arabic languages due to the complex and rich nature of the language. Consisting of twenty-eight letters and written from the right side to left, Arabic comprises of two genders: masculine and feminine, and three number groups: singular, dual and plural that are assigned in nouns, adjectives, pronouns and verbs (Al-Kabi et al., 2011). Therefore, this contributes to the careful decision in applying the appropriate stemming technique, which has been identified as the Sheeren Khojar's stemming. Similar to the research of this thesis, the collected document undergone a pre-processing stage which includes the removal of symbols, punctuation marks (:,"), numbers (0-9), and stop words such as *from* (من), *to* (الى), *when* (متى), *over* (على). Once cleaned, the collected texts are then stemmed by detaching the longest suffix and longest prefix as according to the function of Khoja's stemmer, for instance, the word (الشارب) will be (شارب) after removing the (ال) prefix from it, and the word (شاربان) will be (شارب) after removing the suffix (ان) from it. The remaining stemmed words are matched with verbal and noun patterns in order to obtain the roots. Particularly, the fact that the Indonesian language has assimilated foreign words and phrases as a result of shared histories with Arabic and European influences has to be taken into a consideration (Adriani et al., 2007). This has had an impact on Indonesian prefixes and suffixes, either retained as original forms or transliterated. A characteristic of Indonesian language is that accented characters are transliterated without accents, such as déjà vu and naïve adapted to deja vu and naive (Adriani et al., 2007). Affixes

are more extensively emphasised in Indonesian language, which possibly requires meticulous planning of rules within the stemming process. The complexity of this language is also associated with infixes (insertions) and confixes, or commonly known as circumfixes, for example from the base word *perintah* (rule, order) to be embedded by affixes into *perintahnya* (their rule), *diperintah* (to be ruled), and *pemerintah* (ruler). This signifies that the application of the stemming approach on Indonesian language has its complexities, and hence, requires an enhanced stemming technique. As shaped by this requirement of an appropriate stemming technique, the research of this thesis has developed a stem (dictionary) of Indonesian insulting words as according to the Indonesian stem dictionary technique, proposed by Nazief and Adriani (1996). As mentioned previously, the Nazief and Adriani's approach implemented various morphological rules by eliminating the prefix, suffix, infix, and group of prefixes and suffixes into root words (Asian et al., 2005; Adriani et al., 2007).

In 2012, Chen et al.'s (2012) research on online safety among the adolescent community also acknowledged the unstructured and informal nature of cyberbullying textual content in the English context and developed a Lexical Syntactic Feature-based language architecture for detecting offensive language, and user patterns of posting. Their research noted the structural use of sentences in the offensive messages obtained, where a heavy emphasis of punctuation and uppercase letters subjectified emotions and speaking volume, similar to other cyberbullying research within the English context. Be that as it may, the research of Chen et al. (2012) took users' patterns of posting as an additional factor in cyberbullying detection, where the time length of users' conversation history, despite containing many offensive words, indicates the validity of offensiveness of a user. This feature may be able to distinguish regular offensive users from occasional users. Chen et al. (2012) have also analysed

the characteristics of content established in cyberbullying messages. While their research focused on an English language community, it recognised a large proportion of possible multi-background English speakers. They identified more content-specific features, which included: race, religion, violence, sexual orientation, clothes, accent, appearance, intelligence, and special needs or disabilities. Noting the comparison of cyberbullying content characteristics with the research of this thesis, special needs or disability shares a common ground as cyberbullying content in both the English language and Indonesian language. As previously mentioned, this thesis identified insulting words related to animals, stupidity or psychological needs, disability and general as characteristics of cyberbullying content on the basis of Indonesian context. This emphasises that the research of this thesis focuses specifically on Indonesia, as a country with Indonesian speakers.

In connection to the analysis of English language, Kontostathis et al. (2013) have also studied cyberbullying detection of social media text. During indexing the corpus of the insulting words, they converted the text characters into lowercase format and eliminated numeric and special symbols. On the basis of English language, capital letters are used to replicate a raised voice in direct conversation, including the use of emoticons to express one's feelings (Kontostathis et al., 2013). Apart from the punctuations and alpha-numeric conversion, Kontostathis et al. (2013) categorised the contents of the English insulting terms into common-place terms (*shithead*), esoteric terms (*unclefucker*) and hyphenated terms (*ass-hole*). However, the hyphenated terms were often separated into two words in considering the weighting system of precision and recall. This presents a particular characteristic of cyberbullying terms in the English context, compared to Indonesian insulting words that are lacking of

hyphenated terms. As identified in this thesis, the insulting words in the Indonesian context are often in format of singular terms, such as *anjing* (dog) and *goblok* (stupid).

Furthermore, a prior research by Cao et al. (2014) have explored the use of traffic sentiment analysis (TSA) in examining and analysing the collected Web-based data of sentiments from the traffic of Chinese websites; Sina Weibo, Tencent Weibo, Tianya and autohome. The selected websites are similar to the social network of Twitter with a restriction of 140 characters per post, marking as a similarity of the chosen social media of Twitter in the research of this thesis. The aim of their proposed algorithm is to achieve a method for a more safe and efficient informational exchange through identifying the advantages and disadvantages of rule-based and learning-based approaches. Cao et al. (2014) have come to a decision in opting for a rule-based approach due to its independency of the size of the document and the unchanging syntax rule of the language despite the variety of expressive features of many users, which is suitable for analysing the Chinese language. As explained by Cao et al. (2014), the linguistic characteristics of the Chinese language are present with obstacles for analysis procedure. Taking into an account that the Chinese language do not have spaces to separate words in sentences and its common use of numerous adverbs; 更 ‘more’ and 最 ‘most’, this language poses subtlety and ambiguity in sentences. As compared to the language chosen to be analysed in the research of this thesis, the Indonesian language has prefixes, suffixes, infixes (insertions) and confixes (a mix of prefixes and suffixes), which is similar to the English language. For instance, the root word "makan" (to eat) is added with a suffix of "-lah" to convey "makanlah" (please eat). A prefix of "mem-" and "per-" can also be added to a base word of "jelek" (bad) to create "memperjelek" (the action of ill-favouring). Due to this nature of the Indonesian language, the approach of Nazief and Adriani are

adopted and expanded as a language stemmer to reduce words into their base format in the research of this thesis.

The research by Cao et al. (2014) have designed the framework of their proposed TSA method involving stages of collecting Web-based data, pre-processing, extracting subjects and objects, identifying features of sentiments, computing and classifying sentiment data, evaluating and feedback. The stage that is worthy of being observed and investigated is the pre-processing stage, which can be compared with the pre-processing methodology from the research of this thesis. As Cao et al. (2014) have based their study on the Chinese language, their initial filtering phase is built based on the linguistic features of the chosen language, comprising of segmenting texts, such as an abbreviation of "中石油" is segregated as "中/j, 石油/n". Additionally, the filtering process involves labelling of words and replacing the synonymous expressions. In detail, the synonymous expressions in the Chinese language can be shown in an example of the use of "d", symbolised by the character "顶" (support). The use of this feature can influence the precision of the later processing stages hence, it is essential to decrease the complexity. Thus, with the assistance of Chinese Lexical Analysis System 3, Cao et al. (2014) are able to accomplish the tasks of the filtering phase. On another hand, the pre-processing stage of the research of this thesis consists steps of tokenise, transform case, stop words, stem dictionary and n-grams to eliminate abbreviations, punctuations marks and emoticons. Following the data retrieval from the repository in Rapid Miner by the retrieve operator, the data are tokenised, where the textual data are separated into words, phrases, symbols and other related features. Then, the next step is the implementation of transform case, in which the characters are respectively converted into lower and upper cases where appropriate. The prepositions, articles and pronouns, or referred as stop words are

also removed from the data. Once accomplished, the data are stemmed by using the developed stem dictionary of the Indonesian insulting words from the research of this thesis, based on the adoption of Nazief and Adriani's approach (1996). The developed stemming is to reduce the words into their stems, base or root formats. The final stage is the application of N-grams where this technique functions to produce term n-Grams of tokens in a file.

Moreover, in another development of cyberbullying detection research through the machine learning techniques, Huang et al. (2014) have implemented integrated textual and social network features that can boost the accuracy of the detection of cyberbullying messages. By analysing the social network structures among users and extracting features of the number of friends, network embeddedness, and relationship centrality, detection of cyberbullying can progress. However, the issue of the accuracy of cyberbullying message detection still remains. Huang et al. (2014) found that detecting cyberbullying by textual and social networks features still requires further development to achieve accuracy of detection. For this reason, using the enhancement of the stemming techniques and the identified patterns of insulting words led to a high accuracy of cyberbullying message detection, as shown in the results of this thesis where a total of 152,843 data, around 80.37% were detected as cyberbullying messages. This indicates that the developed frameworks from the research in this thesis successfully facilitated the detection of cyberbullying messages. Although the detection of cyberbullying techniques in this thesis focused on Indonesian text, the stemming techniques and the identification approach of the insulting word patterns developed could be implemented in other common languages, such as English. Therefore, the methods developed in this research have the flexibility to be adopted in other languages.

Another previous research based on Japanese language by Hatakeyama et al. (2016) has experimented the approach on acquiring seed words to enhance the conduct of methods for cyberbullying detection. Valuing the approach of Web mining technique, Hatakeyama et al. (2016) have introduced developments to increase the adopted methods' performances to acquire the highest results of precision and recall. This is achieved by advocating corpus-based approaches or approaches that offer consistent updates of relevant seed words following to the reason cause of rapid dynamics of Websites content. In similarity to the research of this thesis, this thesis has approached a calculative method of precision and recall as parameters to predict outcomes of Twitter posts dataset into classifications of cyberbullying class and non-cyberbullying class. These two parameters employed in the research of this thesis are for obtaining great outcomes of predictive label classes of data set, while also aiding in the selection of suitable prediction of data classes. For good measures, precision and recall are executed to evaluate the performance of the classifier model with regard to speed, robustness and scalability. In this case, the classifications that Hatakeyama et al. (2016) have arranged using a large Japanese database of electronic bulletin board system (BBS) entries are harmful basic word candidates and non-harmful basic word candidates. The basic word candidates are calculated for their Semantic Orientation on Point Mutual Information by Information Retrieval (SO-PMI-IR), or abbreviated as SO value. The greater SO value, the higher potential for the basic words to attain the perfect relevance score among themselves. This signifies the severity of the harmfulness of a word in the document, thus the categorisation of harmful or non-harmful basic words. The comparisons that can be observed in data categorisation between the research of Hatekayama et al. (2016) and the research of this thesis are the approaches that were used and the sequential process that was

performed. The research of this thesis has implemented the approach of clustering algorithm to aid in classifying messages by the identification of characteristics and features from the data collected. In technical terms, as the messages are distributed and allocated into groups, the clustering technique assists in advancing the process of grouping based on the similarities of objects. Once accomplished, the sequential series in progressing to labelling data involves the prior cleaned data undergone through the proposed validations of *If-then-else* mathematical expressions; the Indonesian insulting word patterns from prior data analysis stage by association rules, and the two Indonesian dictionaries to produce two classes: cyberbullying and non-cyberbullying.

Additionally, Hatakeyama et al. (2016) have managed the collected basic words by filtering and verifying whether the words are considered to be harmful or non-harmful according to the judgements of three human annotators. This was achieved by their development of the labelling method from Ishizaka et al. (2011). An initial result of 76 harmful word candidates were agreed upon all three annotators and calculated to have the strongest relevance to harmful BBS database. These words are implemented as a base for automatic selection. A second filtering procedure was then performed using the 17 harmful basic words from the first filtering, such as *bakasayo* (stupid), *fun-nyou* (dung), *majikimo* (seriously gross) and 17 non-harmful basic words of *shiborikomi* (narrowing), *kaiage* (purchase), *furikae* (transfer) and so on. The addition of non-harmful words serves a purpose to authenticate if the harmful basic words produce a bias in scoring.

To compare, the research of this thesis has acquired data of Indonesian insulting words from Twitter posts, or ‘tweets’ that has undergone filtering and cleaning tasks through the developed stemming technique in this thesis. The stemming

technique exists to remove suffixes and prefixes of collected words, including derivational and inflectional suffixes. This is accomplished by means of decomposing words from their affixed formats to their root format. Simple illustrations are shown by the term *anjinglo* (you dog) converted into its root origin of *anjing* (dog), and the suffix "an" was detached from the word *bangsatan*, which has no definite meaning in order to produce the word of *bangsat* (bastard). By extending the capability of Nazief and Adriani (1996) stemming approach to Indonesian stem of insulting words, the research of this thesis is able to expand the implementation of insulting words that were cleaned from their abbreviation, incorrect spellings and punctuation marks. However, unlike Hatakeyama et al. (2016)'s use of human annotators to establish judgement whether the acquired words are harmful or non-harmful, this research has created a process analysis model to classify the insulting words to be either into cyberbullying or non-cyberbullying categories, as previously mentioned. This was achieved by continuing the processing stage using the cleaned insulting words which to be headed for labelling process. In this stage, the mathematical expressions of *If-Then-Else* are implemented to identify the words into cyberbullying class or non-cyberbullying class. This also cannot be accomplished without the use of data patterns of relationships between insulting words that have been previously obtained from generating the association rules techniques. The next stage of data processing leads to the use of one of the clustering techniques which is more robust and effective in terms of coping noise and outliers, K-medoids (Han et al., 2012). K-medoids assists in the division and grouping of data training set into outcomes of two clusters, cyberbullying and non-cyberbullying. To establish a stronghold in the labelling outcomes, the data training set are then verified for their meanings through the application of the *Kamus Besar Bahasa Indonesia* (A Great Dictionary of the Indonesian Language) (Setiawan, 2016)

and *Kamus online* (The Online Indonesian Dictionary) (Indahnesia, 2016). Aside from confirming the translations of each insulting word, the use of two Indonesian dictionaries are also to reduce the ambiguity in the meaning of the words. This is similar to Hatakeyama et al. (2016)'s intention to reduce the bias in scoring, however, their approach involves the addition of non-harmful words in the processing stage.

5.9 Contributions

The analysis model developed by the author of this research contributes to enriching knowledge of social issues analysis and the practical knowledge of the learning process by using data mining techniques. Since the focus of this research was on identifying cyberbullying messages, the development of an analysis model using data mining techniques to identify harmful messages from social networks is a significant contribution as it refines the method of accurately detecting activities related to cyberbullying messages on social networks. Thus, it is anticipated that this research will inspire further studies to create a more effective and efficient analysis model for the detection of cyberbullying messages, especially in the Indonesian context.

Although the analysis model required memory space and the completeness of data classes in the training data set, the results indicated that this research can enhance the understanding of how to discover data patterns from textual documents and to detect cyberbullying messages accurately. For example, this research was able to develop an analysis model to discover insulting word patterns using association rules techniques of FP-growth and cosine similarity. This research has also advanced an analysis model to accurately discover cyberbullying messages, using data mining classification techniques.

This advancement of the analysis model offers a technique to prevent the spread of cyberbullying messages by generating a machine learning model using association rules techniques of FP-growth and cosine similarity sequentially to identify the characteristics of word patterns that have the potential to be cyberbullying messages. Moreover, the machine learning approach using naïve Bayes, decision tree and neural network to detect early cyberbullying messages aids in preventing the distribution of cyberbullying on social networks. Hence, using the data mining approach, the proposed analysis model makes a significant contribution to the analysis of cyberbullying messages emerging from social networks.

5.10 Strengths and Limitations

The rationale for this research is the broader applicability of the analysis model to social studies using data mining techniques. The identification of characteristics relating to cyberbullying messages has been advanced by finding the patterns of insulting words which can be construed as swear words imbedded in cyberbullying messages. By discovering the insulting word patterns, a complete picture of the characteristics of cyberbullying messages emerged. The results of a methodological study in terms of the saturation point in cyberbullying analysis have shown that, in general, using three classification techniques sequentially is more complicated and requires additional time, although results are more accurate and precise. The author of this research has explored and developed the analysis model explained in Chapters 3 and 4. The methodological exploration in Chapters 3 and 4 adequately describes the scientific research analysis of social issues. The identified number of insulting words and their density can be a starting point for identifying patterns in cyberbullying messages and can help with the design and construction of the analysis model. It is

reasonable that cyberbullying messages can be identified through the detection of texts or terms that often occur in cyberbullying messages. For example, some Indonesian swear words associated with cyberbullying are *bangsat* (bastard), *bajingan* (scoundrel), *anjing* (dog), *babi* (pig), *iblis* (devil), and *setan* (satan). The result of the insulting word patterns in this research shows a strong correlation between terms that are associated with cyberbullying messages that can be measured using minimum support of threshold, confidence, lift, laplace and conviction as described in Chapter 3, and accuracy, precision and recall for detecting cyberbullying messages explored in Chapter 4.

When examining the experimental results of sequentially using the association rules techniques, at least three points must be considered:

1. Recognising the number of insulting words and the density of the number of insulting words
2. Identifying the patterns of insulting words that can be considered as swear words in cyberbullying messages
3. Finding the strongest relationships between the insulting words.

The strength of the analysis model constructed through the use of association rules techniques of FP-growth and cosine similarity is that it can be used to discover the frequent patterns of terms in messages such as the offensive Indonesian words.

The other considerations when examining the experimental result via the sequential application of three techniques for classification purposes are as follows:

1. Accuracy of closeness in respect to the measured value when machine learning was used to predict messages which contained various offensive words

2. Precision of the result when using machine learning to detect cyberbullying messages.

The appropriately designed analysis model in respect to classification was also employed as a strong measurement in this study in order to detect cyberbullying messages with accuracy and precision. Therefore, the complexity of using some measurements can only be properly analysed and understood when the development analysis model is investigated from different perspectives by using various research tools such as data mining techniques.

This research focused only on offensive Indonesian words, and therefore the offensive words common in some local areas in Indonesia have yet to be covered. Therefore, future research for cyberbullying detection in the Indonesian context needs to consider the expansion of geographical areas in respect to the identification of insulting words. By using machine learning, many new offensive words on Twitter that may potentially be included in cyberbullying messages can be detected (Raisi and Huang, 2016). Numerous offensive words from different local areas in Indonesia can also be problematic when examining the stem dictionary of offensive Indonesian words. Nevertheless, this dictionary could be developed to include various local dialects in Indonesia or languages similar to Indonesian, such as the Malay language. The findings presented in this thesis might be limited by the specific selection of offensive terms rather than general terms. This needs to be taken into consideration when the results are applied to situations where general terms are associated with cyberbullying.

Another potential variable that must be considered is only the data from Twitter messages analysed in the study are presented in this thesis. Therefore, future research could include data analysis from Twitter in various international languages,

which have the potential to be cyberbullying. As a result, the conclusions drawn from this thesis are applicable to other researchers, including those who are analysing Twitter messages, especially in the Indonesian context.

5.11 Conclusion

This chapter has explained both the results and the design development process of the analysis model used to find insulting word patterns and detect cyberbullying messages. This research contributes to the knowledge of cyberbullying, the development of social analysis and the extended design process of social analysis issues using data mining techniques. Moreover, these results provide important data that can be used as a resource in further research in respect to the combination of data mining techniques and the analysis model of social issues.

A well-designed analysis model plays a major role in ensuring accuracy, efficiency, and precision when detecting cyberbullying. Few studies have focused exclusively on cyberbullying messages in respect to the Indonesian context. Detecting cyberbullying messages in the Indonesian context is challenging and they need to be mined more deeply using various tools such as data mining techniques. This research demonstrates that the development of a well-designed analysis model both in finding insulting patterns using association rules techniques and in detecting cyberbullying through classification model has successfully produced results with accuracy, efficiency, and precision.

This research has discovered contents associated with cyberbullying using the designed analysis model for detecting cyberbullying messages from Twitter that can be advanced further. This advancement would help in the development of effective techniques, algorithms and processes that can be used to efficiently detect

cyberbullying messages in Indonesian social media. The search for insulting word patterns by using association rules techniques of FP-growth and cosine similarity is the first stage in the process of detecting cyberbullying. Naïve Bayes, decision tree and neural network were applied to the patterns in the next stage of detecting cyberbullying messages sent via Indonesian social media. The conclusions drawn from the research will be presented and discussed in the next chapter.

Chapter 6
Conclusion

As the Internet is no the new medium of communication, people use this medium to express their opinions and ideas in the form of texts, images, and video and often convey messages and opinions through social networks such as Facebook, Twitter, YouTube, and so on. These social networks allow users to make positive, neutral and negative comments when they upload their status. The positive comments can bring some benefits to users who have uploaded their status, such as finding new ideas, making new friends, and improving users' writing skills. Meanwhile, neutral comments are often lacking in opinion, and are labelled as objective (Pang and Lee, 2008). Naradhipa and Purwarianti (2012) have stated that neutral comments do not contain any sentiment sentences, and are usually initiated with greetings, news or quotes. However, negative comments have a negative impact on users who upload their status, such as social isolation and feeling unsafe, stress, physical harm, loss of self-esteem, feelings of shame and anxiety, and difficulties in concentrating and learning (Kowalski et al., 2008; Thomas et al., 2013).

The conclusions are presented in this chapter. Any recommendations of this thesis will be included in reference to the literature in Chapter 2. The chapter also identifies and explains the strengths and limitations of this study, and provides further suggestions for developing the research presented in this thesis. The chapter concludes with suggestions for future research on the application of data mining techniques for the analysis of cyberbullying messages.

This research has explored the means of finding cyberbullying patterns on social networks by identifying insulting words and detecting cyberbullying messages

within the Indonesian context by using machine learning in data mining techniques. As a tool for analysing social issues, data mining techniques that provide some statistical calculation can be used to find patterns and discover classes of data.

This research is important because it provides new insights into the use of data mining techniques in an analysis model for cyberbullying. After the research gap was identified, this cyberbullying analysis model was applied to recognise cyberbullying messages on social networks through the implementation of data mining techniques. Further research that builds upon these insights is suggested. Furthermore, this concluding chapter explains the major contributions of this research in regard to analysing cyberbullying messages on social networks through the application of various data mining techniques.

6.1 Reflection on the Research Questions

The research questions and the sub-questions were introduced in Chapter 1 to understand the scope of cyberbullying in Indonesia and find important characteristics in an effective identification of cyberbullying messages. The research questions addressed the elements of finding the Indonesian cyberbullying patterns on social networks and detecting cyberbullying messages.

The identified components from the research questions are (a) the content of the messages containing insulting words; (b) cleaning words using the Indonesian stemming dictionary; (c) relationships of insulting words in a message to others forming frequent patterns of dataset; and (d) the identification of cyberbullying using the insulting word patterns.

The approach of detecting message content, which consists of insulting words, emerged as an adaptation from previous research. This approach involved a process

of identifying the number and density of insulting words in messages (Reynolds et al., 2011). These elements were grouped into categories based on shared common definitions. For example, Indonesian insulting words referring to animals were grouped into one category, such as monyet (monkey), anjing (dog), and babi (pig).

The second component of the research questions that has been found is the cleaning process for data using the Indonesian stemming dictionary. This tool was adapted from the Nazief and Adriani (1996) technique of eliminating the prefix, suffix, infix, and combination of prefixes and suffixes into root words. Such an example can be removing the suffix “an” from the meaningless word of bangsatan into the root of bangsat that gives a meaning of bastard.

Another identified component of the research questions is the relationship of insulting words in a message to another message, which leads to the form of frequent patterns. In order to find the interesting relationships between insulting words, this thesis developed an analysis model. The association rules techniques of FP-growth and cosine similarity were employed along with the developed analysis model. Moreover, several parameters were also implemented that covered support_count, confidence, lift, laplace and conviction. The purpose was to find the strongest relationship between data. Support_count measured the frequency of itemsets appearing in the database and established a minimum threshold for the number of itemsets in the database. The confidence parameter indicated the frequency with which the rule was discovered to be true. In other words, confidence was used as the minimum threshold in regard to data transactions consisting of itemsets *A* and *B*. (Han et al., 2012). The function of Lift parameter is to calculate the frequency of itemsets *A* and *B* that have independently occurred (Brin et al., 1997). Meanwhile, the laplace

functions to estimate the confidence that takes support into a consideration, which can turn pessimistic if the support of itemset *A* declines (Azevedo and Jorge, 2007).

The use of insulting words in order to detect cyberbullying is one of the components of the research question. When the insulting word appeared in one message, that particular message was classified as a cyberbullying message. To achieve this process, the implementation of naïve Bayes, decision tree and neural network was undertaken. In addition, parameters of accuracy, precision and recall were employed to measure the extent to which the model was able to predict the data classes accurately. The accuracy of a classifier, say, on a given test set is the precise percentage of the tuples test set that is classified. This precision parameter is established as the restoration percentage, which later classified documents based on the query. Meanwhile, the recall is the percentage of each relevant document retrieved from the data set (Han et al., 2012). These parameters ensured the effectiveness of the analysis and the accuracy of the results.

6.2 Reflection on Cyberbullying

As modern communication technology develops, an inverse side which also emerges is cyberbullying, and the number of cyberbullying cases has increased sharply across the globe. One of the factors is the abundant number of open social network sites such as Twitter where access is offered to users for freedom of speech (Kowalski et al., 2012). Studies on cyberbullying detection focus on analysing the content of the messages (Kasture, 2015; Zhao et al., 2016). Hence, the analysis of message content associated with cyberbullying on social networks is an interesting research issue to be studied.

According to Zhao et al. (2016), in cyberbullying detection processes cyberbullying messages often comprise of insulting words. As cyberbullying messages consist of insulting words, these curse words are appropriate indicators that signify the presence of cyberbullying on social networks. In this thesis, the cyberbullying messages were discovered to contain insulting words within the Indonesian context and have been proven to indicate the presence of cyberbullying.

For example, cyberbullying messages contained insulting words within the Indonesian context such as "*Dasar anak gendut nggak pakai otak kayak babi, makanya lo goblok. makan tuh kotoran. Jangan belagu punya bapak jendral terus berlaku seenaknya sendiri.* (What a fat kid without brains like a pig, that's why you're stupid that eats faeces. Don't be too full of yourself and act whatever you want just because your dad is a General). This message is considered to be a cyberbullying message. Since, in the Indonesian context when referring to a person's attitude or habit to animals and people with disability, as previously mentioned in section 3.2.1, the use of offensive terms in messages can be categorised as impolite and inappropriate to all age groups. Safaria et al. (2016) asserted that name-calling and denigration of a person, as indicated in the examples above, are cyberbullying acts within the Indonesian context.

This research makes a contribution to the study of cyberbullying by identifying the characteristics of cyberbullying messages through discovering patterns of offensive terms. These patterns indicate the harmful content of the messages that are potentially instance of cyberbullying.

The characteristics of cyberbullying explained above provide a rich opportunity to examine common findings and emerging models regarding the identification of other characteristics of cyberbullying messages. This research suggests that there is another method that can be used to recognise the characteristics of cyberbullying

messages; that is, detecting the pattern of insulting words frequently appearing in the messages sent through social networks.

Although some previous research had been conducted to detect cyberbullying messages, such as the textual data categorisation into cyberbullying texts or non-cyberbullying texts (Dinakar et al., 2011), the analysis of sentiment messages (Sanchez and Kumar, 2011), the identification of cyberbullying messages from social networks (Reynolds et al., 2011; Kontostathis et al., 2013), and the detection of tweets as cyberbullying messages (Kasture, 2015; Nalini and Sheela, 2015), the analysis of cyberbullying tweets in the Indonesian context is an interesting research issue that needed to be more studied in depth (Safaria et al., 2016). Moreover, cyberbullying identification research using vocabulary approach in Indonesian context is an interesting focus for research (Riadi and Hariani, 2017).

Therefore, the research in this thesis has taken an interesting focus on detecting cyberbullying messages from Twitter using a content analysis approach based on the density of Indonesian insulting terms. To detect cyberbullying messages, this research employed association rules techniques of FP-growth and cosine similarity to discover the interesting insulting terms patterns associated with cyberbullying. Meanwhile, machine learning such as naïve Bayes, decision tree, and neural network were implemented to detect cyberbullying messages.

Based on experimental results from this thesis, the insulting term patterns that were identified indicate characteristics of Indonesian cyberbullying messages. In turn, these cyberbullying message characteristics signify the presence of cyberbullying. Moreover, the characteristics of cyberbullying messages based on the identified insulting term patterns were applied to detect messages on social networks as either

cyberbullying or non-cyberbullying. This research has been successful in identifying Indonesian insulting term patterns and detecting cyberbullying messages from tweets.

6.3 Research Contributions

This research was carried out to produce knowledge that is scientifically and practically useful in detecting cyberbullying messages with the use of data mining techniques. The contributions of this research are as follows:

- This research has extended the cyberbullying characteristics framework by identifying new characteristics. This was done by identifying the patterns of offensive words often used in cyberbullying. These extended characteristics were adjusted to suit the Indonesian context of cyberbullying. The extended characteristics of cyberbullying enhance an understanding of how to recognise cyberbullying messages sent via social networks in Indonesia.
- Based on this experimental research, the research has contributed to the development of analysis models for cyberbullying messages on social networks. This research has enhanced a new construct for the analysis model through the use of several parameters to measure the strength of the relationship between data sets. As a result, frequent patterns of data sets were discovered. This can be referred to the insulting term patterns analysed in this research.
- Another contribution of this research is the extended measurement of relationships between data through collaborating association rules techniques of FP-growth and cosine similarity. Several parameters were also implemented involving support_count, confidence, lift, laplace and conviction. This research

discovered an interesting relationship between data sets that were considered to be the data patterns in the database.

- A further contribution of this research is the measurement of data class prediction that has been advanced through combining naïve Bayes, decision tree and neural network. Furthermore, parameters of accuracy, precision and recall were used for the purpose of measuring the extent to which the model could predict data classes with precision. In this case, data was generated along with the implementation of parameters of accuracy, precision and recall, therefore, the outcomes were precisely classified into their own data classes. As a result, the implementation of the three classification techniques has effectively and accurately identified cyberbullying messages.

This research also makes contributions in terms of practical applications as follows:

- The thesis has provided data about patterns of cyberbullying messages in the Indonesian context. These patterns are meaningful since they were extracted from a huge number of messages collected in Twitter. The patterns were also processed structurally and were measurable. Meaningful information has been presented to the research communities who may have taken an interest in the analysis of Indonesian cyberbullying messages from social networks sites. The results in terms of Indonesian cyberbullying patterns can also be applied by public institutions and the research community to identify offensive messages that containing Indonesian insulting words.
- This thesis has contributed to practical knowledge of the analysis process in respect to cyberbullying messages in relation to the use of data mining techniques, including some parameters as thresholds. The analysis process was arranged into structures and conducted sequentially, thus can be taken

into an account academically. Also, the enhanced analysis process resulted in outcomes with high precision to reach successful cyberbullying detection. Therefore, this process can be adopted by other practitioners and researchers in various fields or disciplines.

- As a further contribution, the research has expanded the Indonesian stem dictionary that functions to deconstruct words into their morphological root format, in this particular case, the Indonesian insulting words. A similar approach to the stem dictionary developed in this research can be adopted by practitioners and researchers for various word topics.

6.4 Reflection on Research Gap

In the experimental process of this research, theories, implementations and practice were applied to analyse cyberbullying messages from social networks. In order to discover a research gap and to define the research problem, relevant literature and resources on cyberbullying were examined and reviewed. The review also enabled the author of this research to narrow the scope of the research so that the issue of cyberbullying could be investigated more thoroughly.

Cyberbullying has been investigated thoroughly in relation to the characteristics of cyberbullying messages. Those characteristics are the anonymity of the perpetrator (O'Brien and Moules, 2010), timing of the messages (Privitera and Campbell, 2009), the intention of sending the messages (Spears et al., 2009), and the content of the messages (Mishna et al., 2009). The characteristics of cyberbullying messages enable users to determine whether or not the messages are instances of cyberbullying. Even though users can utilise the four listed characteristics to recognise cyberbullying messages, another characteristic can be used. The message contents are interpreted

according to the definitions of each word as well as phrases and word combinations. Also, each word has its own accurate definition. On the other hand, word combinations are used to customise word definitions in order to generate positive or negative sentences. Hence, the solution is to provide an appropriate analysis model of the process for finding the patterns of message contents. The identification of patterns of terms in the messages can be a challenge in recognising the characteristics of cyberbullying messages. Based on experimental results, the research in this thesis found Indonesian insulting term patterns to be applied in recognising characteristics of cyberbullying messages. The identified Indonesian insulting term patterns can contribute to the characteristics of cyberbullying messages in detecting the presence of cyberbullying.

As Ellis (2013) and Hackett (2015) have pointed out, cyberbullying often occurs in social media on Facebook. Twitter, and Ask.FM. The research in this thesis has identified Twitter is a social media platform where cyberbullying is prevalent among Indonesian users. It was found that 80.37% of messages in this research database were cyberbullying, which is consistent with the findings of Yulianti (2014) and Gottfried (2012).

To detect cyberbullying messages, research by Reynolds et al. (2011) and Dinakar et al. (2011) involved the use of a decision tree. Meanwhile, Kontostathis et al. (2013) and Nahar (2014) have employed support vector machine techniques for the same purpose. The research in this thesis, however, has applied six data mining techniques to analyse cyberbullying messages in order to discover the interesting relationship between data sets. These association rules techniques were FP-growth and cosine similarity which were used to find insulting word patterns; naïve Bayes, decision tree and neural network were used to classify data.

Theories should be applied to practice. Since the analysis model was designed based on theory, the combination of using association rules techniques of FP-growth and cosine similarity in a sequential manner proved to be challenging. However, every obstacle had to be overcome in order to address the research gap and discover patterns of insulting words which indicate cyberbullying messages. In analysing cyberbullying messages, the greatest challenge appeared to be discovering the patterns of offensive words in the content of cyberbullying messages. The research findings of the insulting term patterns can be employed to detect whether messages are cyberbullying or non-cyberbullying. Moreover, this cyberbullying detection has implemented naïve Bayes, decision tree and neural network techniques along with the parameters of accuracy, precision and recall.

As previously mentioned, Indonesia is one several Asian countries that has a high prevalence in cyberbullying (Gottfried, 2012). However, there is limited research focusing on cyberbullying messages in Asian languages. Therefore, the detection of cyberbullying messages in the Indonesian context is a challenge to overcome and needs comprehensive research (Naradhipa and Purwarianti, 2012; Lunando and Purwarianti, 2013; Safaria, 2016). To fill the gap, the research in this thesis focused on cyberbullying messages in the Indonesian context and successfully identified insulting word patterns that ultimately resulted in cyberbullying detection with high precision.

6.5 Change in the Field

Generally, the recognition of cyberbullying messages has been developed during the last decade. In 2009, Spears et al. (2009) indicated that cyberbullying messages can be recognised by the intention of sending messages. Mishna et al.

(2009) also reported that cyberbullying can be detected through the content of messages, while Privitera and Campbell (2009) stated that cyberbullying can be identified through place and time in which direct contact and interaction with the victim was not required. In 2010, O'Brien and Moules (2010) reported that cyberbullying can be recognised by the fact that the perpetrator is often anonymous or is purposely hiding his/her personal identification. Therefore, this research is a new breakthrough in the recognition of cyberbullying messages by identifying patterns of insulting terms. Every message which bears some similarity to one of the patterns from these research findings can be identified as a harmful message which potentially may be a cyberbullying message. Moreover, the findings assist users to enrich their knowledge about the diversity of cyberbullying messages on social networks, especially in the Indonesian and Malay contexts.

On the other hand, over the last decade data mining has been applied in several different fields such as science (Chen and Liu, 2004), counter-terrorism (Thuraisingham, 2002; Thuraisingham, 2003; Thuraisingham, 2004; DeRosa, 2004), the security-liberty debate (Solove, 2008), prediction of the default of a client's credit card (Yeh and Lien, 2009), and the analysis of oriental medicine (Yang et al., 2013). Another field in which data mining techniques can be applied is that of cyberbullying via social networks. Previous research conducted analysis of negative and positive messages from social networks by applying data mining techniques (Go et al., 2009; Reynolds et al., 2011; Nahar et al., 2012; Kontostathis et al., 2013; Nahar, 2014). Hence, this research has contributed to the research field of socio-computing by extending the identification process of cyberbullying characteristics through insulting term patterns. Additional changes to the research field are the collaboration of suitable data mining techniques and the analysis model of social issues in identifying

Indonesian insulting term patterns as one of the cyberbullying characteristics. Therefore, cyberbullying detection in the Indonesian context was conducted.

6.6 Addressing Research Objectives

In Chapter 1, the four crucial objectives of this research were stated. These are: (1) providing theoretical knowledge and practice on how the patterns of insulting terms support the identification of the cyberbullying messages characteristics; (2) offering practical knowledge and lessons regarding how to develop the stemming process to find the morphological roots of several words; (3) developing an analysis model for the detection of cyberbullying messages by applying data mining techniques; and (4) contributing to early detection of cyberbullying messages on social networks.

The objectives of this research have been achieved as follows:

1. The first objective has been achieved by applying theories and practice in order to discover patterns of insulting terms by recognising the characteristics of cyberbullying messages. The theoretical and practical contributions were discussed in section 6.3 above. This research provides an insight into the various patterns of insulting terms that emerged, how they emerged, as well as the patterns that can assist in the detection of cyberbullying messages.
2. The second objective of developing a stem dictionary of Indonesian insulting terms has been successfully undertaken. This stem dictionary was used to eliminate prefixes and suffixes, and removed the abbreviations, leaving the morphological roots of the terms. This stem has successfully deconstructed several insulting words into their morphological roots.
3. The third objective has been accomplished as a successful analysis model was developed for the detection of cyberbullying messages (see Figures in sections

5.6.2 and 5.6.3). The experimental results indicate that the analysis model performs with a high level of accuracy in finding patterns of data and in detecting the classes of data labels.

4. The fourth objective has been achieved through the successful application of the analysis model used for finding the patterns of insulting words in the content of cyberbullying messages, and to detect cyberbullying messages containing insulting words. The patterns of insulting terms have provided a new insight on the characteristics of cyberbullying messages.

6.7 Suggestion for Future Research

The breakthrough of the application of data mining techniques to social issues analysis has contributed to the development of socio-computing research. Although the data mining techniques for the analysis of social issues such as cyberbullying messages have yet to be perfected, the use of other types of data mining techniques such as anomaly or outlier detection and regression analysis are needed to find diverse and hidden interesting information that has yet to be fully mined.

Therefore, in the direction of future research, other social network sites such as Facebook, YouTube, and Instagram can be mined for their comments to detect any cyberbullying posts. In this future work, the collected comments from social networks will be analysed firsthand in order to obtain the density of cyberbullying terms and to detect the presence of cyberbullying in the posts. This process can be undertaken by adopting a similar approach to the research in this thesis, or by developing a different analysis model with a range of data mining techniques, for instance, support vector machines (SVM). SVM is useful in categorising text and hypertext data as their implementation assists in decreasing the requirement for labelled training data sets.

Moreover, SVM can be used in classifying data into accurate classifications in order to identify features of cyberbullying patterns. To be specific, the data of cyberbullying terms will be mapped to a high-dimensional feature space in order for the data points to be classified with the use of SVM.

Furthermore, the author of this research has established several requirements that should be analysed, including other attributes that should be investigated more thoroughly in respect to cyberbullying databases.

In addition, the author suggests that future research focus on the broader context of insulting terms including those in local Indonesian dialects that will enrich the knowledge of the types of swearing words often used in cyberbullying messages. In the Indonesian language, there are various local dialects across the vast archipelago where most of the local terms are adopted into the national Indonesian language. This also leads to the numerous swear words derived from the local dialects. The different range of dialects are often used in cyberbullying messages on social networks. Therefore, swear words from various dialects that usually appear in cyberbullying messages need to be mined in more detail for complete cyberbullying detection.

Future research should also expand studies into a broader dimension includes other variables which may influence the distribution of cyberbullying messages. Variables that can be investigated may involve the locations where the cyberbullying messages are sent, the genre of the cyberbullying messages, and the motive behind the cyberbullying messages. Such variables can assist in identifying the behaviour patterns of sending cyberbullying messages.

6.5 Final Concluding Remarks

Cyberbullying messages can be recognised by the various contents of messages. The recognition process of cyberbullying messages with the well-developed analysis model will discover patterns of insulting terms based on the content of messages, since cyberbullying messages are comprised of sentences formed from a series of words. Messages on social networks contain much of information that needs to be explored in detail. Therefore, a well-designed analysis model can assist in identifying cyberbullying messages. The effective analysis process and the challenges in respect to accurate findings can be addressed employing a combination of several techniques used to recognise cyberbullying messages.

These findings are relevant to recognising the characteristics of cyberbullying messages that were discussed in Chapter 2. These are the anonymity of the perpetrator (O'Brien and Moules, 2010), the timing of messages (Privitera and Campbell, 2009), the intention of sending messages (Spears et al., 2009), and the content of messages (Mishna et al., 2009). This research has provided another characteristic of cyberbullying messages, in that the pattern of insulting terms in the content of messages can be used to identify whether or not the messages are cyberbullying. These patterns can assist users to cope with harmful messages that often contain swear words in order to bully victims. The findings of this research can also support the Indonesian Government in its aim of combating cyberbullying messages on social networks.

References

Abaya, S. A. (2012). Association Rule Mining based on Apriori algorithm in minimizing candidate generation. *International Journal of Scientific & Engineering Research,* 3**,** 1-4.

Acemoglu, D., Ozdaglar, A. & ParandehGheibi, A. (2010). Spread of (mis) information in social networks. *Games and Economic Behavior,* 70**,** 194-227.

Achananuparp, P., Hu, X. & Shen, X. (2008). The evaluation of sentence similarity measures. International Conference on Data Warehousing and Knowledge Discovery. Springer, 305-316.

Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M. & Williams, H. E. (2007). Stemming Indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP),* 6**,** 1-33.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O. & Passonneau, R. (2011). Sentiment analysis of twitter data. Proceedings of the workshop on languages in social media. Association for Computational Linguistics, 30-38.

Aggarwal, A., Xie, B., Vovsha, I., Rambow, O. & Passonneau, R. (2011). Sentiment analysis of twitter data. Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 30-38.

Aggarwal, C. C. & Han, J. (2014). *Frequent pattern mining*, Springer.

Aggarwal, C. C. & Zhai, C. (2012). *Mining text data*, Springer.

Agichtein, E., Castillo, C., Donato, D., Gionis, A. & Mishne, G. (2008). Finding high-quality content in social media. Proceedings of the 2008 international conference on web search and data mining. ACM, 183-194.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. & Verkamo, A. I. (1996). Fast Discovery of Association Rules. *Advances in knowledge discovery and data mining,* 12**,** 307-328.

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. 487-499.

Ahmad, F., Yusoff, M. & Sembok, T. M. (1996). Experiments with a stemming algorithm for Malay words. *Journal of the American Society for Information Science,* 47**,** 909-918.

Ahmad, T. & Doja, M. N. (2013). Opinion Mining Using Frequent Pattern Growth Method from Unstructured Text. International Symposium on Computational and Business Intelligence (ISCBI) IEEE, 92-95.

Akbar, M. A. & Utari, P. (2014). Cyberbullying pada media sosial.

Akbar, M. A. & Utari, P. (2015). Cyberbullying pada Media Social (Cyberbullyin on Social Media).

Akthar, F. & Hahne, C. (2012). RapidMiner 5 operator reference. *Rapid-I GmbH*.

Al-Kabi, M., Al-Shawakfa, E. & Alsmadi, I. (2011). The Effect of Stemming on Arabic Text Classification: An Empirical Study. *Information Retrieval Methods for Multidisciplinary Applications***,** 207.

Alfred, R., Paskaleva, E., Kazakov, D. & Bartlett, M. (2007). Hierarchical agglomerative clustering of English-Bulgarian parallel Corpora. Proceedings of International Conference of Recent Advances in Natural Languages Processing.

Alghamdi, A. S. A. (2011). Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data. *International Journal of Computer Science and Network Security,* 11**,** 7-16.

Angiulli, F. (2009). Outlier Detection Techniques for Data Mining.

Arifin, A. & Setiono, A. (2002). Classification of event news documents in Indonesian language using single pass clustering algorithm. Proceedings of the Seminar on Intelligent Technology and its Applications (SITIA)', Teknik Elektro, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia.

Asian, J., Williams, H. E. & Tahaghoghi, S. M. (2005). Stemming Indonesian. Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38. Australian Computer Society, Inc., 307-314.

Azevedo, P. J. & Jorge, A. M. (2007). Comparing rule measures for predictive association rules. European Conference on Machine Learning. Springer, 510-517.

Bakar, Z. A., Mohemad, R., Ahmad, A. & Deris, M. M. (2006). A comparative study for outlier detection techniques in data mining. 2006 IEEE conference on cybernetics and intelligent systems. IEEE, 1-6.

Barbier, G. & Liu, H. (2011). Data Mining in Social Media. *Social Network Data Analytics.* Springer.

Basterretxea, K., Tarela, J. & Del Campo, I. (2004). Approximation of sigmoid function and the derivative for hardware implementation of artificial neurons. *IEE Proceedings-Circuits, Devices and Systems,* 151**,** 18-24.

Bauman, S. (2015). Types of Cyberbullying. *Cyberbullying***,** 53-58.

Bauman, S., Cross, D. & Walker, J. (2012). *Principles of Cyberbullying Research: Definitions, Measures, and Methodology*, Routledge.

Bayardo, R. J., Ma, Y. & Srikant, R. (2007). Scaling up all pairs similarity search. Proceedings of the 16th international conference on World Wide Web. ACM, 131-140.

Beran, T. & Li, Q. (2005). Cyber-harassment: A study of a new method for an old behavior. *Journal of educational computing research,* 32**,** 265-277.

Beran, T. & Li, Q. (2008). The relationship between cyberbullying and school bullying. *The Journal of Student Wellbeing,* 1**,** 16-33.

Berry, M. J. & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*, John Wiley & Sons.

Best, P., Manktelow, R. & Taylor, B. (2014). Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review,* 41**,** 27-36.

Bhagat, S., Cormode, G. & Muthukrishnan, S. (2011). Node classification in social networks. *Social network data analytics.* Springer.

Bhattacharya, A. (1946). On a measure of divergence of two multinomial populations. *Sankhya. v7***,** 401-406.

Bifet, A. & Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. Discovery Science. Springer, 1-15.

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. (1984). *Classification and regression trees*, CRC press.

Brin, S., Motwani, R. & Silverstein, C. (1997). Beyond market baskets: Generalizing association rules to correlations. ACM SIGMOD Record. ACM, 265-276.

Cahyaningtyas, E. N. (2014). *Peranan Kepolisian Daerah Istimewa Yogyakarta Dalam Menanggulangi Tindakan Cyberbullying* Fakultas Ilmu Sosial.

Campbell, M. A. (2005). Cyber bullying: An old problem in a new guise? *Australian Journal of Guidance and Counselling,* 15**,** 68-76.

Cao, J. & Li, L. (2009). Cluster synchronization in an array of hybrid coupled neural networks with delay. *Neural Networks,* 22**,** 335-342.

Cao, J., Zeng, K., Wang, H., Cheng, J., Qiao, F., Wen, D. & Gao, Y. (2014). Web-based traffic sentiment analysis: Methods and applications. *IEEE transactions on Intelligent Transportation systems,* 15**,** 844-853.

Chai, S., Yang, J. & Cheng, Y. (2007). The research of improved Apriori algorithm for mining association rules. 2007 International Conference on Service Systems and Service Management. IEEE, 1-4.

Chapman, C. & Feit, E. M. (2015). Association Rules for Market Basket Analysis. *R for Marketing Research and Analytics.* Springer.

Chen, S. Y. & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science,* 30**,** 550-558.

Chen, Y., Zhou, Y., Zhu, S. & Xu, H. (2012). Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, 71-80.

Cheung, C. M., Chiu, P.-Y. & Lee, M. K. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior,* 27**,** 1337-1343.

Cooper, G. (2012). Data Mining and Social Media.

Cortis, K. & Handschuh, S. (2015). Analysis of cyberbullying tweets in trending world events. Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business. ACM, 7.

David, E. G. (2009). Doing research in the real world.

Defu, Z., Bo, W., Qihua, L. & Jiemin, Z. (2008). An Efficient Frequent Patterns Mining Algorithm Based on Apriori Algorithm and the FP-Tree Structure. Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on. 1099-1102.

DeRosa, M. (2004). *Data mining and data analysis for counterterrorism*, CSIS Press.

Diffen. (2016). *Facebook vs. Twitter* [Online]. Diffen LLC. Available: http://www.diffen.com/difference/Facebook_vs_Twitter [Accessed 3/11 2016].

Dinakar, K., Reichart, R. & Lieberman, H. (2011). Modeling the Detection of Textual Cyberbullying. The Social Mobile Web.

Dipa, A. (2016). *Most Youth Unaware of Cyberbullying* [Online]. Jakarta: Jakarta Post. Available: https://www.pressreader.com/indonesia/the-jakarta-post/20160903/281539405394354 [Accessed 20/2 2017].

Donegan, R. (2012). Bullying and cyberbullying: History, statistics, law, prevention and analysis. *The Elon Journal of Undergraduate Research in Communications,* 3**,** 33-42.

Dooley, J. J., Pyżalski, J. & Cross, D. (2009). Cyberbullying versus face-to-face bullying. *Zeitschrift für Psychologie/Journal of Psychology,* 217**,** 182-188.

dos Santos, C. N. & Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. COLING. 69-78.

Dredge, R., Gleeson, J. & de la Piedad Garcia, X. (2014). Cyberbullying in social networking sites: An adolescent victim's perspective. *Computers in Human Behavior,* 36**,** 13-20.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A. & Madden, M. (2015). Social media update 2014. *Pew Research Center,* 9.

Dumitru, C. & Maria, V. (2013). Advantages and Disadvantages of Using Neural Networks for Predictions. *Ovidius University Annals, Series Economic Sciences,* 13.

Ellis, M. (2013). *Cyber-bullying: 5.4m kids in UK are potential victims on Facebook, Twitter and Ask.fm* [Online]. The United Kingdom: Mirror. Available: http://www.mirror.co.uk/news/uk-news/cyber-bullying-facebook-twitter-askfm-2328238 [Accessed 3/11 2016].

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication,* 13**,** 210-230.

Erlandsson, F., Bródka, P., Borg, A. & Johnson, H. (2016). Finding Influential Users in Social Media Using Association Rule Learning. *Entropy,* 18**,** 164.

Febrianti, R. & Hartana, G. (2014). *Cyberbullying pada Mahasiswa Universitas Indonesia* [Online]. Jakarta: Fakultas Psikologi Universitas Indonesia. Available: http://lib.ui.ac.id/naskahringkas/2016-06/S56877-Rianda%20Febrianti [Accessed 28/3 2017].

Fitzgerald, B. (2012). Bullying On Twitter: Researchers Find 15,000 Bully-Related Tweets Sent Daily (STUDY). *The Huffington Post*, 08/02/2012.

Frakes, W. B. & Fox, C. J. (2003). Strength and similarity of affix removal stemming algorithms. ACM SIGIR Forum. ACM, 26-30.

Gadia, K. & Bhowmick, K. (2015). Parallel Text Mining in Multicore Systems Using FP-tree Algorithm. *Procedia Computer Science,* 45**,** 111-117.

Gao, Y. & Er, M. J. (2005). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy sets and systems,* 150**,** 331-350.

Gaur, P. (2013). Neural Networks in Data Mining. *International Journal of Electronics and Computer Science Engineering,* 1.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2014). *Bayesian data analysis*, Chapman & Hall/CRC Boca Raton, FL, USA.

Ghiassi, M., Skinner, J. & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications,* 40**,** 6266-6282.

Gillespie, A. A. (2006). Cyber-bullying and Harassment of Teenagers: The Legal Response. *Journal of social welfare & family law,* 28**,** 123-136.

Gindl, S., Weichselbraun, A. & Scharl, A. (2010). Cross-domain contextualisation of sentiment lexicons.

Go, A., Bhayani, R. & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford***,** 1-12.

Goebert, D., Else, I., Matsu, C., Chung-Do, J. & Chang, J. Y. (2011). The impact of cyberbullying on substance use and mental health in a multiethnic sample. *Maternal and child health journal,* 15**,** 1282-1286.

Gomaa, W. H. & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications,* 68.

Gottfried, K. (2012). *One in Ten (12%) Parents Online, Around the World Say Their Child Has Been Cyberbullied, 24% Say They Know of a Child Who Has Experienced Same in Their Community* [Online]. Montreal, Canada.: Ipsos. Available: http://www.ipsoshk.com/wp-content/uploads/2012/04/Cyberbullying-factum-AP.pdf [Accessed 1/4/2013 2013].

Gradinger, P., Strohmeier, D. & Spiel, C. (2009). Traditional bullying and cyberbullying: Identification of risk groups for adjustment problems. *Zeitschrift für Psychologie/Journal of Psychology,* 217**,** 205-213.

Gu, X.-F., Hou, X.-J., Ma, C.-X., Wang, A.-G., Zhang, H.-B., Wu, X.-H. & Wang, X.-M. (2015). Comparison and improvement of association rule mining algorithm. Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015 12th International Computer Conference on. IEEE, 383-386.

Gualdo, A. M. G., Hunter, S. C., Durkin, K., Arnaiz, P. & Maquilón, J. J. (2015). The emotional impact of cyberbullying: Differences in perceptions and experiences as a function of role. *Computers & Education,* 82**,** 228-235.

Hackett, L. (2015). Bullying Statistics in the UK. *The Annual Bullying Survey 2015.* The United Kingdom: Ditch the Label.

Hämäläinen, W. (2006). Descriptive and Predictive Modelling Techniques for Educational Technology.

Han, J., Kamber, M. & Pei, J. (2012). *Data mining: concepts and techniques,* Amsterdam, Elsevier.

Han, J. & Pei, J. (2000). Mining frequent patterns by pattern-growth: methodology and implications. *ACM SIGKDD explorations newsletter,* 2**,** 14-20.

Han, J., Pei, J. & Yin, Y. (2000). Mining frequent patterns without candidate generation. ACM Sigmod Record. ACM, 1-12.

Hatakeyama, S., Masui, F., Ptaszynski, M. & Yamamoto, K. (2016). Statistical analysis of automatic seed word acquisition to improve harmful expression extraction in cyberbullying detection. *International Journal of Engineering and Technology Innovation,* 6**,** 165-172.

Hoover, J. & Miller, J. S. (2016). Decision tree machine learning. Google Patents.

Hsieh, N.-C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert systems with applications,* 27**,** 623-633.

Hu, L.-Y., Hu, Y.-H., Tsai, C.-F., Wang, J.-S. & Huang, M.-W. (2016). Building an associative classifier with multiple minimum supports. *SpringerPlus,* 5**,** 528.

Hu, Y.-H. & Chen, Y.-L. (2006). Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. *Decision Support Systems,* 42**,** 1-24.

Huang, A. (2008). Similarity measures for text document clustering. Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand. 49-56.

Huang, Q., Singh, V. K. & Atrey, P. K. (2014). Cyber bullying detection using social and textual analysis. Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. ACM, 3-6.

Indahnesia. (2016). *Kamus Online (Indonesian Dictionary)* [Online]. Indahnesia.com. Available: http://www.kamus-online.com/?lang=en [Accessed 12 November 2016].

Indonesia, H. o. R. (2008). Undang-undang Republik Indonesia nomor 11 tahun 2008 tentang Informasi dan Transaksi Elektronik. Jakarta: Indonesian House of Representative.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters,* 31**,** 651-666.

Jong, H. N. (2016). RI children and adolescents prone to cyberbullying. *The Jakarta Post*, 28/07.

Kalucki, J. (2010). Twitter streaming API.

Kaman, C. (2007). *What country has the most bullies?* [Online]. Cambridge: Latitude News. Available: http://www.latitudenews.com/story/what-country-has-the-most-bullies/ [Accessed 29/4/2013 2013].

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*, John Wiley & Sons.

Kasture, A. S. (2015). *A predictive model to detect online cyberbullying.* Auckland University of Technology.

Kaufman, L. & Rousseeuw, P. (1987). *Clustering by means of medoids*, North-Holland.

Khashei, M., Hejazi, S. R. & Bijari, M. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems,* 159**,** 769-786.

Kim, S., Jeon, S., Kim, J., Park, Y.-H. & Yu, H. (2012). Finding Core Topics: Topic Extraction with Clustering on Tweet. Cloud and Green Computing (CGC), 2012 Second International Conference on. IEEE, 777-782.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Klinkenberg, R. (2013). *RapidMiner: Data mining use cases and business analytics applications*, Chapman and Hall/CRC.

Kontostathis, A., Edwards, L. & Leatherman, A. (2010). Text mining and cybercrime. *Text Mining: Applications and Theory. John Wiley & Sons, Ltd, Chichester, UK*.

Kontostathis, A., Reynolds, K., Garron, A. & Edwards, L. (2013). Detecting cyberbullying: query terms and techniques. Proceedings of the 5th annual acm web science conference. ACM, 195-204.

Kossinets, G. & Watts, D. J. (2006). Empirical analysis of an evolving social network. *science,* 311**,** 88-90.

Kotsiantis, S. B., Zaharakis, I. & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.

Kowalski, R., Limber, S. & Agatston, P. (2008). Cyber Bully: Bullying in the Digital Age. *Australia: Blackwell*.

Kowalski, R. M., Limber, S. P. & Agatston, P. W. (2012). *Cyberbullying: Bullying in the digital age*, John Wiley & Sons.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*, Sage.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 1097-1105.

Kros, J. F., Lin, M. & Brown, M. L. (2006). Effects of the neural network s-Sigmoid function on KDD in the presence of imprecise data. *Computers & operations research,* 33**,** 3136-3149.

Kryszkiewicz, M. (2014). The Cosine Similarity in Terms of the Euclidean Distance.

Langos, C. (2012). Cyberbullying: The challenge to define. *Cyberpsychology, Behavior, and Social Networking,* 15**,** 285-289.

Lapata, M. & Barzilay, R. (2005). Automatic evaluation of text coherence: Models and representations. IJCAI. 1085-1090.

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*, John Wiley & Sons.

Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A. & Choudhary, A. (2011). Twitter trending topic classification. Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. IEEE, 251-258.

Leers, W. (2011). *FP-Growth-powered association rule mining with support for constraints* [Online]. Available: http://wimleers.com/article/fp-growth-powered-association-rule-mining-with-support-for-constraints [Accessed 9 January 2017].

Lenhart, A. (2015). Teens, social media & technology overview 2015. *Pew Research Center,* 9.

Lenhart, A., Madden, M., Smith, A. & Macgill, A. (2009). Teens and social media: An overview. *Washington, DC: Pew Internet and American Life*.

Leong, L. C., Basri, S. & Alfred, R. (2012). Enhancing Malay stemming algorithm with background knowledge. Pacific Rim International Conference on Artificial Intelligence. Springer, 753-758.

Li, B. & Han, L. (2013). Distance weighted cosine similarity measure for text classification. International Conference on Intelligent Data Engineering and Automated Learning. Springer, 611-618.

Lippmann, R. P. (1989). Pattern classification using neural networks. *IEEE communications magazine,* 27**,** 47-50.

Lisboa, P. J., Vellido, A. & Edisbury, B. (2000). *Business applications of neural networks: the state-of-the-art of real-world applications*, World scientific.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies,* 5**,** 1-167.

Livingstone, S., Haddon, L., Görzig, A. & Ólafsson, K. (2010). Risks and safety on the internet: the perspective of European children: key findings from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries.

Lo, Y. W. & Potdar, V. (2009). A review of opinion mining and sentiment classification framework in social networks. 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies. Ieee, 396-401.

Lovinger, J. & Valova, I. (2015). Scrubbing the Web for Association Rules: An Application in Predictive Text. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, 439-442.

Lunando, E. & Purwarianti, A. (2013). Indonesian social media sentiment analysis with sarcasm detection. Advanced Computer Science and Information Systems (ICACSIS). IEEE, 195-198.

Maheshwari, A., Kharbanda, G. & Patel, H. (2014). Association Rules in Data Mining.

Mahoto, N. A., Shaikh, A. & Nizamani, S. (2014). Association Rule Mining in Social Network Data. *Communication Technologies, Information Security and Sustainable Development.* Springer.

Marius, P. e. & Anggoro, S. e. (2016). Profile Internet Indonesia 2015. Jakarta: Association of Indonesian Internet Service Providers (APJII).

Mather, P. & Tso, B. (2016). *Classification methods for remotely sensed data*, CRC press.

Mazer, J. P., Murphy, R. E. & Simonds, C. J. (2007). I'll see you on “Facebook”: The effects of computer-mediated teacher self-disclosure on student motivation, affective learning, and classroom climate. *Communication Education,* 56**,** 1-17.

McGuigan, N. (2013). Detecting Text Message Spam. *RapidMiner: Data Mining Use Cases and Business Analytics Applications***,** 199.

Mhashakhetri, U. & Sheikh, R. (2016). Frequent Pattern Mining Implementation on Social Network for Business Intelligence. *International Journal on Recent and Innovation Trends in Computing and Communication,* 4**,** 39-42.

Mihalcea, R., Corley, C. & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. AAAI. 775-780.

Miner, G. (2012). *Practical text mining and statistical analysis for non-structured text data applications*, Academic Press.

Mishna, F., Saini, M. & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review,* 31**,** 1222-1228.

Mishra, M. R. & Choubey, M. A. (2012). Discovery of frequent patterns from web log data by using FP-growth algorithm for web usage mining. *International Journal of Advanced Research in Computer Science and Software Engineering,* 2.

Müller, B., Reinhardt, J. & Strickland, M. T. (2012). *Neural networks: an introduction*, Springer Science & Business Media.

Murray, K. E. & Waller, R. (2007). Social networking goes abroad. *International Educator,* 16**,** 56.

Nahar, V. (2014). *On detection of cyberbullying in social networks.* PhD, The University of Queesland

Nahar, V., Li, X. & Pang, C. (2013). An effective approach for cyberbullying detection. *Communications in Information Science and Management Engineering,* 3**,** 238.

Nahar, V., Unankard, S., Li, X. & Pang, C. (2012). Sentiment analysis for effective detection of cyber bullying. *Web Technologies and Applications.* Springer.

Nalini, K. & Sheela, L. J. (2015). Classification of tweets using text classifier to detect cyber bullying. Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2. Springer, 637-645.

Nancy, P., Ramani, R. G. & Jacob, S. G. (2013). Mining of Association Patterns in Social Network Data (Face Book 100 Universities) through Data Mining Techniques and Methods. *Advances in Computing and Information Technology.* Springer.

Nandhini, B. & Sheeba, J. (2015). Cyberbullying detection and classification using information retrieval algorithm. Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015). ACM, 20.

Naradhipa, A. R. & Purwarianti, A. (2012). Sentiment classification for Indonesian message in social media. Cloud Computing and Social Networking (ICCCSN), 2012 International Conference on. IEEE, 1-5.

Nasreen, S., Azam, M. A., Shehzad, K., Naeem, U. & Ghazanfar, M. A. (2014). Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams: A Survey. *Procedia Computer Science,* 37**,** 109-116.

Nazief, B. A. & Adriani, M. (1996). Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. *Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta,* 41.

Nettleton, D. (2014). *Commercial data mining: Processing, analysis and modeling for predictive analytics Projects*, Elsevier.

Neuman, W. L. & Robson, K. (2012). Basics of social research: Qualitative and quantitative approaches.

Nitta, T., Masui, F., Ptaszynski, M., Kimura, Y., Rzepka, R. & Araki, K. (2013). Detecting cyberbullying entries on informal school websites based on category relevance maximization. Proceedings of the Sixth International Joint Conference on Natural Language Processing. 579-586.

O'Keeffe, G. S. & Clarke-Pearson, K. (2011). The impact of social media on children, adolescents, and families. *Pediatrics,* 127**,** 800-804.

O'Brien, N. & Moules, T. (2010). The impact of cyber-bullying on young people's mental health. *Chelmsford: Anglia Ruskin University*.

Omiecinski, E. R. (2003). Alternative interest measures for mining associations in databases. *IEEE Transactions on Knowledge and Data Engineering,* 15**,** 57-69.

Ott, R. L. & Longnecker, M. T. (2015). *An introduction to statistical methods and data analysis*, Nelson Education.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval,* 2**,** 1-135.

Parime, S. & Suri, V. (2014). Cyberbullying detection and prevention: Data mining and psychological perspective. Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on. IEEE, 1541-1547.

Parker, D. R., Gustafson, S. C. & Ross, T. D. (2005). Receiver operating characteristic and confidence error metrics for assessing the performance of automatic target recognition systems. *Optical Engineering,* 44**,** 097202-097202-9.

Police, I. (2016). *Cara Membuat Laporan Pengaduan Online (How to Report a Complaint Online)* [Online]. Jakarta: Indonesian Police. Available: https://www.polisionline.info/tag/cara-membuat-laporan-pengaduan-online/ [Accessed 3/11 2016].

Porter, M. (2006). Stemming algorithms for various European languages. *http://snowball. tartarus. org/texts/stemmersoverview. html As seen on May,* 16**,** 2013.

Porter, M. F. (2001). Snowball: A language for stemming algorithms.

Porter, M. F. (2005). Lovins revisited. *Charting a New Course: Natural Language Processing and Information Retrieval.* Springer.

Privitera, C. & Campbell, M. A. (2009). Cyberbullying: the new face of workplace bullying? *CyberPsychology & Behavior,* 12**,** 395-400.

Putra, D. F. (2014). *Ketika Bullying Berujung Maut* [Online]. Jakarta: CNN. Available: http://www.cnnindonesia.com/gaya-hidup/20140910112008-255-2906/ketika-bullying-berujung-maut/ [Accessed 4/4 2017].

Putra, E. N. (2016). *Urgency of Regulating CyberbullyingOn Indonesian Law* [Online]. Academia. Available: https://www.academia.edu/11602151/Urgency_of_Regulating_Cyberbullying_On_Indonesian_Law [Accessed 11/10 2016].

Qian, G., Sural, S., Gu, Y. & Pramanik, S. (2004). Similarity between Euclidean and cosine angle distance for nearest neighbor queries. Proceedings of the 2004 ACM symposium on Applied computing. ACM, 1232-1237.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR),* 28**,** 71-72.

Rahmadi, D. (2014). *Ini kata ABG tak kenal empati ibu hamil soal statusnya di Path* [Online]. Jakarta: Merdeka. Available: https://www.merdeka.com/peristiwa/ini-kata-abg-tak-kenal-empati-ibu-hamil-soal-statusnya-di-path.html [Accessed 3/11 2016].

Raisi, E. & Huang, B. (2016). Cyberbullying Identification Using Participant-Vocabulary Consistency. *arXiv preprint arXiv:1606.08084*.

Ramasubramanian, C. & Ramya, R. (2013). Effective pre-processing activities in text mining using improved porter's stemming algorithm. *International Journal of Advanced Research in Computer and Communication Engineering,* 2**,** 2278-1021.

Raorane, A., Kulkarni, R. & Jitkar, B. (2012). Association rule–extracting knowledge using market basket analysis. *Research Journal of Recent Sciences* 2277**,** 2502.

Raschka, S. (2014). Naive Bayes and Text Classification I-Introduction and Theory.

Razak, N. (2014). *Study: Most children in Indonesia are online now, but many are not aware of potential risks* [Online]. Geneva: United Nations Children's Fund UNICEF. Available: https://www.unicef.org/indonesia/media_22167.html [Accessed 28/8 2015].

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. Proceedings of the ACL Student Research Workshop. Association for Computational Linguistics, 43-48.

Reynolds, K., Kontostathis, A. & Edwards, L. (2011). Using machine learning to detect cyberbullying. Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on. IEEE, 241-244.

Riadi, I. & Hariani (2017). Detection Of Cyberbullying On Social Media Using Data Mining Techniques. *International Journal of Computer Science and Information Security,* 15**,** 244.

Richards, J. A. (2013). Supervised classification techniques. *Remote Sensing Digital Image Analysis.* Springer.

Riff, D., Lacy, S. & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research*, Routledge.

Ripley, B. D. (2007). *Pattern recognition and neural networks*, Cambridge university press.

Robert, C. (1999). *The Essence of Neural Networks*, Pearson Education India.

Rojas, R. (2013). *Neural networks: a systematic introduction*, Springer Science & Business Media.

Rosa, K. D., Shah, R., Lin, B., Gershman, A. & Frederking, R. (2011). Topical clustering of tweets. Proceedings of the ACM SIGIR.

Russell, M. A. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*, O'Reilly Media.

Safaria, T. (2016). Prevalence and Impact of Cyberbullying in a Sample of Indonesian Junior High School Students. *Turkish Online Journal of Educational Technology-TOJET,* 15**,** 82-91.

Safaria, T., Tentama, F. & Suyono, H. (2016). Cyberbully, Cybervictim, and Forgiveness among Indonesian High School Students. *Turkish Online Journal of Educational Technology,* 15.

Sánchez, D. & Batet, M. (2013). A semantic similarity method based on information content exploiting multiple ontologies. *Expert Systems with Applications,* 40**,** 1393-1399.

Sánchez, D., Vila, M., Cerda, L. & Serrano, J.-M. (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications,* 36**,** 3630-3640.

Sanchez, H. & Kumar, S. (2011). Twitter bullying detection (UCSC ISM245). Retrieved from

Sanders, J., Smith, P. & Cillessen, A. (2009). Cyberbullies: Their motives, characteristics, and types of bullying. fourteenth European Conference on Developmental Psychology, Vilnius, Lithuania.

Sarma, P. K. D. & Mahanta, A. K. (2012). Reduction of number of association rules with inter itemset distance in transaction databases. *International Journal of Database Management Systems,* 4**,** 61.

Savage, M. (2012). *Developing a Measure of Cyberbullying Perpetration and Victimization.* Arizona State University.

Schneider, S. K., Smith, E. & O'Donnell, L. (2013). *Social Media and Cyberbullying: Implementation of School-Based Prevention Efforts and Implications for Social Media Approaches*, EDC.

Setiawan, E. (2016). *Kamus Besar Bahasa Indonesia (the Indonesian Dictionary)* [Online]. Jakarta: Kemendikbud. Available: http://kbbi.web.id/ [Accessed 12 December 2016].

Setyawan, D. (2014). *KPAI : Kasus Bullying dan Pendidikan Karakter* [Online]. Jakarta: Indonesia's National Child Protection Commission (KPAI). Available: http://www.kpai.go.id/berita/kpai-kasus-bullying-dan-pendidikan-karakter/ [Accessed 2/2 2017].

Severyn, A. & Moschitti, A. (2015). Twitter sentiment analysis with deep convolutional neural networks. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 959-962.

Shacheng, W. (2012). Data Mining: Study on Intelligence-Led Counterterrorism. Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science. Springer, 87-93.

Shaikh, M. A., Wang, J., Liu, H. & Song, Y. (2007). Investigative data mining for counterterrorism. *Advances in Hybrid Information Technology.* Springer.

Shiells, K. & Pham, P. (2010). Unsupervised Clustering for Language Identification.

Singh, Y. & Chauhan, A. S. (2009). Neural networks in data mining. *Journal of Theoretical and Applied Information Technology,* 5**,** 36-42.

Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.,* 24**,** 35-43.

Sleglova, V. & Cerna, A. (2011). Cyberbullying in adolescent victims: Perception and coping. *Cyberpsychology: journal of psychosocial research on cyberspace,* 5**,** 23-46.

Slonje, R. & Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian journal of psychology,* 49**,** 147-154.

Smirnov, I. (2008). Overview of stemming algorithms. *Mechanical Translation,* 52.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S. & Tippett, N. (2008). Cyberbullying: its nature and impact in secondary school pupils. *J Child Psychol Psychiatry,* 49**,** 376-385.

Snakenborg, J., Van Acker, R. & Gable, R. A. (2011). Cyberbullying: Prevention and intervention to protect our children and youth. *Preventing School Failure: Alternative Education for Children and Youth,* 55**,** 88-95.

Soeriaatmadja, W. (2011). Bullying in Schools a Worry in Indonesia. *Jakarta Globe*.

Solove, D. J. (2008). Data mining and the security-liberty debate. *The University of Chicago Law Review***,** 343-362.

Solskinnsbakk, G. & Gulla, J. A. (2010). Combining ontological profiles with context in information retrieval. *Data & Knowledge Engineering,* 69**,** 251-260.

Song, T., Song, J., An, J.-Y. & Woo, J.-M. (2014). Social Risk Factor Prediction Utilizing Social Big Data.

Spears, B., Slee, P., Owens, L. & Johnson, B. (2009). Behind the scenes and screens: Insights into the human dimension of covert and cyberbullying. *Zeitschrift für Psychologie/Journal of Psychology,* 217**,** 189-196.

Spiegel, J. R. (2011). Mining of user event data to identify users with common interests. Google Patents.

Squicciarini, A., Rajtmajer, S., Liu, Y. & Griffin, C. (2015). Identification and characterization of cyberbullying dynamics in an online social network. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. ACM, 280-285.

Srikant, R., Vu, Q. & Agrawal, R. (1997). Mining Association Rules with Item Constraints. KDD. 67-73.

Su, M.-C., DeClaris, N. & Liu, T.-K. (1997). Application of neural networks in cluster analysis. Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1-6.

Subrahmanyam, K., Reich, S. M., Waechter, N. & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of applied developmental psychology,* 29**,** 420-433.

Sulianta, F. & Hendrawan, W. (2015). *Cyberbully, Cybervictim, and Forgiveness among Indonesian High School Students,* Bandung, Tablo.

Syam, A. A. (2015). *Tinjuan Kriminologis terhadap Kejahatan Cyberbullying*

Tagarelli, A. & Greco, S. (2010). Semantic clustering of XML documents. *ACM Transactions on Information Systems (TOIS),* 28**,** 3.

Tata, S. & Patel, J. M. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record,* 36**,** 7-12.

Thomas, L., Falconer, S., Cross, D., Monks, H., Brown, D. & Commission, A. H. R. (2013). *Cyberbullying, Human rights and bystanders* [Online]. Sydney: Australian Human Rights Commision Available: https://bullying.humanrights.gov.au/cyberbullying-human-rights-and-bystanders-0 [Accessed 29/4/2013 2013].

Thuraisingham, B. (2002). Data mining, national security, privacy and civil liberties. *ACM SIGKDD Explorations Newsletter,* 4**,** 1-5.

Thuraisingham, B. (2004). Data mining for counter-terrorism. *Data Mining: Next Generation Challenges and Future Directions***,** 157-183.

Thuraisingham, B. M. (2003). *Web data mining and applications in business intelligence and counter-terrorism*, CRC.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology,* 49**,** 1225-1231.

UNDP. (2013). *Internet and Racism: Cyberbullying* [Online]. United Nations Development Programe (UNDP). Available: https://munap2013.files.wordpress.com/2013/02/internet-and-racism-cyberbullying.pdf [Accessed].

Vega, V. B. (2001). Information retrieval for the indonesian language. *Master's thesis, National University of Singapore*.

Verma, T. & Renu, D. G. (2014). Tokenization and Filtering Process in RapidMiner. *International Journal of Applied Information Systems (IJAIS)–ISSN***,** 2249-0868.

Weisstein, E. W. (2002a). Gaussian function.

Weisstein, E. W. (2002b). Heaviside step function.

Weisstein, E. W. (2002c). Sigmoid function.

Whittaker, E. & Kowalski, R. M. (2015). Cyberbullying via social media. *Journal of School Violence,* 14**,** 11-29.

Wijaya, H., Erwin, A., Soetomo, A. & Galinium, M. (2013). Twitter Sentiment Analysis and Insight for Indonesian Mobile Operators. *ISICO 2013*.

Willard, N. (2006). Cyberbullying and cyberthreats. *Eugene, OR: Center for Safe and Responsible Internet Use*.

Witten, I. H. & Frank, E. (2011). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B. & Philip, S. Y. (2008). Top 10 algorithms in data mining. *Knowledge and information systems,* 14**,** 1-37.

Xhemali, D., Hinde, C. J. & Stone, R. G. (2009). Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages.

Xiong, H., Tan, P.-N. & Kumar, V. (2006). Hyperclique pattern discovery. *Data Mining and Knowledge Discovery,* 13**,** 219-242.

Yang, D. H., Kang, J. H., Park, Y. B., Park, Y. J., Oh, H. S. & Kim, S. B. (2013). Association rule mining and network analysis in oriental medicine. *PloS one,* 8**,** e59241.

Yang, Q., Garofalo, C., Gupta, Y., Cass, R., Wilson, K. & Sedukhin, I. (2008). Using neural networks for data mining. Google Patents.

Ybarra, M. L., Mitchell, K. J., Wolak, J. & Finkelhor, D. (2006). Examining characteristics and associated distress related to Internet harassment: findings from the Second Youth Internet Safety Survey. *Pediatrics,* 118**,** e1169-e1177.

Ye, J. (2011). Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling,* 53**,** 91-97.

Yeh, I.-C. & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications,* 36**,** 2473-2480.

Yuan, J., Wu, Y. & Yang, M. (2007). Discovery of collocation patterns: from visual words to visual phrases. 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 1-8.

Yulianti, K. Y. (2014). *Cyberbullying in Indonesian senior high schools: a study of gender differences.* Institute of Education, University of London.

Zhang, G. P. (2009). Neural networks for data mining. *Data mining and knowledge discovery handbook.* Springer.

Zhang, W., Yoshida, T. & Tang, X. (2008). Text classification based on multi-word with support vector machine. *Knowledge-Based Systems,* 21**,** 879-886.

Zhao, R., Zhou, A. & Mao, K. (2016). Automatic detection of cyberbullying on social networks based on bullying features. Proceedings of the 17th International Conference on Distributed Computing and Networking. ACM, 43.

Zhao, Y., Karypis, G. & Du, D.-Z. (2005). Criterion functions for document clustering. Retrieved from

Zhong, N., Li, Y. & Wu, S.-T. (2012). Effective pattern discovery for text mining. *IEEE transactions on knowledge and data engineering,* 24**,** 30-44.

Zhou, L. & Yau, S. (2007). Efficient association rule mining among both frequent and infrequent items. *Computers & mathematics with applications,* 54**,** 737-749.

Zhu, S., Wu, J., Xiong, H. & Xia, G. (2011). Scaling up top-K cosine similarity search. *Data & Knowledge Engineering,* 70**,** 60-83.

Zhu, X. (2007). *Knowledge Discovery and Data Mining: Challenges and Realities: Challenges and Realities*, Igi Global.

Appendix

No.	Time	User	Language	Text	location	re-tweet count
1	Fri Sep 02 15:53:06 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Jakarta	5
2	Fri Sep 02 15:51:49 AEST 2016	hidden	in	@syofaasavr bangsat ni anak gasopan amat sama adenya lupa trus.musnah kek lu	Bangdung	0
3	Fri Sep 02 15:50:58 AEST 2016	hidden	in	ini juga Bangsat sih, RS yang nolak pasien BPJS https://t.co/EySNIirPXd	Jakarta	0
4	Fri Sep 02 15:50:29 AEST 2016	hidden	in	RT @AHMADDHANIPRAST: Cuma si Bangsat yg gak mau cuti masa kampanye...pendukungnya idiot...tragis #ADP	Tagerang	85
5	Fri Sep 02 15:50:10 AEST 2016	hidden	in	halah gitu gitu plottwist nya ikon ke icn dulu baru deh ke jakarta. bangsat nyampenya malem aja udahlah	Bekasi	0
6	Fri Sep 02 15:46:06 AEST 2016	hidden	in	Fb aing aya nu nge hack tai bangsat!!!!	Jakarta	0
7	Fri Sep 02 15:45:24 AEST 2016	hidden	in	RT @fffffrhan: @akmalrazak7 haha bangsat aku dah mintak maaf setan. Kau je yg tak pernah mintak maaf dkt aku 👉👈	Bangdung	1
8	Fri Sep 02 15:43:34 AEST 2016	hidden	in	@akmalrazak7 haha bangsat aku dah mintak maaf setan. Kau je yg tak pernah mintak maaf dkt aku 👉👈	Jakarta	1
9	Fri Sep 02 15:42:20 AEST 2016	hidden	in	@masked569 @officialJKT48 lah bangsat gue kalah	Tagerang	0
10	Fri Sep 02 15:42:12 AEST 2016	hidden	in	saat sudah dirumah aku melihat awan dan ada orang yang berhubungan intim. namanya itu kan bangsat.	Bekasi	0
11	Fri Sep 02 15:41:19 AEST 2016	hidden	in	bangsat.	Jakarta	0
12	Fri Sep 02 15:38:53 AEST 2016	hidden	in	@abashvilla ANJING BANGSAT IYA	Bangdung	0
13	Fri Sep 02 15:36:06	hidden	in	RT @mata_indigo: BANGSAT..!!!! jelas siapa yg RASIS sekarang.!! https://t.co/csoSODWeFu	Jakarta	18

	AEST 2016					
14	Fri Sep 02 15:34:03 AEST 2016	hidden	in	RT @Eyyza9: Gentle la hadap depan2 jangan main bangsat kat belakang ☺	Tagerang	1
15	Fri Sep 02 15:33:24 AEST 2016	hidden	in	dasar loe #Mukidi kampung, gaya loe juga norak ,pantes cewek loe ngejar2 gw tpi sorri gw kagak terima #barangbekas #bangsat #jiah	Bekasi	0
16	Fri Sep 02 15:32:48 AEST 2016	hidden	in	Seniority Bangsat Diktator Anjing. — mendengarkan Superiots	Jakarta	0
17	Fri Sep 02 15:32:17 AEST 2016	hidden	in	Gentle la hadap depan2 jangan main bangsat kat belakang ☺	Bangdung	1
18	Fri Sep 02 15:32:12 AEST 2016	hidden	in	RT @TakPanasKe: The bangsat people bila kita dah janji dengan dia nak jumpa . Last2 dia sembang kita This we call cibai.	Jakarta	2301
19	Fri Sep 02 15:31:34 AEST 2016	hidden	in	RT @mata_indigo: Luar biasa KEBIADABAN APARAT BANGSAT..!! tertawa bangga selfa selfie selepas penggusuran.!! https://t.co/9zMNH65l8N	Tagerang	62
20	Fri Sep 02 15:30:02 AEST 2016	hidden	in	Brisek bangsat wkwk https://t.co/djuXo431Pc	Bekasi	0
21	Fri Sep 02 15:28:38 AEST 2016	hidden	in	RT @_zafrankhn: Panjang sebenar memang bangsat. Harapan 100% lk berterabur Hahahha	Jakarta	1
22	Fri Sep 02 15:28:11 AEST 2016	hidden	in	@mimiskr_ dari yang first sampai last semua nye macam bangsat ahahaha	Bangdung	0
23	Fri Sep 02 15:27:52 AEST 2016	hidden	in	@riski_prby so lu bangsat	Jakarta	0
24	Fri Sep 02 15:27:44 AEST 2016	hidden	in	RT @Mandala03309173: Jika ini benar, maka sangat pantas mereka disebut "Persekongkolan para Bangsat " tanah air https://t.co/CptewVellf	Tagerang	4
25	Fri Sep 02 15:26:43 AEST 2016	hidden	in	dulu sahabat sekarang jadi bangsat	Bekasi	0

26	Fri Sep 02 15:25:48 AEST 2016	hidden	in	Lebih seneng temen yang ngomong bangsat, tai, eek, dan sebagainya di depan daripada yang ngomong manis, tapi di belakang ngejelek-jelekin.	Jakarta	0
27	Fri Sep 02 15:24:19 AEST 2016	hidden	in	RT @akunkaget: ANJRI APA BANGSAT!!!!!	Bangdung	121
28	Fri Sep 02 15:21:52 AEST 2016	hidden	in	Gini nih yg bangsat banget.. Udah gusur udh diem aja gitu,ga usah ngeluarin kata2 yg bikin sakit. Dasar taik idup https://t.co/raBvo0KTG5	Jakarta	0
29	Fri Sep 02 15:15:10 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Tagerang	5
30	Fri Sep 02 15:12:46 AEST 2016	hidden	in	RT @zhouujq: bangsat gue ikutan penasaran :) https://t.co/B5bgBc6vO3	Bekasi	1
31	Fri Sep 02 15:09:49 AEST 2016	hidden	in	RT @cxpxjxn: Ni la jenis bangsat yang bwk kereta besar tapi bodoh, signal pun susah nak bagi.. Nasib xmati makcik tu.. Bangang http://t.co/…	Jakarta	15499
32	Fri Sep 02 15:08:08 AEST 2016	hidden	in	RT @OKWSsu: Rupamu Lugu Lugu ,Bangsat -'-	Bangdung	56
33	Fri Sep 02 15:07:54 AEST 2016	hidden	in	Ngurusin eKTP itu bangsat lamanya	Jakarta	0
34	Fri Sep 02 15:06:10 AEST 2016	hidden	in	empire damansara tempat paling bangsat bodoh takde atm apa lancau	Tagerang	0
35	Fri Sep 02 15:05:21 AEST 2016	hidden	in	RT @FridaRatih: bangsat	Bekasi	1
36	Fri Sep 02 15:05:10 AEST 2016	hidden	in	RT @Mandala03309173: Jika ini benar, maka sangat pantas mereka disebut "Persekongkolan para Bangsat " tanah air https://t.co/CptewVellf	Jakarta	4
37	Fri Sep 02 15:04:20 AEST 2016	hidden	in	RT @Mandala03309173: Jika ini benar, maka sangat pantas mereka disebut "Persekongkolan para Bangsat " tanah air https://t.co/CptewVellf	Bangdung	4
38	Fri Sep 02 14:59:10 AEST 2016	hidden	in	@alfaonline_ID min mhn info. Dlabel hrga brang 12.500 srwlh d chek out hrga jdi 16.000 knp bsa bgtu? Apa cma akalan alfa online yg bangsat??	Jakarta	0

39	Fri Sep 02 14:56:46 AEST 2016	hidden	in	Itu bangsat. BANGSAT. https://t.co/2lOzUJ2Sf5	Tagerang	1
40	Fri Sep 02 14:56:34 AEST 2016	hidden	in	temen kadang-kadang bisa bangsat juga yak	Bekasi	0
41	Fri Sep 02 14:55:58 AEST 2016	hidden	in	enjinir forward enjinir nanya doang enjinir bacot doang enjinir teamviewer enjinir gabut enjinir bangsat	Jakarta	0
42	Fri Sep 02 14:55:30 AEST 2016	hidden	in	RT @sapiqaiman: Aaaa bangsat 😂😂😂 https://t.co/eJvEppDBhP	Bangdung	1
43	Fri Sep 02 14:54:44 AEST 2016	hidden	in	Aaaa bangsat 😂😂😂 https://t.co/eJvEppDBhP	Jakarta	1
44	Fri Sep 02 14:54:21 AEST 2016	hidden	in	bangsat	Tagerang	1
45	Fri Sep 02 14:53:02 AEST 2016	hidden	in	DSR BANGSAT OH SEHUN AOSOSKXKSMXNCNX https://t.co/qvRvJjbFFG	Bekasi	0
46	Fri Sep 02 14:52:17 AEST 2016	hidden	in	ADA TEMAN YANG SUKA CERITA JELEK DIBELAKANG? ITU BUKAN TEMAN, TAPI BANGSAT. :)	Jakarta	0
47	Fri Sep 02 14:51:16 AEST 2016	hidden	in	BANGSAT BODAT BABIK AIR KENTOD https://t.co/0NZT4ji0dO	Bangdung	0
48	Fri Sep 02 14:48:23 AEST 2016	hidden	de	Lrt bangsat	Jakarta	0
49	Fri Sep 02 14:47:50 AEST 2016	hidden	in	Mamat india ni jalan depan aku dok buat muka bangsat apehal ..ade kene blasah jap gi kang .	Tagerang	0
50	Fri Sep 02 14:47:09 AEST 2016	hidden	in	RT @cxpxjxn: Ni la jenis bangsat yang bwk kereta besar tapi bodoh, signal pun susah nak bagi.. Nasib xmati makcik tu.. Bangang http://t.co/…	Bekasi	15499
51	Fri Sep 02 14:46:08 AEST 2016	hidden	in	Bangsat ndok ndok	Jakarta	0
52	Fri Sep 02 14:45:52 AEST 2016	hidden	in	BANGSAT PT 2 https://t.co/seQrAUuKPy	Bangdung	0

53	Fri Sep 02 14:45:00 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Jakarta	5
54	Fri Sep 02 14:44:30 AEST 2016	hidden	in	RT @di_2_kan: Ceritanya bangun pagi biar rejekinya gak dipatok ayam eh malah gebetan yg dipatok temen sendiri. Bangsat !!!	Tagerang	2
55	Fri Sep 02 14:43:06 AEST 2016	hidden	in	RT @mata_indigo: Luar biasa KEBIADABAN APARAT BANGSAT..!! tertawa bangga selfa selfie selepas penggusuran.!! https://t.co/9zMNH65l8N	Bekasi	62
56	Fri Sep 02 14:42:04 AEST 2016	hidden	in	@tvOneNews @VIVAcoid bangsat... rakyat diusir kayak anjing,diburu,dipukuli....ente semua digaji dr duit rakyat.dsr anjing	Jakarta	0
57	Fri Sep 02 14:41:23 AEST 2016	hidden	in	@apinksnxe bangsat lu emg	Bangdung	0
58	Fri Sep 02 14:40:59 AEST 2016	hidden	in	RT @_flow3rp3t: how come ppl can crush on me dah lah selekeh nokharom hahahahaha perangai cam bangsat gak ah kdg kdg hahahahahahahahahahaha…	Jakarta	2
59	Fri Sep 02 14:39:28 AEST 2016	hidden	in	RT @_flow3rp3t: how come ppl can crush on me dah lah selekeh nokharom hahahahaha perangai cam bangsat gak ah kdg kdg hahahahahahahahahahaha…	Tagerang	2
60	Fri Sep 02 14:38:11 AEST 2016	hidden	in	Jika ini benar, maka sangat pantas mereka disebut "Persekongkolan para Bangsat " tanah air https://t.co/CptewVellf	Bekasi	4
61	Fri Sep 02 14:36:36 AEST 2016	hidden	in	RT @Gemacan70: wkkwkwkwk @EndodiK kejar Proyek makanya @SuaraGolkar dukung BANGSAT	Jakarta	1
62	Fri Sep 02 14:35:06 AEST 2016	hidden	in	bangsat ganteng banget (3) https://t.co/r95G361iez	Bangdung	0
63	Fri Sep 02 14:34:48 AEST 2016	hidden	in	bangsat ganteng banget (2) https://t.co/gSx4wbRRke	Jakarta	0
64	Fri Sep 02 14:34:23 AEST 2016	hidden	in	bangsat ganteng banget https://t.co/d6N2WgeWLf	Tagerang	0
65	Fri Sep 02 14:32:16 AEST 2016	hidden	in	how come ppl can crush on me dah lah selekeh nokharom hahahahaha perangai cam bangsat gak ah kdg kdg	Bekasi	2

				hahahahahahahahahahahaha lawokk		
66	Fri Sep 02 14:29:30 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Jakarta	5
67	Fri Sep 02 14:27:58 AEST 2016	hidden	in	RT @AHMADDHANIPRAST: Cuma si Bangsat yg gak mau cuti masa kampanye...pendukungnya idiot...tragis #ADP	Bangdung	85
68	Fri Sep 02 14:25:26 AEST 2016	hidden	in	RT @mata_indigo: BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Jakarta	5
69	Fri Sep 02 14:25:03 AEST 2016	hidden	in	BANGSAT Ahok.!!!! #SaveIlyasKarim #SaveIlyasKarim #SaveIlyasKarim https://t.co/nG1IfRJex8	Tagerang	5
70	Fri Sep 02 14:21:53 AEST 2016	hidden	in	RT @cxpxjxn: Ni la jenis bangsat yang bwk kereta besar tapi bodoh, signal pun susah nak bagi.. Nasib xmati makcik tu.. Bangang http://t.co/…	Bekasi	15499
71	Fri Sep 02 14:20:27 AEST 2016	hidden	in	Ah sial punya temen mulutnya kntl, anjing, bangsat jd latah gue	Jakarta	0
72	Fri Sep 02 14:17:58 AEST 2016	hidden	in	Si bangsat ngomong nih https://t.co/jOFnYiCtT8	Bangdung	0
73	Fri Sep 02 14:17:44 AEST 2016	hidden	in	RT @Altefalken: si bangsat https://t.co/ode8bW3cz9	Jakarta	2
74	Fri Sep 02 14:16:12 AEST 2016	hidden	in	wkkwkwkwk @EndodiK kejar Proyek makanya @SuaraGolkar dukung BANGSAT	Tagerang	1
75	Fri Sep 02 14:11:38 AEST 2016	hidden	in	Apa masalah customer local ni eh ? Aku tanya elok elok kau jawab macam bangsat , aku dah start bangsat kau pulak tak boleh terima . Bodoh	Bekasi	0
76	Fri Sep 02 14:10:12 AEST 2016	hidden	in	bangga sangat lantik @amriyahyah_17 jadi kapten... apa @OngKimSwee tak terfikir nak develop pemain muda ke? atau @fam bangsat forever?	Jakarta	0
77	Fri Sep 02 14:05:07 AEST 2016	hidden	in	Sumpek bangsat	Bangdung	0

78	Fri Sep 02 14:04:58 AEST 2016	hidden	in	Bukan tanak berkawan tapi bila friendly sangat semuanya macam bangsat ! Sis dah serik 🙆	Jakarta	0
79	Fri Sep 02 14:02:21 AEST 2016	hidden	in	Dan semua ny terasa bangsat .. Smenjak ada masa lalu yg berusaha hadir kembalii !!!!	Tagerang	0
80	Fri Sep 02 14:02:13 AEST 2016	hidden	in	RT @qitmr: Wkwkwkwkwkwk bangsat iki ngamuk karo jok @lingganesia @onaldjahad (koyo pas bar ngamlet kae) https://t.co/CMUDnZOCPw	Bekasi	1
81	Fri Sep 02 14:01:45 AEST 2016	hidden	und	@bangsat_io	Jakarta	0
82	Fri Sep 02 14:00:43 AEST 2016	hidden	in	Bangsat @__ji95 https://t.co/J3z8rpXeVt	Bangdung	0
83	Fri Sep 02 14:00:12 AEST 2016	hidden	in	memang aku pandang dia buat muka jelaa . dasar melayu macam tu memang dasar bangsat . aku fikir tempat kerja aku ada policy je .	Jakarta	0
84	Fri Sep 02 13:54:52 AEST 2016	hidden	in	Bapak aku ji prangai bangsat	Tagerang	0
85	Fri Sep 02 13:52:19 AEST 2016	hidden	in	Kan emg bangsat sistah ^^ https://t.co/XOXdfNrmsU	Bekasi	0
86	Fri Sep 02 13:48:23 AEST 2016	hidden	en	@bangsat_io dm jum	Jakarta	0
87	Fri Sep 02 13:48:17 AEST 2016	hidden	und	@bangsat_io	Bangdung	0
88	Fri Sep 02 13:47:41 AEST 2016	hidden	in	Kl jar buhan lasga tu lugu lugu bangsat 😂	Jakarta	0
89	Fri Sep 02 13:45:04 AEST 2016	hidden	in	CJR BANGSAT! PERUSAK MORAL ! Masih kecil udh ngajarin cinta"an, BELUM SUNAT aja sombong ! Dah Sono Pergi kelaut aja ! Cc : All Comati	Tagerang	0
90	Fri Sep 02 13:44:23 AEST 2016	hidden	in	bangsat gue ikutan penasaran :) https://t.co/B5bgBc6vO3	Bekasi	1
91	Fri Sep 02 13:44:03 AEST 2016	hidden	in	Parah! Supaya gak bangsat lagi harus diapakan nih kak? https://t.co/BhL5Pa1Boy	Jakarta	0

92	Fri Sep 02 13:42:57 AEST 2016	hidden	in	@takyuyaki bangsat sih emang aku bayangin ceye pake wrangler rasanya pengin bergelantungan di pohon cabe 😣😣	Bangdung	0
93	Fri Sep 02 13:41:38 AEST 2016	hidden	in	@exokiso @Nmjaxxd bangsat 😊 bawel lu bedua	Jakarta	0
94	Fri Sep 02 13:40:55 AEST 2016	hidden	in	RT @Altefalken: si bangsat https://t.co/ode8bW3cz9	Tagerang	2
95	Fri Sep 02 13:38:45 AEST 2016	hidden	in	RT @missciccone: Udah ngezoom foto John Mayer di Instagram karena penasaran, ternyata tulisannya "Follow me on Snapchat". Bangsat. :))))	Bekasi	1
96	Fri Sep 02 13:37:18 AEST 2016	hidden	in	Bangsat gua penasaran setengah mati -_- https://t.co/ZQg6RcmTVN	Jakarta	0
97	Fri Sep 02 13:32:51 AEST 2016	hidden	in	si bangsat https://t.co/ode8bW3cz9	Bangdung	2
98	Fri Sep 02 13:30:14 AEST 2016	hidden	in	Semalam aku tau. Kamu pun lega. Bisa bernafas panjang skrg? Dasar bangsat!	Jakarta	0
99	Fri Sep 02 13:27:26 AEST 2016	hidden	in	RT @Brammantio_: Bangsat tp S2 pernah jd Bupati, DPR, WaGub & GubDKI, nek awakmu opo prestasinya, mikir taah..?!😁 #AyoSekolah https://t.co/…	Tagerang	33
100	Fri Sep 02 13:26:47 AEST 2016	hidden	in	Ulah nganggap mantan runtah, karna saterkutuk-kutukna mantan manehna pernah mere kabagjaan ka urang. Hatur nuhun nya BANGSAT!!! :((	Bekasi	0