Human Depression Analysis:

An Experimental Study of the Use of AI Botics for Early Detection

Thesis submitted in fulfilment of the requirements of the degree of
Master by Research
College of Engineering and Science
Victoria University

by

Payam Kaywan

February 2022

Principal Supervisor: Khandakar Ahmed

Supervisors: Ayman Ibaida, Yuan Miao

Intelligent Technology Innovation Lab, Victoria University, Melbourne, Australia

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Declaration of Authenticity

"I, Payam Kaywan, declare that the Master by Research thesis entitled Human Depression Analysis an Experimental Study of the Use of AI Botics for Early Detection is no more than 60,000 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 award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

"I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University's Higher Degree by Research Policy and Procedures.

All research procedures reported in the thesis were approved by the Victoria University Human Research Ethics Committee office for Research with ethics approval number of HRE20-184".

Signature: *Payam Kaywan* Date: 22 February 2022

Acknowledgements

I would like to share my sincere appreciations to the staff and the supervisors at Victoria University who supported me all the way through to successfully pass the Confirmation of Candidature (CoC), Towards Submissions and in preparation of this final thesis. They all assisted with a robust and caring support.

I appreciate Dr. Khandakar Ahmed for his knowledgeable dedication throughout the Master by Research journey. I have learnt a wide range of skills by working closely with him and he dedicates his time to arrange meetings when and where they were required. Dr. Ahmed has a unique character that he would like to help anyone who needs a guidance or support at any stages. I have benefited from his ethics and manner to approach the research.

I would like to appreciate Dr. Ayman Ibaida for his professional and thoughtful dedications. He is target oriented and he has a set of skills to solve even very complicated tasks. His attention to details and how to approach the issues have been a role model for me and I always look forward to working with him whenever the chances are available.

Moreover, although Dr. Yuan Miao has a very busy schedule, he was contacting me to guide me through my research from time to time. I would like to thank him for his hints and helpful guidance.

During the master's journey, I had the privilege to be working with Dr. Bruce Gu, Dr. Rose Lucas and Professor Ron Adams. Dr. Bruce Gu assisted me on how to prepare the final version of a paper and how to submit with following the requirements of a conference. Dr. Lucas and Professor Adams presented the prerequisites courses for both PhD and Masters' students and they familiarized us with the points to be cautiously have in mind during the research. I also would like to appreciate Dr. Deborah Zion for her course presentation and especially with the ethics approval for my research. You all are brilliant and together you form the staff of Victoria University with a promising commitment for the future endeavors.

I want to take this opportunity to appreciate my family in Canada, my mom and my sister, for their endless support and motivation even though we are miles away from each other. Due to Covid-19, the same as most of the families around the world, we have been losing the chance to meet each other for about two years, however, their love and care have always been there for me through video calls, calls and messages. I am confident they would be delighted to find out that I have managed to receive my master's degree especially after years of hard working and passing through the challenges. I love you all from bottom of my heart.

Abstract

The world is facing a shortage of professional medical staff, a situation which has been exacerbated by the COVID-19 pandemic which has significantly increased challenges globally and has had an adverse impact on the health care system. This has also led to additional barriers to patient care access, specifically for individuals who are in need of constant special care. To address the issue of limited access to medical professionals, medical assistance can be provided to patients in the form of a chatbot which acts as a proxy between psychiatrists and patients and is available and accessible 24/7.

Although there has been a degree of success in developing medical chatbots, many medical professionals believe that the use of chatbots in early depression detection needs to be more practical which will require further research. In this research, we address the well-known and common shortcomings which have been discussed in the recent literature. Three of these shortcomings are as follows: firstly, there is a lack of open-ended questions to enable participants to interact openly and without any restrictions about their moods and emotions as most bots in the literature constrain the participants' responses by limiting them to multiple choice questions which means the participants are not able to open up and describe their real feelings freely. Secondly, there is a lack of semantic analysis to draw exact meaning from a text. Thirdly, there is a requirement for participants to make a long-term commitment in terms of their involvement in the research.

This research introduces a depression analysis chatbot, DEPRA, which aims to resolve some of these shortcomings and challenges by asking open-ended questions, providing semantic analyses and automatic depression scoring. DEPRA is developed using contemporary bot platforms, Dialogflow on Google cloud-based infrastructure, and is integrated with social network platforms such as Facebook. Most chatbots today are designed for therapeutic purposes. However, the DEPRA chatbot is designed with a focus on the detection of depression in its early stages. DEPRA is designed based on a structured early detection depression Standard Interview Guideline the Hamilton Depression Scale (SIGH-D) and Inventory of Depressive Symptomatology (IDS-C), which is used by professional psychiatrists in

triage sessions with patients. DEPRA has been trained with personalized utterances from a focus group. This research utilizes Natural Language Processing (NLP) to identify the depression level of participants based on their recorded conversation. DEPRA uses a scoring system to determine the participant's depression level and severity.

This research also details a non-clinical trial with 50 participants who interacted with the DEPRA chatbot. Due to the ethical limitations of this research, such as only residents of Australia and participants to be in the age group of 18 to 80 years old, we have approached a dataset with only 50 participants. This size of dataset was suitable to conduct and run the research. However, the future studies will target a more comprehensive dataset. This study was a first stage of utilizing Chatbot for early detection of depression. In this stage our goal was to develop the system not to run the clinical trial. Therefore, we required a sample that could assist us mainly to identify the accuracy of the system developed. Future work is to access evaluation by human expert which goes into the next phase of the project and could also include extending the sample and/or enhancing the system further and the assistance would be offered to western health. Therefore, at this stage 50+ sample sufficed to capture various responses by the people that had participants of different level of depression.

To evaluate the autoscoring feature of DEPRA, the accuracy of the Machine Learning (ML) algorithms is calculated. Accordingly, manual scoring is compared with the calculated depression scores. The average accuracy of the 27 questions related to the linear SVC of the 26 participants' experiment is 88%, the SGD algorithm of 40 participants' experiment is 80%, and the linear SVC of 50 participants' experiment is 87%. Furthermore, the overall satisfaction rate of using DEPRA was 79% indicating that the participants had a high rate of user satisfaction and engagement.

Table of Contents

DECLARATION OF AUTHENTICITY II
ACKNOWLEDGEMENTS III
ABSTRACT V
TABLE OF CONTENTS VII
LIST OF FIGURES IX
LIST OF TABLES X
ACRONYMS X

CHAPTER 1 INTRODUCTION 11

1.1 Introduction 12
1.2 Background 13
1.3 Overall Aim and Motivation 14
1.4 Research Questions 15
1.5 Statement of Significance 16
1.6 Contributions 17
1.7 List of Publications 18
1.8 Thesis Composition 19

CHAPTER 2 LITERATURE REVIEW 21

2.1 Introduction 22
2.2 Literature Review 23

CHAPTER 3 RESEARCH DESIGN 39

3.1 Introduction 40
3.2 Methodology and Conceptual Framework 41
3.3 Experimental Design 65
3.4 DEPRA Chatbot Implementation 70
3.5 DEPRA Features 71

CHAPTER 4 NON-CLINICAL TRIAL USING DEPRA 72

4.1 Introduction 73
4.2 Data Collection 73
4.3 Analysis and Interpretation 79

CHAPTER 5 AUTOMATIC SCORING 91

5.1 Introduction 92
5.2 AutoScoring Model 92
5.3 Results and Discussion 105

CHAPTER 6 CONCLUSION 112

6.1 Limitations and Future Work 113

APPENDICES 115

APPENDIX A – FULFILLMENT CODE FOR DEPRA CHATBOT DESIGN 115
APPENDIX B – ETHICS APPLICATION RELATED TO THIS STUDY 115
APPENDIX C – CONSENT FORM FOR PARTICIPANTS INVOLVED IN RESEARCH 115
APPENDIX D – USER RATING FORM 115
APPENDIX E – ATOMIC DATA OF FOCUSED GROUP RESPONSES 115

APPENDIX F – DEPRA CHATBOT SCORING – 50 PARTICIPANTS 115
REFERENCES........ 116

List of Figures

Figure 1 - DEPRA chatbot Design and Implementation Schema 41
Figure 2 - Conceptual flowchart of Depressive Symptoms........ 48
Figure 3 - Dialogflow Intents Window 51
Figure 4 - Dialogflow Training Phrases Section 52
Figure 5 - Dialogflow Action and Parameters Section 53
Figure 6 - Dialogflow Responses Section 54
Figure 7 - Dialogflow Entites defined for DEPRA 55
Figure 8 - Dialogflow input and output Contexts 56
Figure 9 - Overview of Dialogflow Console 57
Figure 10 - Overview of the Integrations Section on Dialogflow 58
Figure 11 - DEPRA Chatbot Integration with Facebook Messenger 58
Figure 12 - Fulfillment Section on Dialogflow........ 59
Figure 13 - Schema/E-R diagram of the central database 62
Figure 14 - AWS RDS Service for Database 63
Figure 15 - Display of the Schemas Designed in MySQL Workbench Client 64
Figure 16 - DEPRA Participant Recruitment Flow 77
Figure 17 – Closed Group Statistics based on Gender, Age and Profession 80
Figure 18 - IDS-SR Depression Level Statistics 84
Figure 19 - QIDS-SR Depression Level Statistics 85
Figure 20 - IDS-SR vs QIDS-SR 86
Figure 21 - Depression Level Statistics According to Participants Interactions 88
Figure 22 – Auto Scoring Block Diagram 92
Figure 23 – Scatter Plot Sample of Multiclass Classification Dataset 93
Figure 24 – Word Cloud in Python 94
Figure 25 - Logistic Function 98
Figure 26 - SVM Classification .... **Error! Bookmark not defined.**
Figure 27 - SVM Hyper Plane Scenario .... **Error! Bookmark not defined.**
Figure 28 - Applicable Algorithm for DEPRA chatbot for 26 Participants 107
Figure 29 - Applicable Algorithm for DEPRA chatbot for 40 Participants 109
Figure 30 - Applicable Algorithm for DEPRA chatbot for 50 Participants 111

List of Tables

Table 1 - Summary of available chatbots with regards to intervention type 24
Table 2 - DEPRA Chatbot Dialogflow Intents Breakdown 51
Table 3 - Features of Early Depression Detection DEPRA chatbot 71
Table 4 - Demographic Variables of Participants 74
Table 5 - Symptoms Categories and Relevant Questions of Scoring Systems 78
Table 6 - Sequence of Responses 81
Table 7 - Comparison of Total Scores with regard to Depression Levels 84
Table 8 - User Satisfaction and Engagement 90
Table 9 –Number of Features per Question for Count Vectorizer 96
Table 10 - Confusion Matrix 105
Table 11 - Five Algorithms and their Accuracies per each 27 Questions (26 Participants) 106
Table 12 - ML Algorithms Count of Appearance (26 Participants) 107
Table 13 - Five Algorithms and their Accuracies per each 27 Questions (40 Participants) 108
Table 14 - ML Algorithms Count of Appearance (40 Participants) 109
Table 15 - Five Algorithms and their Accuracies per each 27 Questions (50 Participants) 110
Table 16 - ML Algorithms Count of Appearance (50 Participants) 111

Acronyms

The following list of abbreviations are used in all over this thesis:

AI	Artificial Intelligence
AI-Bots	Artificial Intelligent Botics
Amazon RDS	Amazon Relational Database Service
AWS	Amazon Web Services
BAES	Bot – Autonomous Emotional Support
CA	Counselling Agent
CBT	Cognitive Behavioural Therapy
CNN	Convolutional Neural Networks
DADRM	Digitally Advanced Depression Relieving Machine
DEPRA	Depression Analysis Chatbot
Dialogflow CX	Dialogflow Customer Experience

Dialogflow ES	Dialogflow Essentials
DSM-4	Diagnostic and Statistical Manual
ECA	Embodied Conversational Agent
fMRI	Functional Magnetic Resonance Imaging
GAD-7	Generalized Anxiety Disorder 7-item scale
HR	Human Resources
IDS-C	Inventory of Depressive Symptomatology – Clinician Rated
IDS-SR	Inventory of Depressive Symptomatology – Self Rated
Linear SVC	Linear Support Vector Classifier
MCQ	Multiple Choice Questions
MDD	Major Depression Disorder
ML	Machine Learning
MLE	Maximum-Likelihood Estimation
MRI	Magnetic Resonance Imaging
NLP	Natural Language Processing
OSMI	Open Sourcing Mental Illness
PHQ-9	Patient Health Questionnaire
PTDS	Post Traumatic Stress Disorder
QIDS-SR	Quick Inventory of Depressive Symptomatology – Self Rated
RCT	Randomized Controlled Trial
SA	Sentiment Analysis
SGD	Stochastic Gradient Descent

SIGH-D	Structured Interview Guide for the Hamilton Depression Rating Scale
SVM	Support Vector Machine
TAU	Treatment As Usual
WHO	World Health Organization

CHAPTER 1 INTRODUCTION

1.1 Introduction

Worldwide, one person every 40 seconds dies by committing suicide [1]. The flow-on effects of suicide can have a serious impact on the community as it struggles with emotions such as grief, sorrow and pain. With many suicide cases reported every month, early detection and prevention are imperative to reduce fatality [2]. The unprecedented COVID-19 epidemic and the pace of change in society have led to an increased number of mental health issues and consequently an increased number of suicides. Depression is a serious cause of social challenges, such as an inability to function in the workplace or at school and difficulty in meeting daily responsibilities in general. Severe depression is the leading cause of suicide for all age groups around the world. A psychiatric evaluation and diagnosis by specialists is considered the most effective way to detect depression. However, the lack of qualified specialists may make it difficult for depressed individuals to access the help they need. This research uses Artificial Intelligent Botics (AI-Bots) for the early detection of depression to intervene and prevent fatality. One of the common problems for those with mental health issues is the perceived social stigma which is associated with visiting a mental health facility. One of the benefits of AI-Bots is that patients can receive assistance and an early diagnosis for their mental health issues without the need to visit a mental health professional, which is particularly useful given the current lack of qualified psychiatrists. In this research, the initial goal is to detect depression at early stages with the use of a bot. Chatbots simulating human conversions can enable personalized interactions and enhance patients' engagements as opposed to online resources. DEPRA chatbot has the mission to interact with potential individuals to assist them with their symptoms and guide them to the path of having their depression cured. The rationale behind choosing a chatbot to collect the data relies on the fact that the bot can be available 24/7, the data collection phase can be reduced to several minutes instead of hours and days of sampling and the easiness of access the chatbot either in the comfort of an office or readily at home.

ML techniques, a subset of AI, can play a vital role in mass surveillance due to their ability to detect anomalies and provide potential solutions for crisis prevention. Furthermore, a complicated traditional clinical diagnosis requires the extensive participation of experts. This research focuses on the implementation of AI chatbots and the use of ML, NLP and sentiment analysis techniques to diagnose depression so individuals can seek treatment from psychiatrists and medical professionals.

1.2 Background

Loneliness is an unintended consequence of the widespread use of technology around the world. Nowadays, people tend to spend a lot of time on mobile devices in private and public environments rather than socializing. Social isolation during the COVID-19 pandemic has also had a significant negative psychological impact globally. Prolonged isolation, feeling of loneliness, financial challenges due to job losses, and grief over the death of loved ones has potentially resulted in an increasing deterioration in the mental health status of a considerable proportion of the global population. To address the issue of depression, governments around the world are allocating an increased budget to reduce the adverse impacts of psychological problems and to conduct practical research to reduce the side effects of depression. According to WHO, about one million people with depression commit suicide each year [3]. Hence, there is a need to conduct research into the cause of depression in an attempt to reduce the number of suicides globally.

Although the application of AI and ML opens new horizons and brings hope in the delivery of healthcare services to improve the population's physical and mental well-being, as a new paradigm, it requires further study before it is practical in the healthcare domain. More research is also needed to adopt AI and ML to the specific application of depression. ML is useful in three areas, pattern recognition, anomaly detection and prediction. Therefore, the adoption of ML can play a significant role in detecting and preventing depressive crises. ML involves a significant range of algorithms that are promising in terms of diagnosing and assisting clinicians and patients by offering more practical methods.

In this study, we applied AI and ML technologies and implemented a chatbot, DEPRA, which is based on contemporary bot platforms and has the capability to assess mental health symptoms and detect early depression. The severity level of the identified depression is also calculated by the DEPRA chatbot using a structured early detection depression interview guide based on SIGH-D and IDS-C. DEPRA has the capability to emulate a psychiatrist's triage process and act as a proxy between health professionals and patients. DEPRA asks a series of open-ended questions, including the main and sub-questions, to interact with the participants and potential patients to determine their well-being and rate their level of depressive disorder. The participants can interact with the chatbot with no limitations and briefly explain their emotions and moods with no boundaries. If a patient is deemed to be suffering from severe depression, they are advised to visit a medical professional to receive assistance as required.

1.3 Overall Aim and Motivation

The need for an innovative digital-based healthcare system is increasing and the application of AI-based services in diagnostic clinical procedures and the treatment of health issues has gained momentum in recent years. This research focuses on human depression analysis in its early stages.

This thesis explores potential solutions to manage incidents resulting from depressive symptoms. We implemented a mental health bot, DEPRA, to detect depression levels by asking questions to identify the participant's mood, feelings of guilt, suicide ideation, insomnia, agitation or retardation, anxiety, appetite and weight change, interpersonal sensitivity and somatic symptoms and to assist patients who suffer from depression. NLP techniques and ML algorithms are used in this research to conduct multiple experiments and build a prediction model to detect depression and identify the degree of severity. ML is the technology underpinning this study and defines how machines learn from data to solve complex problems and emulate human capabilities using mathematical algorithms to recognise patterns and train the models to make predictions.

The DEPRA chatbot is used as a digital assessor to interact with the participants and collect their data for further analysis and severity rating. DEPRA is available 24/7 and it offers the participants a base with which they can interact anytime and anywhere. This flexibility provides supportive and trustworthy assistance to the patients as they can interact with a non-human agent any time and they will receive advice about the next step and whether they should discuss their well-being with a health care professional. The method of text messaging instead of verbal communication is regarded as a positive feature so the participants feel more connected to the chatbot rather than a psychiatrist's session. Another positive feature is that the participants only need to interact with the chatbot for one 30-minute session.

The outcome of this research will assist patients through emergencies including pandemics, natural disasters and the lack of available and accessible medical professionals. DEPRA is a virtual assistant to medical professionals and cannot replace human medical professionals who are the ones qualified to deal with patients suffering from depression and prescribe the appropriate medications. However, AI Botics such as DEPRA can assist patients, caregivers and medical professionals by easing their burden, especially in a time of emergency.

1.4 Research Questions

This section defines the research questions. By finding the answers to these questions, the research team will achieve the goal of the study, which is to apply a bot to detect human depression and its severity in the early stages with the aid of AI, ML and NLP. The research questions are as follows:

1) How can a chatbot be utilized to conduct a structured interview for human depression rating?
2) How feasible and efficient it is to use AI-enabled chatbots to measure depression in its early stages?

3) How can classical NLP techniques be leveraged to measure depression levels automatically?

1.5 Statement of Significance

"Almost half (45%) of the Australian population will suffer from a form of mental illness in their lifetime"[4]. Depression is a serious mental health disorder with the potential for complications that can adversely affect one's life [5]. Advanced and economically powerful countries have established lifeline call centres for people who are suffering from mental health problems. Although these services are important they are not sufficient to address the burgeoning problem of mental health issues in the community. Suicide mortality is significantly increasing among young and elderly women (approximately 800,000 deaths per year) in low- and middle-income countries [6].

Intentional self-harm is one of the leading causes of death among Australians aged between 15 and 44, with the lowest median age of death at 43.9 [6]. Therefore, early detection and prevention are imperative to reduce fatalities. Early depression detection analysis can provide an early warning in relation to behaviours such as suicidal thoughts, self-harm and declining mental health. Chatbots offer three major advantages to the medical world by reducing the gap between the available resources compared to the number of people, alleviating the social stigma associated with the need to visit a mental health facility, and addressing the fear of being judged by others. The goal of chatbots is to provide access to mental health systems in a time of need. The design of these chatbots is influenced by patients' suggestions and recommendations which are evaluated by researchers and applied in the design of chatbots to serve vulnerable populations. Mental disorders, especially depression, are experienced at least once by approximately 29% of people in the world [7, 8]. Depression has a direct effect on people in all countries around the world and has an adverse impact on their ability to concentrate and their general ability. Consequently, those who are depressed tend not to contribute to society and they avoid social interactions with peers, families, housemates or colleagues in the workplace. Although chatbots cannot replace a human psychiatrist or medical professional, they can interact with a human in the three formats, vocal, text or animation [9]. There are two types of chatbots, rule-based and intelligent chatbots. The first type applies rules or decision trees to manage the conversation while the second type is based on AI [9]. Chatbots are designed and used in a wide variety of applications such as e-commerce and retail, hospitality, real estate, finance and banking and so on. One the main applications of bots is in health care and mental health. The trend of facilitation and access to health care and mental health for the majority of the population had a promising increase in last five years. There are around 41 different bots used in the area of mental health. Mental health chatbots are utilized in areas such as therapy, training, education and so forth [9]. A report published by World Health Organization (WHO) revealed that in 2017, around 4.4% of the world's population had a depression disorders. which equates to 322 million people globally [10]. The prediction of WHO is that by 2030, depression will be the highest cause of death in advanced and third world countries around the globe [11].

The psychological consequences of the COVID-19 pandemic crisis in Australia and around the world are already visible and even by conservative estimates, they are yet to reach their peak. It has had a significant impact on people with pre-existing mental disorders due to the limited access to medical staff for people with mental conditions who may be in immediate need of support as well as the detrimental effects of prolonged social isolation, a lack of interaction, nervousness, and post-traumatic and environmental stress. Medical staff have attempted to mitigate the risks and find ways to support people with depression in these difficult times. However, it is inevitable that the average number of suicides will increase dramatically due to isolation. Research into the use of AI and ML to minimize the burden on patients, caregivers and medical staff is more crucial than ever.

1.6 Contributions

The contribution of this research can be summarized as follows:

- to design a chatbot, DEPRA, which asks a set of questions based on the structured interview guide for the SIGH-D and IDS-C to diagnose depression in the participants according to their responses. The DEPRA chatbot is unique chatbot in that it asks clinically proven questions to deal with participants.

- to conduct a non-clinical trial to investigate the feasibility and efficacy of using an AI Bot to detect depression at an early stage. 50 participants were involved in this trial, aged between 18 to 80. All participants were located in Australia. The participants' demographic information, user experience and manual scoring are the components used to measure depression level severity.

- to automatically score the depression level with the use of NLP classification techniques. Five ML algorithms are examined to find the most applicable algorithm and the accuracy of the algorithms is calculated. The three experiments show that the Linear SVC and SGD classifier are the most applicable algorithms. Automatic scoring was compared to manual scoring in this research. After comparing the accuracy, the most applicable algorithms with the highest accuracy for each of the three experiments was Linear SVC with 22 out of 27 questions for 26 participants, the SGD classifier for 11 out of 27 questions for 40 participants, and Linear SVC with 10 out of 27 questions for 50 participants.

1.7 List of Publications

1. P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao and B. Gu, "DEPRA: An Early Depression Detection Analysis Chatbot," in *10th International Conference on Health Information Science (HIS),* 2021.

2. P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao and B. Gu, "Early Detection of Depression Using Conversational AI Bot: A Non Clinical Trial,"in *PLOS ONE,* 2022 (Submitted).

3. P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao and B. Gu, “DEPRA: Early Depression Detection Chatbot and Automatic Scoring,”in *Frontiers in Psychiatry*, 2022 (Draft).

1.8 Thesis Composition

The rest of this thesis includes the following chapters:

Chapter 2 discusses the literature review with regard to the chatbots which are available on the market. Bot families, including those for psychosocial intervention and early depression detection are introduced. Various intervention types such as CBT, lifestyle management, PTDS management, and work efficiency improvement are discussed. The bots and their application in areas such as mental health are explained and the way each study managed the chatbots in real applications is discussed. Factors such as participant recruitment, the design methodology and the method applied to collect data and run the implementation of the bots are summarized. The literature survey reveals that, to the best of our knowledge, the use of a chatbot for early depression detection has not been previously mentioned in the existing literature.

Chapter 3 summarizes the research design by answering one of the research questions: “How can a chatbot be utilized to conduct a structured interview for a human depression rating?” First, the chatbot design and implementation schema is presented. The process is explained and the components which are required for a successful implementation are discussed. The flowchart which is designed based on the Hamilton Guideline is presented and the questions and their sequence are summarized in this chapter. The reason that Dialogflow, as the contemporary platform, is applied is reviewed and the components and section of this platform are briefly explained. The intents of Dialogflow which are utilized for DEPRA chabot design are detailed. These intents are summarized in accordance with the occurrences of each intent in this research. Fulfillment and the inline editor as the programming environment of the Dialogflow is explained briefly. A sample of the code behind the functions implemented in Fulfillment is presented and the database and the Amazon RDS and the client, MySQL Workbench, is summarized. The four tables which are designed for this study and their relationships are also discussed and the experimental design of the chatbot is reviewed. The ethics approval and the requirements to comply with the rules and regulations for this study are briefly explained. Two forms, the user consent form and the user rating form, are presented and discussed briefly. Finally, in this chapter, the implementation and the features of the chatbot are summarized.

Chapter 4 summarizes the various stages of the chatbot non-clinical trials by reviewing the participants' demographic details, their user experience and manual scoring to determine the depression severity of the participants. The two scoring methods, IDS-SR and QIDS-SR, are discussed. The participants and the details on the recruitment process, addressing the depressive variables and the analysis are briefly explained. At the next stage, chatbot training and the use of a closed group is discussed. The last sections summarize the depression score and user satisfaction and engagement.

Chapter 5 introduces NLP techniques to perform sentiment analysis on the dataset and calculate autoscoring. The ML algorithms and the theory behind the selected five algorithms are explained. The ML algorithms and the most applicable are selected according to their accuracy. This is achieved in the three stages of autoscoring, data pre-processing, feature extraction, and classification. Chapter 5 details the results and discusses the results of the experiments.

Chapter 6 summarizes the conclusion of this study and suggests directions for future work which needs to be completed in the next stages.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

The connection between suicidal thoughts and depression are undeniable. In most cases, severe depression can result in self-harm and sometimes suicidal thoughts [12]. Suicidal trends are significantly increasing among young and aged women in countries with low and middle salaries and these statistics have reached approximately 800,000 cases annually [6]. Intentional self-harm remains one of the top five leading causes of death among Australians aged between 15 and 44, with the lowest median age of death at 43.9 [13]. In addition, as per the plan by the WHO between 2019 to 2023, WHO aims to facilitate access to the mental health system within 12 countries aiming at 100 million more people in vulnerable societies [6]. In this research, as a direct connection with this aim followed by WHO, we establish the idea of early detection of depression to target the best momentum to tackle the issue of the depression at the best time so the mental disorder can be in control by medical professionals. This research is not aiming to replace the traditional diagnostic methods by any medical professionals, such as psychiatrists, but instead, it would like to contribute to the trend of addressing the mental disorder in this case depression at early stages. In developed countries, there are 9 psychiatrists per 100,000 population and it is at an incredibly low rate of 0.1 psychiatrists per 1 million population in underdeveloped countries around the globe [14].

According to the requirements of this era, we are moving towards chatbots that can assist the participants in emergency cases such as commit suicides. The chatbots should be able to respond proactively and precisely to avoid unpleasant consequences. Moreover, due to the nature of the depression itself, participants may attach to the para-social relationship, a one-sided relationship. As a result, therapeutic limitations and interconnections are not addressed properly in this new era of technology and chatbots as well. Most of the chatbots cited in the literature focus on delivering Cognitive Behavioural Therapy (CBT), tracking the adherence of physical activity and medication, and healthy lifestyle recommendations. There is a wide prevailing criticism over the surprisingly sparse clinical trial and therefore lack of confidence, Woebot [15, 16] is an exception though. Additionally, to the best of our knowledge, there is no noticeable contribution toward the use of bot in mass screening and early detection of depression. Almost all of the researches used custom-built bot instead of deploying an application on widely used platforms such as Google Cloud, Dialogflow, Microsoft Azure, AWS, Amazon Alexa or Siri.

2.2 Literature Review

Contemporary mental health bots built using conversational agents over the last half-decade aims to a) provide psychosocial intervention in improving mental health, lifestyle or work efficiency and b) identify physical and/or mental health disorders in patients. The experiment duration time required from participants is considered a feature for any chatbots when it is less than 1 week. The reason behind this feature is that the participants do not need to be monitored for long durations and this can create more reliable results. Some researchers have complex process when the nature of the research requires to track the participants on regular basis or when the participants are required to perform self-monitoring over the research lifecycle or a long time commitment is required which in some cases it resulted in complexity to track the trend and leading to inaccurate consequences. It has been proved that participants lose enthusiasm and concern for the research in cases that they require to interact with a chatbot or any experiments for a long time. Participants prefer to commit to short term and less time-consuming interactions. DEPRA provides this capacity of the interaction to be completed within 30 minutes. Following the completion of the interaction, participants will be prompted to the validation and greeting message by the end of the experiment and they will be informed of a rating page and the fact that their record has been saved into a database. We also provide an automatic scoring by the end of the interaction so the users will be informed of the level of depression, if it exists, and how to deal with it as a comment. Before discussing the additional features of DEPRA to address the existing shortcomings in the current literature, such as lack of semantic analysis of response and restrictions to the short and Multiple Choice Questions (MCQ) based interactions, we will be reviewing the pervious work in the last decade and compare various bot family from different perspective including the intervention type and the time commitment by participants. Table 1 summarizes the bot family concerning intervention types.

Table 1 - Summary of available chatbots with regards to intervention type

Bot Family	Intervention Type	Chatbot	Study Duration	Year
Psychosocial Intervention	CBT	Shim	2 weeks	2017
		Woebot	2 weeks	2017
		Tess	2 to 4 weeks	2018
		Help4Mood	Not Reported	2015
		T-bot	Not Reported	2018
		KokoBot	5 Days	2018
		Mindspa	Not Reported	2021
		Serenity	Not Reported	2021
		Youper	Not Reported	2021
		MindDoc	Not Reported	2021
	Lifestyle Management	Gabby	4 weeks	2017
		Laura	4 weeks	2019
		Tess	2 to 4 weeks	2018
		Wysa	2 consecutive screening time points	2018
		ViviBot	Weeks 2, 4 and 8	2019
		CARO	Not Reported	2020
		BAES (Bot Autonomous Emotional Support)	Not Reported	2018
		InnerHour Self-Care Therapy	Not Reported	2021
	PTSD Management	WorldBuilder	1 day	2017
		Unnamed [39]	1 day	2017
	Work Efficiency Improvement	Unnamed [18]	14 weeks	2015
		WorldBuilder	1 day	2017
		Unnamed [40]	1 day	2019
		Laura	4 weeks	2010
		Unnamed (3D) [39]	1 day	2017
Early Depression Detection	N/A	Unnamed [41]	Not Reported	2021
		DEPRA	30 minutes	2021

a. Cognitive Behavioural Therapy (CBT),
b. Post-traumatic Stress Disorder

2.2.1 Psychosocial Intervention

One aspect of the chatbots established so far would be the specific category they belong to. In this section, we will discuss different intervention types and the chatbots designed for them separately. When it comes to Physical Intervention, we can categorize the type into four categories, CBT which refers to Cognitive Behavioural Therapy, Lifestyle Management, Post Traumatic Stress Disorder (PTSD) Management and Work Efficiency Improvement.

2.2.2 CBT

As the demand for medical professionals rises, the high capability of utilizing chatbots to cooperate and aid the medical field in various possible aspects helps to respond to and ease the demand. Regarding the work that was conducted in this area so far, Ly et al [17] researched to measure the effectiveness of smartphone apps in CBT interventions. Shim chatbot considered 28 participants including males and females and it was based on a text-only method of collecting data. The duration of the research was narrowed down to 2 weeks. The research confirmed that participants' experiences and the output of conversation with the Shim chatbot can promote and target the mental health field. The duration of one face-to-face therapy was 60 minutes. The results show that there is no significant inconsistency between blended treatment (including four face-to-face sessions and the smartphone application) and the full BA (including ten -face-to-face sessions and no smartphone application) on neither of the result variances. Both pre and post measurement and follow up actions were considered in this study. The same as Woebot [14] , [18] which offers the CBT, Sharma et al [19] designed a chatbot for CBT. As per the plan, the researchers claim they are targeting 300 million people worldwide for their study. The programming language used by this research is python along with the android mobile operating system. In order to attract more participants, the chatbot is designed as a virtual psychotherapist. Depression level from 0 to 4 was declared with 0 to be completely healthy and 4 highly depressed. The strategy behind the question design is that for any optimistic response the total score will increase by an x greater than zero value. On the other hand, as per any pessimistic response, the total score increases by y greater than zero and less than x value. It can be calculated that for n number of questions the optimistic total score would be $x*n$ and for the same number of questions (n) and pessimistic responses the value would be $y*n$. So, if we have a participant who is suffering from depression the score would be $y*n$ where it can be derived that $x*n>y*n>0$. Any other medical status for any participants would be $x>z>y>0$. The depression levels can be divided into, a) zero depression, absolutely healthy b) stressed c) highly stressed d) depressed e) highly depressed. As the outcome of the research, it wasaffirmed that it is quite difficult to extend therapy chatbot results. The reason behind this evaluation is that the assessments are not managed constantly. Another research discussed Tess [20] as the CBT intervention type. X2AI designed this psychological AI chatbot and the intention was to provide conversations in three formats: mental health, psychoeducation, and reminders. Tess is not intended to replace a real health care professional and it is regarded as a therapeutic resource. Eliza [21], as the first therapeutic chatbot which applied NLP with machines and people in 1966, was the inspiration of Tess in design and implementation. Tess in fact was the version that could adjust to the latest version of the changes in AI memory, and it was removing the limitations that Eliza was facing. Tess was a novel version of a chatbot and it was customizable in a way that the it could join the content to demographics, such as the information of a group of patients, to be linked to the chatbot. One of the methods for data collection in Tess is the use of the PHQ-9 [22] a self-report questionnaire with only 9 questions on it. PHQ-9 questionnaire evaluates depression symptoms in accordance with severity and frequency in the time frame of 2 weeks. Each of the questions includes the Diagnostic and Statistical Manual of Mental Disorders (DSM-4) criteria for major disorder and the scoring is based on 0 (no symptoms at all) to 3 (symptoms appeared every day). The PHQ-9 is one the most recognized and common methods to evaluate the depression level in the patients. If the participate rate between 0 and 5, they are healthy. If they rate between 5 to 9, they are experiencing mild depression. If they rate between 10 to 14, they have moderate depression. If they rate between 15 to 20, they are experiencing moderately severe. Finally, if the rate is greater than 20, they have severe depression. Tess also measure the anxious thoughts by another method called the Generalized Anxiety Disorder 7-item scale (GAD-7) [23]. With the aid of GAD-7, the frequency and severity of anxious reactions is checked within the past 2 weeks period. Similar to DSM-4, GAD-7 will score from 0 (no anxious symptoms) to 3 (anxious symptoms every day) with is the same for all the 7 questions and criteria. Help4Mood [24] is a another communal CBT system in which by the support of an avatar it assists patients under the depression treatment by self-monitoring, self-report and biometric monitoring. A future Randomized Controlled Trial (RCT) was conducted in three locations throughout Europe. Patients with major depressive disorders as well as patients with mild and moderate symptoms were asked to use the system for the duration of 4 weeks. The patients were selected randomly and there was not a specific method to choose the samples. This method is called Treatment As Usual (TAU). Help4Mood was accepted by some of the patients and they interacted with it, however, they were not using it usually as planned by the research team. Those patients who applied and interacted with the system experienced significant changes in depressive moods which supports the idea that Help4Mood is designed and implemented in the way that is acceptable in the medical applications. T-bot [19] is the next therapy chatbot with the capability of curing common and wide spread diseases such as depression. In design process of T-bot the Digitally Advanced Depression Relieving Machine (DADRM) steps are followed to create a chatbot that can deal with major issues. T-bot features are shortlisted: a) acting as a personal assistance, 24/7 availability. b) combination of technology and psychiatry. c) The resemblance of the communication of human being and a machine, it seems that they are identical and the machine conversation cannot be taken out. d) handling the depression the same as a professional health care such as a psychiatrist. e) assisting participants with special care required. The integration of several features in future implementation of T-bot will turn it into a even more robust version of its kind. The T-bot can be integrated with a messaging platform such as Facebook, Telegram, WhatsApp and so forth to be centralized into one platform for easier access and more handy collection of data. Also, the idea of translating the chatbot into several more common languages will expand the area of effectiveness for a more international access by patients around the globe. Sentiment Analysis and NLP can be utilized to the design so the participants can also share their emotions with the bot. Increasing the quality of chatbot in posing the questions and they method the data is collected is another aspect to be examined in the design and implementation process of upgrading the T-bot. The flow of the conversational process can be a helpful aspect to maintain in the future assessments. A constructive learning atmosphere sounds like a new area to examine and it is feasible by applying the technology behind the implementation of a chatbot.

KokoBot [25] is a chatbot based on emotional statements introduced as a corpus of data, that is, Koko data. Koko platform has a mobile implementation which recommends psychological resilience. This platform reflects the idea suggested by web-based Panoply platform that offers methods to mitigate depression signs and symptoms [26]. KokoBot is interacting with its participants through text messaging. The integrations with a wide range of messaging platforms such as Kik, Facebook Messenger and Twitter makes it a remarkable experience by the participants. They can also use the mobile and desktop applications to interact with the KokoBot. All the interactions follow a simple post response format. With this format KokoBot is passing the message between the participants who ask for help while interacting with the bot. Also, there is an interesting option to activate the participants to assist their peers in case they are facing difficulties to deal with the bot. The interaction on the bot are in the format of posts. When interacting with the bot, the participants are encouraged to describe an unsuccessful experience and the feelings that they have due to this unpleasant issue. The participants are familiar with the fact that they can approach the experience with optimistic view and accept the situation in order to suggest a solution for it by responding to that specific post. The number of characters which are applied for posts are 950 and the responses are limited to 600. This feature of character limitation on KokoBot is unique and it distinguishes the bot from the online peer forums such as Reddit or Facebook. Most of the online peer forums do not limit the number of characters in any format, this leads to a hassle when someone would like to respond to the posts which are lengthy. This creates a challenge on the feature as open-ended posts are regarded as more reliable and they provide a robust background on the fact that the participants can elaborate more on their own words about their real feelings and experiences. About 37,169 participants interacted with the bot within one month study of the research. To make the demography of the participants private and confidential, KokoBot refused to ask for personal information of the users such as age, gender or any other sort of information [27].

A set of common chatbots such as Replika, Inner Hour, Woebot, Wysa, Youper, MindDoc and Pocketcoach [28] are capable of any types of conversation as long as the conversation is in short term and the responses of the participants do not increase in length and content. These chatbots can get involved in emotional and spiritual conversations and the participant can face a mutual interaction about any topics. As the Replika is training through the process of conversation and it is basically an AI chatbot, it can be educated by dealing with the participants. Mindspa is a chatbot introduced and designed in a way to handle short conversations. It is applied for emergency requirements only. This chabot does not have a complicated structure and it simply supports conversation to collect the critical information of participants such as their medical history, allergies and so forth. The Waysa has a fascinating feature that it can add a health care professional in the chat with the participants. Woebot [14], Wysa, Youper, MindDoc, DEPRA and Pocketcoach are the chatbots that do not require to create an account in the app that they are hosted on. The other chatbots ask the participants to create an account before they can proceed into the experiment.

In this research, an interview guide, discipline, was not referred to, however, in the DEPRA chatbot we followed the SIGH-D and IDS-C which is a standard guideline throughout DEPRA chatbot research.

2.2.3 Lifestyle Management

Gardiner et al [29] studied urban women to consider the possibility of applying Embodied Conversational Agent (ECA) to implement changes in their lifestyle. The 3D Gabby chatbot was used in this experiment. The participants, mainly women, were divided into 2 categories a) an ECA which included mindfulness, stress management, physical activities during the day and eating habits. b) The same as the first group but with the mediation CD/MP3 to be considered daily for one month. A total of 61 women participated.

More research is conducted by Bickmore et al [30] with the aid of Laura 3D chatbot. Antipsychotic medicine adherence and change in lifestyle of patients suffering from schizophrenia was examined. The methodology of this research was based on Coyle and Doherty which is the mental health guide integrated with technology. Moreover, the approach of collaboration between the specialists and mental health staff who are practising in the research projects was applied. As the result, the research team includes computer professionals, psychiatric nurses, a psychiatrist, and a biostatistician. Regarding the participants of this study, 20 participants in the age range of 19 to 58 years old enrolled. Out of this range, 67% were females, all single and never got married, 78% were African American and 89% were unemployed. The acceptance to participate in the system was considerably high, with only one participant avoidance to be part of the system, one required to be reoriented and 15 completed the task with no doubts or further assistance. The ratings of satisfaction of the system were 4.5/5, easiness to use the system was 4.3/5, the willingness of continuation to use the system was 4.4/5. Consequently, these statistics suggest that relational agents are trustworthy and reliable assessors to consult with patients who are suffering from schizophrenia. By the end of the research, some women reduced the average level of drinking alcohol in order to control their stress level and they managed to consume more fruits in their daily diet.

Within another research by Bickmore et al [31], the Laura chatbot was designed to use a 3D environment for a patient's hospital discharge plan. The goal was to consider how hospitalized patients would react to a chatbot when they were at the stage of being discharged from the hospital. Moreover, this chatbot was utilized to target patients who are experiencing a high level of depression interacted with the bot. A total of 131 patients were examined and as they interacted with the Laura chatbot for their discharge process, they declared a high rate of satisfaction regarding the easiness of using the chatbot, fluency of receiving discharge information from the agent compared to the medical staff at the hospital. Besides, patients who were suffering from depressive disorders ranked the chatbot remarkably higher on therapeutic alliance. It was concluded that people who need care in an inpatient environment will get the most out of conversational agents in the areas of assessment, education and counselling.

Also, Podrazhansky et al [32] designed a chatbot mobile app for mental health. It is meant to collect data from the users in natural ways such as text, voice, or video. ML algorithms along with NLP, neural network, computing methodologies, information systems and mobile information processing systems applied to locate a mental disorder and offer prevention methods. For text-based chatbot, the researchers designed their practical chatbot and they did not use an existing chatbot with the aid of the Dialogflow. Dialogflow is used for design and implementation to expedite the process and to avoid inconsistencies with the available chatbot. With this strategy the boundaries of limiting the participants to avoid predefined responses and by analysing the participants' responses the allocation of the mood to the participants were more practical. The role of NLP is vital as the research has removed unwanted letters and changed the orientation of letters to all lowercase. By analysis of the participants' responses, it is possible to realize which words are the most common ones and the calculation of the frequency is at hand. They used a profile to reflect the details related to each participant while collecting and analysing the data for text, audio and video. The traditional Patient Health Questionnaire (PHQ-9) survey which pertains to 9 depression questions to monitor the presence and severity of depression in patients was applied to the participants. A neural network is utilized to collect data from various resources. In order to compute the participant's mood, the second to last layer includes pessimistic and optimistic measures relevant to each emotion, that is, sick is pessimistic and successful is optimistic. At the final stage, the results were summed up and the overall mood was considered to be positive or negative. It is possible to conclude if a participant's mood has increased or decreased by the final stage of the process.

Besides, Sharma et al [33] introduced a chatbot designed to predict and manage stress. Based on questions and responses on the chatbot, the stress can be addressed within this process. A feature of this chatbot will be the integration with a website that has been designed for this purpose. For this study, they have used the python and Rasa framework and applied other tools which were required for this research. Continued stress can be the main cause of some serious illnesses such as headaches, heart attacks and depression. As a result, it is essential to find the stress level of the participants and avoid putting them at the risk of more challenging diseases. Within a survey conducted by Open Sourcing Mental Illness (OSMI), it proved that employees in high tech and IT corporations were in danger of establishing stress-related symptoms more than any other employees. As a resolution to this situation, it was mentioned that more realistic Human Resources (HR) policies and regulations can contribute to the employees in high tech companies. Among the tools used by the study, the decision tree was a robust tool for classifying and considering the prediction phase. The hierarchical structure of the decision tree makes it possible to visualize the entire structure of the stress management process. With this approach, this study managed to illustrate depression and its relevant levels. The paper suggests that future works can include the study of more diseases such as anxiety, insomnia, and so forth. Also, a couple of other datasets would be applied in future words.

In another study ViviBot [34] and its effects on psychological factors on young adults suffering from cancer was discussed. A group of adults aged between 18 and 29 years were involved within the period of five years to interact with active cancer treatment which was deployed over the Facebook messenger. The participants were divided into two major groups: a) experimental group and b) control group. In experimental group, the participants were randomly selected to have immediate access to the Vivibot content or to have daily access to spiritual ratings. In control group, the full access was authorized to participants to check the chatbot full content after four weeks' time. ViviBot is designed as a human-based chatbot to deal with young adults suffering from cancer which includes optimistic psychological skis, daily status evaluations, educational material from the supervisors and usual feedbacks to the participants. On weeks 2, 4, and 8 a survey was circulated with all the participants to assess the mental well-beings. Chatbot involvement and open-ended feedback prepared a basis to analyze the participants well-being and how helpful the process was for experimental group and control group. This analysis checked the level of anxiety and depression and emotional changes within the groups for the period of four weeks commencing from the day one of the experiment. A further follow-up analysis conducted in control group within the period of four weeks to eight week, to reveal the side effects of interacting with VivoBot. ViviBot provides a CBT structure to escalate optimistic emotion constructed by Moskowitz et al [35].

In a study by Harilal et al, the features of CARO [36] chatbot was discussed. CARO is a chatbot which can perform emotional conversations along with medical consultation for patients who suffer from major depressive disorders. This chatbot has a capability to extract content of a conversation and analyze the intention by the text and derive the emotional concept from the text. Then, CARO consult the participant with an emotional message or advise them to seek a health care professional. The feature of CARO is that it is the first attempt within related work to bridge between emotional message to the participants and the medical advice at the same time. To reach the goals of this approach, the researchers trained two models with two separate datasets and they considered the intent to reunite the outcomes of these two models.

BAES [37] which stands for Bot – Autonomous Emotional Support is a chatbot that can increase the mental health of the people suffering from depression. There are well known chatbots such as Apple Siri, Google Allo, Microsoft Cortana and so forth that can interact with the users and assist with responses which are practical and useful. However, they apply NLP and do not have a system to deal with emotional responses from the users. Moreover, Woebot, Pepper.ai, Wysa, Joy and Evei are the set of chatbots that support the emotional correspondents and they are accessible through the market. BAES is also an example of emotional supportive chatbots that can elevate the senses of its users. As a feature of BAES it is open source and it has the capability of sensing the way the users communicate with the chatbot over a period of time. BAES gathers information from various sources, database and websites and it matches the text input with the gathered data. As soon as more data is collected, BAES improves the method of interacting with the users by directing them through the most tangible way. This chatbot is available in two languages, English and Hindi-English. This characteristic makes it available to more users and participants to derive a more successful outcomes. BAES is considered as an online assistant to the public so people who are suffering from depression but they are not aware of it as a serious disease will have the opportunity to access the chatbot and have a review over their mental health conditions. This chatbot can be cheap and effective compared to the expensive psychological sessions run by medical health professionals such as psychiatrists and psychologists. When people feel lonely and they are seeking help over the internet they will reach out for online chatbots such as BAES and this way it would be promoted to potential users. The status of the users which is reflected on social websites or a direct response on the chatbot helps to improve the way the chatbot can assist the users to be escalated emotionally and they feel satisfied with their mental health.

2.2.4 PTSD Management

In another study, Tielman et al [38] promoted a therapy system, 3MR_2, that includes a digital diary and a 3D tool, WorldBuilder, where patients can recreate their traumatic memories intending to treat Post Traumatic Stress Disorder (PTSD). Due to the cost and stigma involved in this treatment, numerous patients avoided attending the medical sessions. 3D WorldBuilder virtual agent provides a basis to support the patients with at-home therapy sessions. This system also includes a virtual agent employing ontology-based questions to guide patients throughout their sessions. A live therapist is only to monitor the progress remotely. As for the participants, there were 4 in the focus group, 2 males and 2 females. All of the participants had previous PTSD follow-up therapies. A 3MR_2 system was introduced for PTSD disorder sufferers. As patients work alone with the 3MR_2 without any direct supervision, it is not recommended for patients with depression and a history of suicide. In addition, the digital agent unlike a therapist has limitations to steer a conversation and bring patients back on track during a therapy session. That said, despite the limitations, the 3MR_2 system can have a noticeable effect on societies around the world. Another phase of the Tielman et al [15] research focuses further on PTSD If this disease is diagnosed at any stage, there are positive approaches that it can be cured. However, due to stigma and cost most of the patients avoid consulting with a health care professional. This research is promising in the aspect of providing therapy at the ease of the patients' homes. They can monitor their progress and a psychiatrist might be required remotely in minor cases to check the progress and encourage the participation by consulting the patients. The participants utilize the virtual daily diary to record their experiences and then they simulate the traumatic experience in 3D WorldBuilder tool. The purpose behind the 3D WorldBuilder tool simulation is that the experience can be rebuilt as completely as possible with most of the details so the mental health specialists can trigger the point which caused the damage to the patients. The system and the way it dealt with the symptoms creates hopes that it is practical and effective to assist the recovery of its users along with PTSD treatments sessions. This is basically a novel method to treat patients with home therapy with the aid of a virtual agent.

Lucas et al [39] mention two studies with the same 3D chatbot. This chatbot provides a consultant to deal with military veterans with PTSD symptoms. They had the conversation in an autonomous method to keep their identities private. Military veterans with distinctive symptoms could not participate in the experiment as they were suffering from serious side effects of PTSD. For this group, the virtual simulation was so traumatic that they could not recover from the process.

2.2.5 Work Efficiency Improvement

Another study by Tielman et al [40] revealed the fact that whether verbal or text-based psychoeducation is more effective for improving adherence where text-based agents result in higher adherence than voice-based agents. In mental health this psychoeducation can be targeted with two approaches, by an embodied tool through voice and verbal communication, or by textual content. A sample of 46 participants were recruited to collect their writings on their worst experiences in a virtual diary after receiving psychoeducation recommendations before commencement of writing. The results showed that textual psychoeducation had a priority to vocal approach. The participants were selected from university staff and students who completed the writing task. All the participants were from Germany, English was their second language of communication. The details of the experiment were shared with all the participants including signing the consent form prior to the participation. Only one participant withdraws from the experiment which is not included in the 46 statics of participants.

Moreover, Shinozaki et al [18] conducted a research on IT employees in a study to analyse anxiety and emotional problems that they face as the consequence of the low success rate of development projects in the IT industry. As a result, counselling became a significant issue for the staff in this field. A context-based conversational agent was designed to replace live counsellors serving IT professionals. Results indicate that the Counselling Agent (CA) was two times more popular to interact with compared to ELIZA-style conversational agents. This popularity along with context-based feature of the CA increases the user friendliness of the bot compared to similar chatbots offered by other studies and research manuscript.

2.2.6 Depression Detection

A study by Kharel et al [41] researched early detection of depression with the aid of ML that was performed on patients' MRI images. The Functional Magnetic Resonance Imaging (fMRI) dataset consisted of brain images of 30 participants who were on antidepressants or CBT. Due to limitations in the small dataset for the experiment, the use of ML algorithms and features were restricted. It was shared that Deep Learning algorithms and in particular Convolutional Neural Networks (CNN) were the most popular methodology to analyse medical images. Also, Support Vector Machines (SVM) was applied in order to provide a measure and value of accuracy. This accuracy targeted the standard size of a dataset. Classification method of facial processing as well as developing depression neurobiological patterns were the initial purpose of this study. Cerebral formation over the entire surface of the brain in healthy participants was examined. The idea behind this was to predict neurocognitive in a healthy population. In the data collection phase, 38 participants were recruited. Half of the individuals (N=19) were suffering from depression; however, they were not on any medications, another half (N=19) were healthy populations willing to follow the subjects and instructions of the study. The fMRI data were utilized at each stage of the process with the aid of an SVM pattern. As a result, the work in the area of early detection of depression remains an intact base of knowledge that requires further work and study. They utilized methods to collect data through which are not currently applied in other studies such as audio, video and even movement.

Finally, Philip et al [42] conducted research with 179 participants for a period of 1-day participation. A 3D chatbot was used with consideration of Embodied Conversational Agents (ECAs) which are software with high potential but they have not been used more often in the studies of mental disorders. Initially, 221 outpatients were consulted, however, only 179 were interested to participate in the data analysis phase. Exactly 90 participants experienced the medical interview with the specialist, namely psychiatrist, before the actual session with ECA and the rest 89 participants were visited by the psychiatrist after the interview with the ECA. The average age was 46.5±12.9 years old, females formed 57.5% of the participants and the average educational level was 13.3±3.0. Also, 35 participants, 19.6%, were diagnosed with Major Depression Disorder (MDD) under the confirmation of the psychiatrist. On the other hand, in the group that was consulted by the psychiatrist before the ECA session, the average age was 45.7±13.6 years old, females were 55.6% of the participants' ratio and the average educational level was 13.1±3.2. In the group which was examined by the psychiatrist after the ECA session, the average age was 47.3±12.1 years old, females formed 59.6% of the participants and the average educational level was 13.1±2.9. As per the information, the two groups are almost identical and there are no major differences between them. Furthermore, the ECA interview revealed the statics as Mild Depression (58%), Moderate Depression (77%) and Severely Depression (84%). There were two goals for this study: a) testing efficiency of diagnostic system related to Major Depressive Disorders (MDD). b) evaluation of acceptability on ECA.

The conclusion of this study suggests that ECAs have a high potential to be utilized for standard and well-performed applications in the interviews.

To conclude, DEPRA also falls under this bot family category as the Depression Detection Bot which is to circumvent some of the existing challenges that have been discussed in this section. The main difference between this research and the related works conducted so far is that DEPRA chatbot gathers the data based on a series of psychologically approved questions and its purpose is to triage and detect the early signs of depression. However, the available chatbots mostly focus on the depression disorder itself and they are trying to find a better cure. In other words, DEPRA chatbot intents to assist the medical professionals to detect and cure the depression at its initial formation and it does not intent to replace a medical professional. DEPRA was implemented to further investigate the application of AI to interact with users, collect their responses to questions on depression symptoms and identify depression and estimate the level of severity.

- DEPRA is capable of assisting participants to find out about their mental health and medical professionals to realize their role in achieving the cure for the patients who are suffering from depression.
- It has been implemented following a conversation flow based structured early detection depression interview guide for the Hamilton Depression Scale (SIGH-D) and Inventory of Depressive Symptomatology (IDS-C) which is derived from the experience of a group of psychiatrists. This makes DEPRA chatbot a unique chatbot that shares the same stem of clinically proved questions.
- The validity of the questions used in the DEPRA chatbot has been measured by applying closed group participation. This is arranged by designing open-ended questions for the participants. Regarding social media, DEPRA is benefiting from integration with Facebook Messenger. So, the participants can interact with the chatbot through social media to share their responses.
- DEPRA takes approximately 30 minutes to complete. As the results suggested there are high

satisfaction rates reported by the participants in relevance to the responses. The questions were easy to comprehend and respond to, not as time consuming as a real psychiatrist session, text messaging via social media platforms has a higher degree of preference compared to talking in a consultation session.

CHAPTER 3 RESEARCH DESIGN

3.1 Introduction

In this chapter, we intend to find the answer to this question: “How a chatbot can be utilized to conduct a structured interview for human depression rating?”. The participants of DEPRA chatbot have the opportunity to be looked upon with care so they can report their daily and weekly moods to assist the researchers and, as a result, their personal wellbeing with a series of clinically proven enquiries. The conversational bots, as the virtual assistance of mental health, are recently receiving remarkable consideration [14] by the societies as they bring three major advantages to the medical world: a) reducing the gap between the available resources compared to the number of people who are suffering from depression and mental illnesses, b) mitigating the risk caused by the social stigma, and c) addressing the fear of being judged by others.

Although chatbots are commercially advertised to the customers, we know that academic psychiatric study is dramatically in short supply. More studies should be conducted in this area of knowledge. As the concept of AI and the technology beneath it are in its initial formation it is considered that more studies are required to prove the possibility of applying these emerging technologies at a more practical level. By using chatbots or any conversational agents the risk of self-harm is reduced to a remarkable level. For example, a study conducted in controlled lab settings concerning the use of smartphones to respond to a crisis such as suicide [43]. The results were not accurate and in some cases they were improper. The chatbots on the smartphones were useless as they could only provide web searches and helpline details.

With DEPRA, it removes the reliance on a short set of multiple-choice responses and includes open-ended responses for the participants to explain their emotions and feelings without any limitations. The conversations are designed as per the SIGH-D and IDS-C [44]. SIGH-D and IDS-C are based on medical psychiatrist achievements where a set of questions are shared with the patients to rate the severity of their depression by probing mood, feelings of guilt, suicide ideation, insomnia, agitation or retardation, anxiety, weight change, and somatic symptoms [44]. The privacy of the participants and any future patients is precisely secured. Only the researcher and the supervisory team have access to the collected data. The dataset is backed up by an R drive and it is stored in a safe place to avoid access by intruders. As per the privacy policy, we considered to de-identify the data related to our participants. To manage this, patients' personal information was limited to only age and gender identifiers and responses were collected in different tables in the database and stored separately in a secured password-protected drive. De-identification and separate storage of identification and relevant data that only authorized people to have access to make it a trustworthy source for this study. Furthermore, the data that was used to generate the statistical details, such as graphs, charts and tables, relevant to the participants were all based on de-identified forms of the collected data.

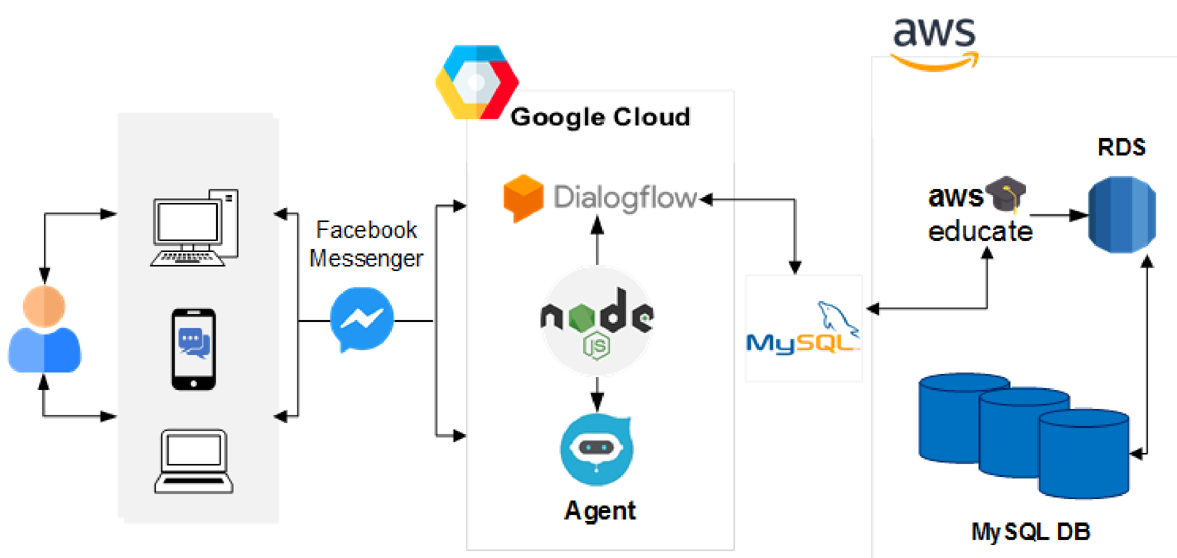


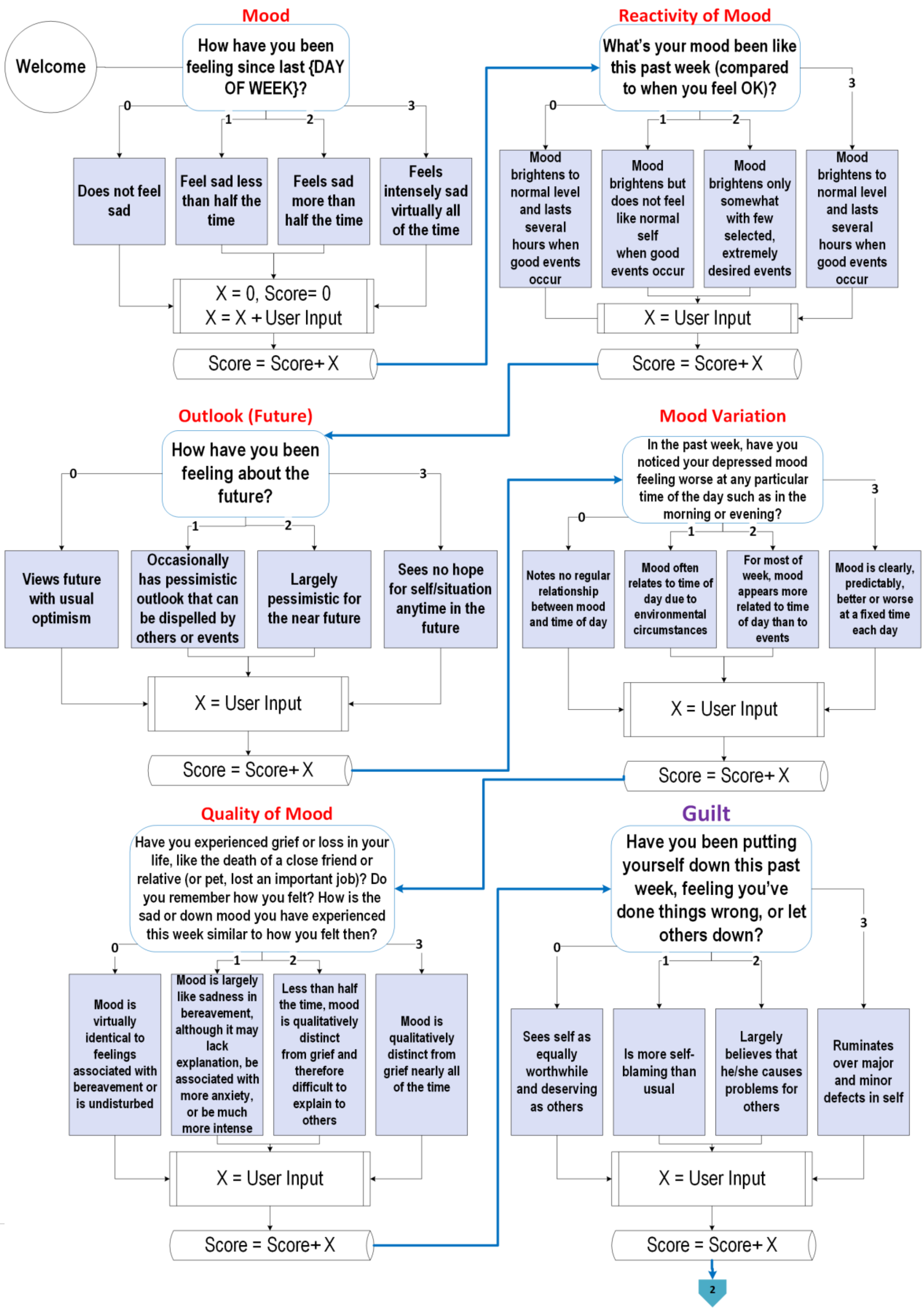
Figure 1 - DEPRA chatbot Design and Implementation Schema

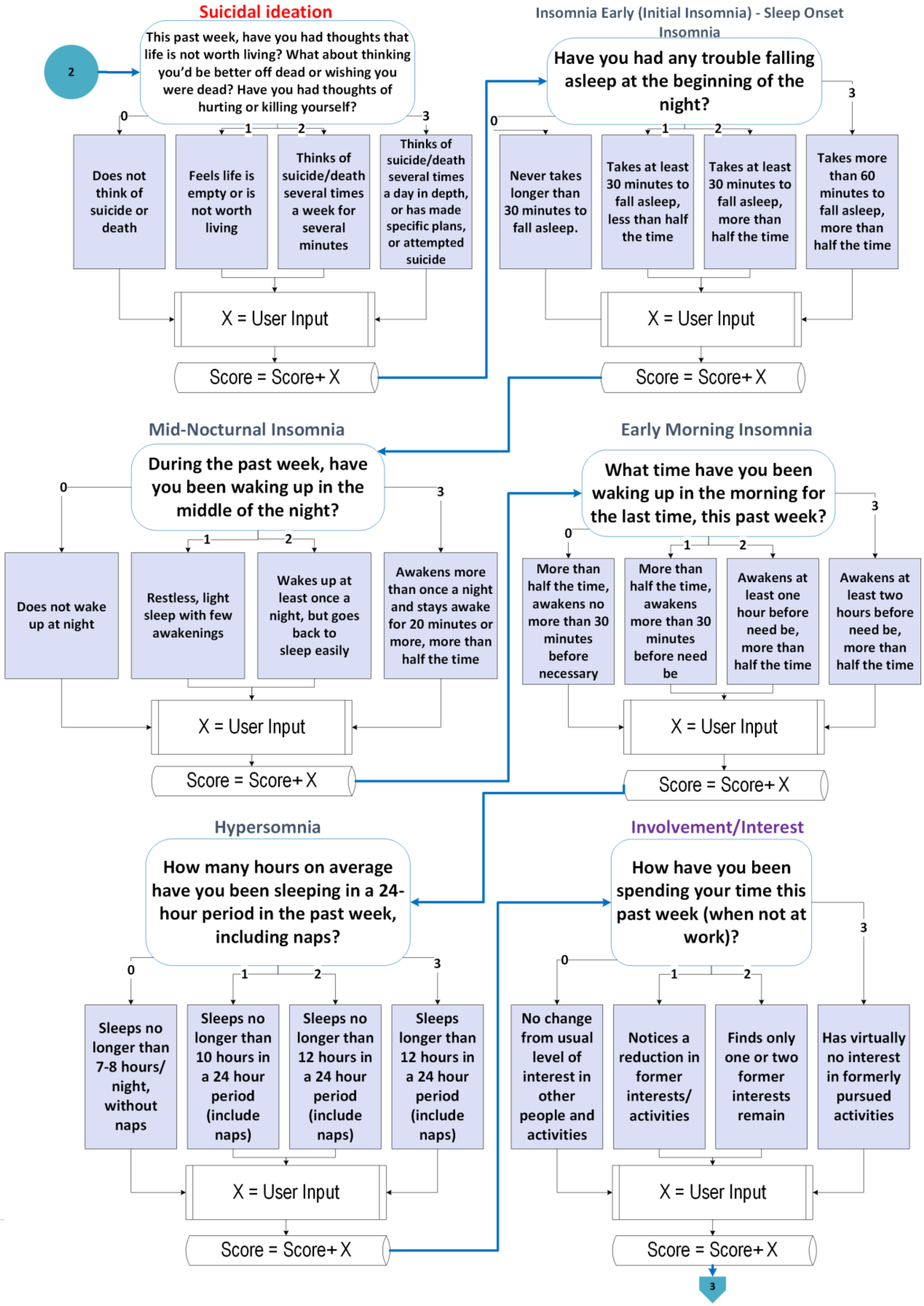
3.2 Methodology and Conceptual Framework

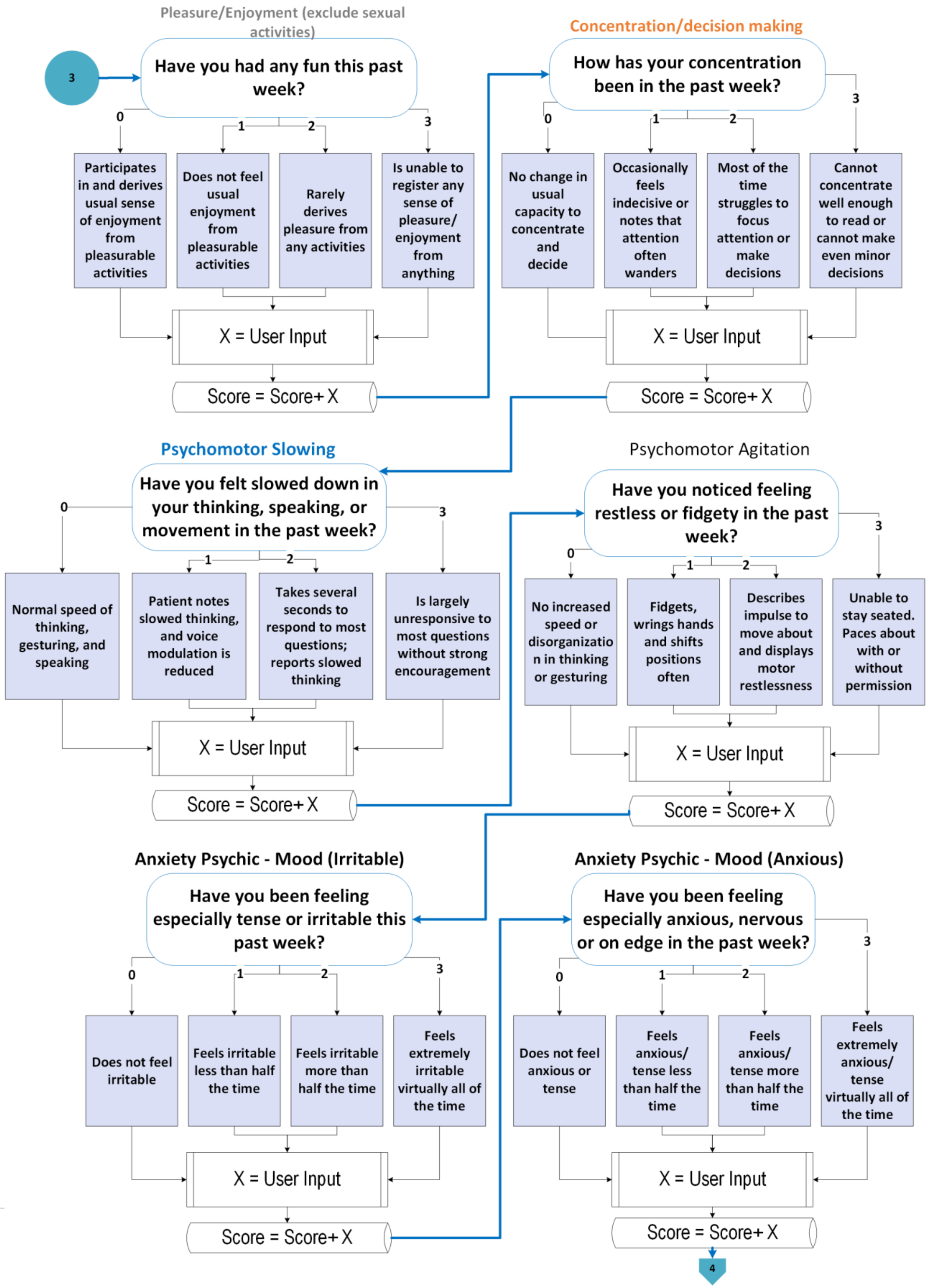
In this section, we have designed the DEPRA as an agent under Dialogflow. The participant has access to its smartphone, laptop, desktop or other devices. Facebook Messenger is a connection between the participant and the agent. In order to establish the connectivity, the agent, in this case DEPRA, which has been designed in Dialogflow under Google Cloud and Node.js receives a record from the participant by Facebook Messenger. At the backend, the agent will save the data into AWS RDS database. In order to create and modify the tables in the database, we have applied MySQL Workbench client. The AWS RDS is the core database that saves all the records when the interaction takes place between the agent and the participant. Figure 1 summarizes the general trend of the process.

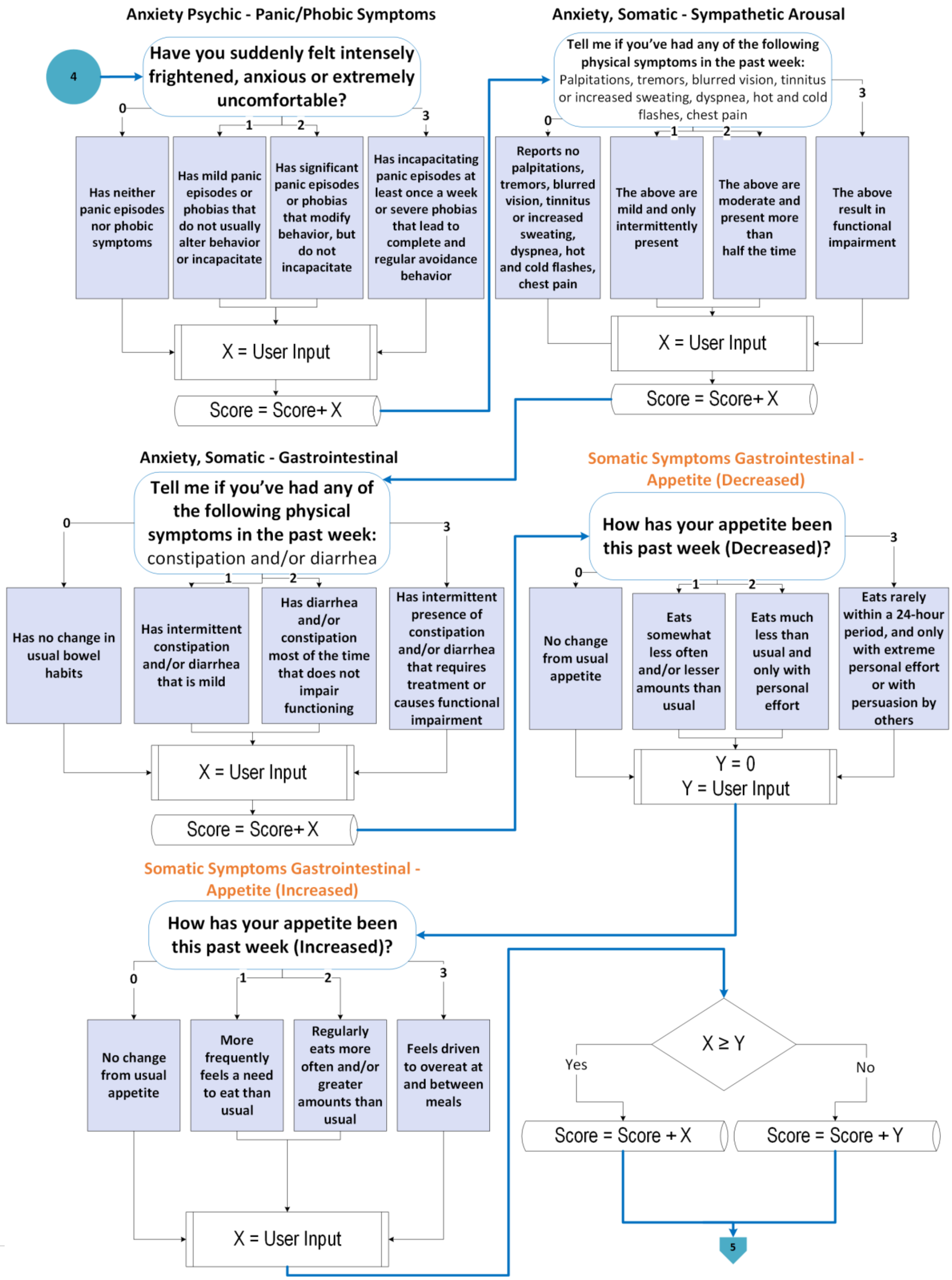
3.2.1 DEPRA Conversation Flow Guidelines

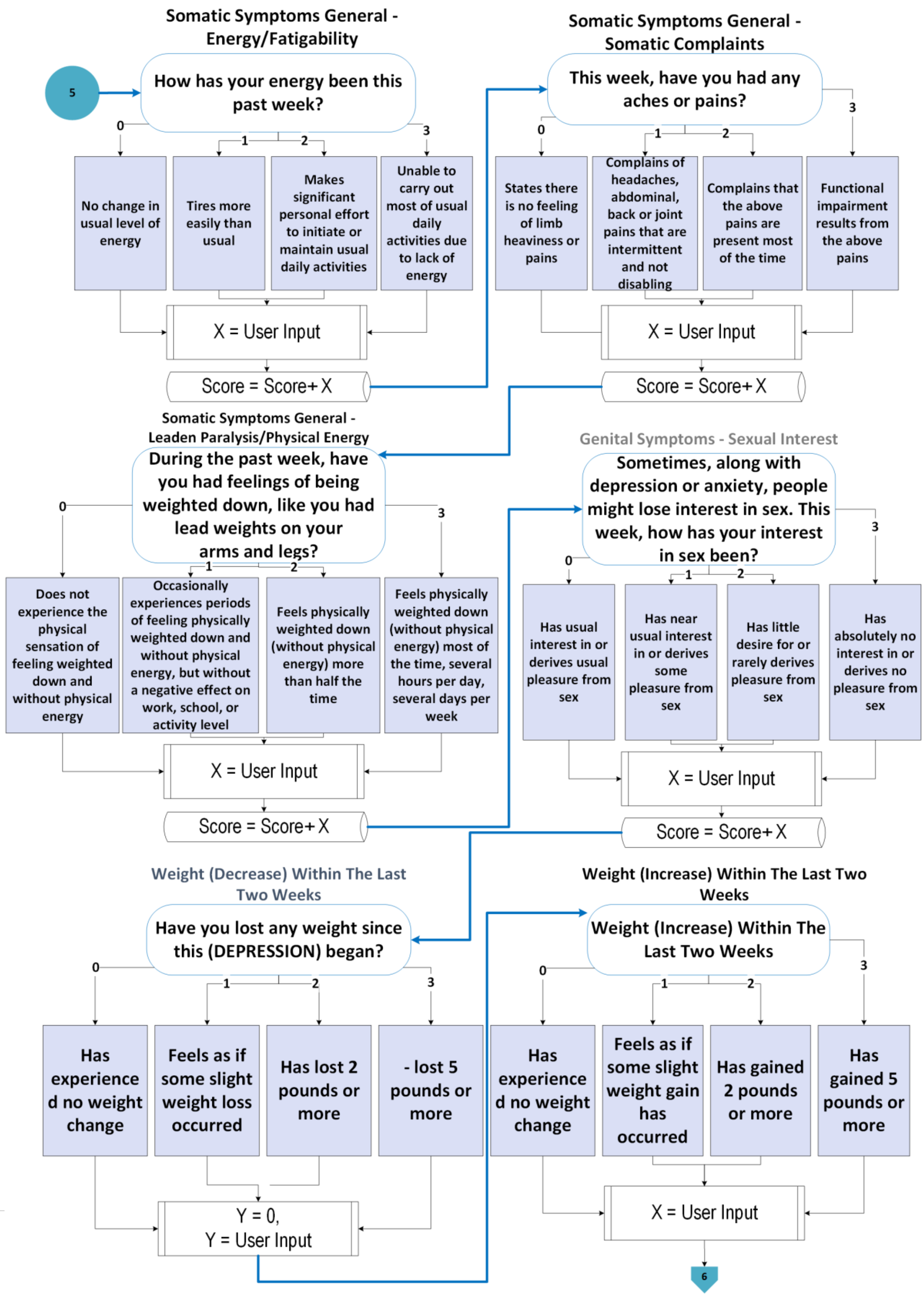
DEPRA is implemented on the Dialogflow platform as a conversational environment for communication and interaction between the bot and the participants. The analysis of the responses from the potential patients and the participants, in general, makes it possible to have a review of the participants and derive the symptoms rating scale out of a set of responses. In order to design the conversation, we developed a conceptual model with 27 psychometric questions. In the case of DEPRA chatbot, we have integrated the design into Facebook Messenger. There are several options to choose from, for example, Dialogflow Messenger, Web demo and so forth. Moreover, the flowchart in Figure 2 illustrates the design of conversational flow and how the sequence of questions deal with the depression symptoms such as mood, reactivity of mood, outlook, guilt, insomnia, suicidal ideation, anxiety, involvement, concentration, psychomotor, somatic, symptoms, general, pleasure, weight, interpersonal sensitivity and genital symptoms. There are 17 mental conditions as per the Hamilton Depressive guideline.


Welcome
Mood
How have you been feeling since last {DAY OF WEEK}?
0
1
2
3
Does not feel sad
Feel sad less than half the time
Feels sad more than half the time
Feels intensely sad virtually all of the time
X = 0, Score= 0
X = X + User Input
Score = Score+ X
Reactivity of Mood
What's your mood been like this past week (compared to when you feel OK)?
0
1
2
3
Mood brightens to normal level and lasts several hours when good events occur
Mood brightens but does not feel like normal self when good events occur
Mood brightens only somewhat with few selected, extremely desired events
Mood brightens to normal level and lasts several hours when good events occur
X = User Input
Score = Score+ X
Outlook (Future)
How have you been feeling about the future?
0
1
2
3
Views future with usual optimism
Occasionally has pessimistic outlook that can be dispelled by others or events
Largely pessimistic for the near future
Sees no hope for self/situation anytime in the future
X = User Input
Score = Score+ X
Mood Variation
In the past week, have you noticed your depressed mood feeling worse at any particular time of the day such as in the morning or evening?
0
1
2
3
Notes no regular relationship between mood and time of day
Mood often relates to time of day due to environmental circumstances
For most of week, mood appears more related to time of day than to events
Mood is clearly, predictably, better or worse at a fixed time each day
X = User Input
Score = Score+ X
Quality of Mood
Have you experienced grief or loss in your life, like the death of a close friend or relative (or pet, lost an important job)? Do you remember how you felt? How is the sad or down mood you have experienced this week similar to how you felt then?
0
1
2
3
Mood is virtually identical to feelings associated with bereavement or is undisturbed
Mood is largely like sadness in bereavement, although it may lack explanation, be associated with more anxiety, or be much more intense
Less than half the time, mood is qualitatively distinct from grief and therefore difficult to explain to others
Mood is qualitatively distinct from grief nearly all of the time
X = User Input
Score = Score+ X
Guilt
Have you been putting yourself down this past week, feeling you've done things wrong, or let others down?
0
1
2
3
Sees self as equally worthwhile and deserving as others
Is more self-blaming than usual
Largely believes that he/she causes problems for others
Ruminates over major and minor defects in self
X = User Input
Score = Score+ X
2


Suicidal ideation
2
This past week, have you had thoughts that life is not worth living? What about thinking you'd be better off dead or wishing you were dead? Have you had thoughts of hurting or killing yourself?
0
1
2
3
Does not think of suicide or death
Feels life is empty or is not worth living
Thinks of suicide/death several times a week for several minutes
Thinks of suicide/death several times a day in depth, or has made specific plans, or attempted suicide
X = User Input
Score = Score+ X
Insomnia Early (Initial Insomnia) - Sleep Onset Insomnia
Have you had any trouble falling asleep at the beginning of the night?
0
1
2
3
Never takes longer than 30 minutes to fall asleep.
Takes at least 30 minutes to fall asleep, less than half the time
Takes at least 30 minutes to fall asleep, more than half the time
Takes more than 60 minutes to fall asleep, more than half the time
X = User Input
Score = Score+ X
Mid-Nocturnal Insomnia
During the past week, have you been waking up in the middle of the night?
0
1
2
3
Does not wake up at night
Restless, light sleep with few awakenings
Wakes up at least once a night, but goes back to sleep easily
Awakens more than once a night and stays awake for 20 minutes or more, more than half the time
X = User Input
Score = Score+ X
Early Morning Insomnia
What time have you been waking up in the morning for the last time, this past week?
0
1
2
3
More than half the time, awakens no more than 30 minutes before necessary
More than half the time, awakens more than 30 minutes before need be
Awakens at least one hour before need be, more than half the time
Awakens at least two hours before need be, more than half the time
X = User Input
Score = Score+ X
Hypersomnia
How many hours on average have you been sleeping in a 24-hour period in the past week, including naps?
0
1
2
3
Sleeps no longer than 7-8 hours/ night, without naps
Sleeps no longer than 10 hours in a 24 hour period (include naps)
Sleeps no longer than 12 hours in a 24 hour period (include naps)
Sleeps longer than 12 hours in a 24 hour period (include naps)
X = User Input
Score = Score+ X
Involvement/Interest
How have you been spending your time this past week (when not at work)?
0
1
2
3
No change from usual level of interest in other people and activities
Notices a reduction in former interests/ activities
Finds only one or two former interests remain
Has virtually no interest in formerly pursued activities
X = User Input
Score = Score+ X
3


Pleasure/Enjoyment (exclude sexual activities)
3
Have you had any fun this past week?
0
1
2
3
Participates in and derives usual sense of enjoyment from pleasurable activities
Does not feel usual enjoyment from pleasurable activities
Rarely derives pleasure from any activities
Is unable to register any sense of pleasure/ enjoyment from anything
X = User Input
Score = Score+ X
Concentration/decision making
How has your concentration been in the past week?
0
1
2
3
No change in usual capacity to concentrate and decide
Occasionally feels indecisive or notes that attention often wanders
Most of the time struggles to focus attention or make decisions
Cannot concentrate well enough to read or cannot make even minor decisions
X = User Input
Score = Score+ X
Psychomotor Slowing
Have you felt slowed down in your thinking, speaking, or movement in the past week?
0
1
2
3
Normal speed of thinking, gesturing, and speaking
Patient notes slowed thinking, and voice modulation is reduced
Takes several seconds to respond to most questions; reports slowed thinking
Is largely unresponsive to most questions without strong encouragement
X = User Input
Score = Score+ X
Psychomotor Agitation
Have you noticed feeling restless or fidgety in the past week?
0
1
2
3
No increased speed or disorganization in thinking or gesturing
Fidgets, wrings hands and shifts positions often
Describes impulse to move about and displays motor restlessness
Unable to stay seated. Paces about with or without permission
X = User Input
Score = Score+ X
Anxiety Psychic - Mood (Irritable)
Have you been feeling especially tense or irritable this past week?
0
1
2
3
Does not feel irritable
Feels irritable less than half the time
Feels irritable more than half the time
Feels extremely irritable virtually all of the time
X = User Input
Score = Score+ X
Anxiety Psychic - Mood (Anxious)
Have you been feeling especially anxious, nervous or on edge in the past week?
0
1
2
3
Does not feel anxious or tense
Feels anxious/ tense less than half the time
Feels anxious/ tense more than half the time
Feels extremely anxious/ tense virtually all of the time
X = User Input
Score = Score+ X
4


Anxiety Psychic - Panic/Phobic Symptoms
4
Have you suddenly felt intensely frightened, anxious or extremely uncomfortable?
0
1
2
3
Has neither panic episodes nor phobic symptoms
Has mild panic episodes or phobias that do not usually alter behavior or incapacitate
Has significant panic episodes or phobias that modify behavior, but do not incapacitate
Has incapacitating panic episodes at least once a week or severe phobias that lead to complete and regular avoidance behavior
X = User Input
Score = Score+ X
Anxiety, Somatic - Sympathetic Arousal
Tell me if you've had any of the following physical symptoms in the past week:
Palpitations, tremors, blurred vision, tinnitus or increased sweating, dyspnea, hot and cold flashes, chest pain
0
1
2
3
Reports no palpitations, tremors, blurred vision, tinnitus or increased sweating, dyspnea, hot and cold flashes, chest pain
The above are mild and only intermittently present
The above are moderate and present more than half the time
The above result in functional impairment
X = User Input
Score = Score+ X
Anxiety, Somatic - Gastrointestinal
Tell me if you've had any of the following physical symptoms in the past week:
constipation and/or diarrhea
0
1
2
3
Has no change in usual bowel habits
Has intermittent constipation and/or diarrhea that is mild
Has diarrhea and/or constipation most of the time that does not impair functioning
Has intermittent presence of constipation and/or diarrhea that requires treatment or causes functional impairment
X = User Input
Score = Score+ X
Somatic Symptoms Gastrointestinal - Appetite (Decreased)
How has your appetite been this past week (Decreased)?
0
1
2
3
No change from usual appetite
Eats somewhat less often and/or lesser amounts than usual
Eats much less than usual and only with personal effort
Eats rarely within a 24-hour period, and only with extreme personal effort or with persuasion by others
Y = 0
Y = User Input
Somatic Symptoms Gastrointestinal - Appetite (Increased)
How has your appetite been this past week (Increased)?
0
1
2
3
No change from usual appetite
More frequently feels a need to eat than usual
Regularly eats more often and/or greater amounts than usual
Feels driven to overeat at and between meals
X = User Input
X ≥ Y
Yes
No
Score = Score + X
Score = Score + Y
5


Somatic Symptoms General - Energy/Fatigability
5
How has your energy been this past week?
0
1
2
3
No change in usual level of energy
Tires more easily than usual
Makes significant personal effort to initiate or maintain usual daily activities
Unable to carry out most of usual daily activities due to lack of energy
X = User Input
Score = Score+ X
Somatic Symptoms General - Somatic Complaints
This week, have you had any aches or pains?
0
1
2
3
States there is no feeling of limb heaviness or pains
Complains of headaches, abdominal, back or joint pains that are intermittent and not disabling
Complains that the above pains are present most of the time
Functional impairment results from the above pains
X = User Input
Score = Score+ X
Somatic Symptoms General - Leaden Paralysis/Physical Energy
During the past week, have you had feelings of being weighted down, like you had lead weights on your arms and legs?
0
1
2
3
Does not experience the physical sensation of feeling weighted down and without physical energy
Occasionally experiences periods of feeling physically weighted down and without physical energy, but without a negative effect on work, school, or activity level
Feels physically weighted down (without physical energy) more than half the time
Feels physically weighted down (without physical energy) most of the time, several hours per day, several days per week
X = User Input
Score = Score+ X
Genital Symptoms - Sexual Interest
Sometimes, along with depression or anxiety, people might lose interest in sex. This week, how has your interest in sex been?
0
1
2
3
Has usual interest in or derives usual pleasure from sex
Has near usual interest in or derives some pleasure from sex
Has little desire for or rarely derives pleasure from sex
Has absolutely no interest in or derives no pleasure from sex
X = User Input
Score = Score+ X
Weight (Decrease) Within The Last Two Weeks
Have you lost any weight since this (DEPRESSION) began?
0
1
2
3
Has experience d no weight change
Feels as if some slight weight loss occurred
Has lost 2 pounds or more
- lost 5 pounds or more
Y = 0,
Y = User Input
Weight (Increase) Within The Last Two Weeks
Weight (Increase) Within The Last Two Weeks
0
1
2
3
Has experience d no weight change
Feels as if some slight weight gain has occurred
Has gained 2 pounds or more
Has gained 5 pounds or more
X = User Input
6

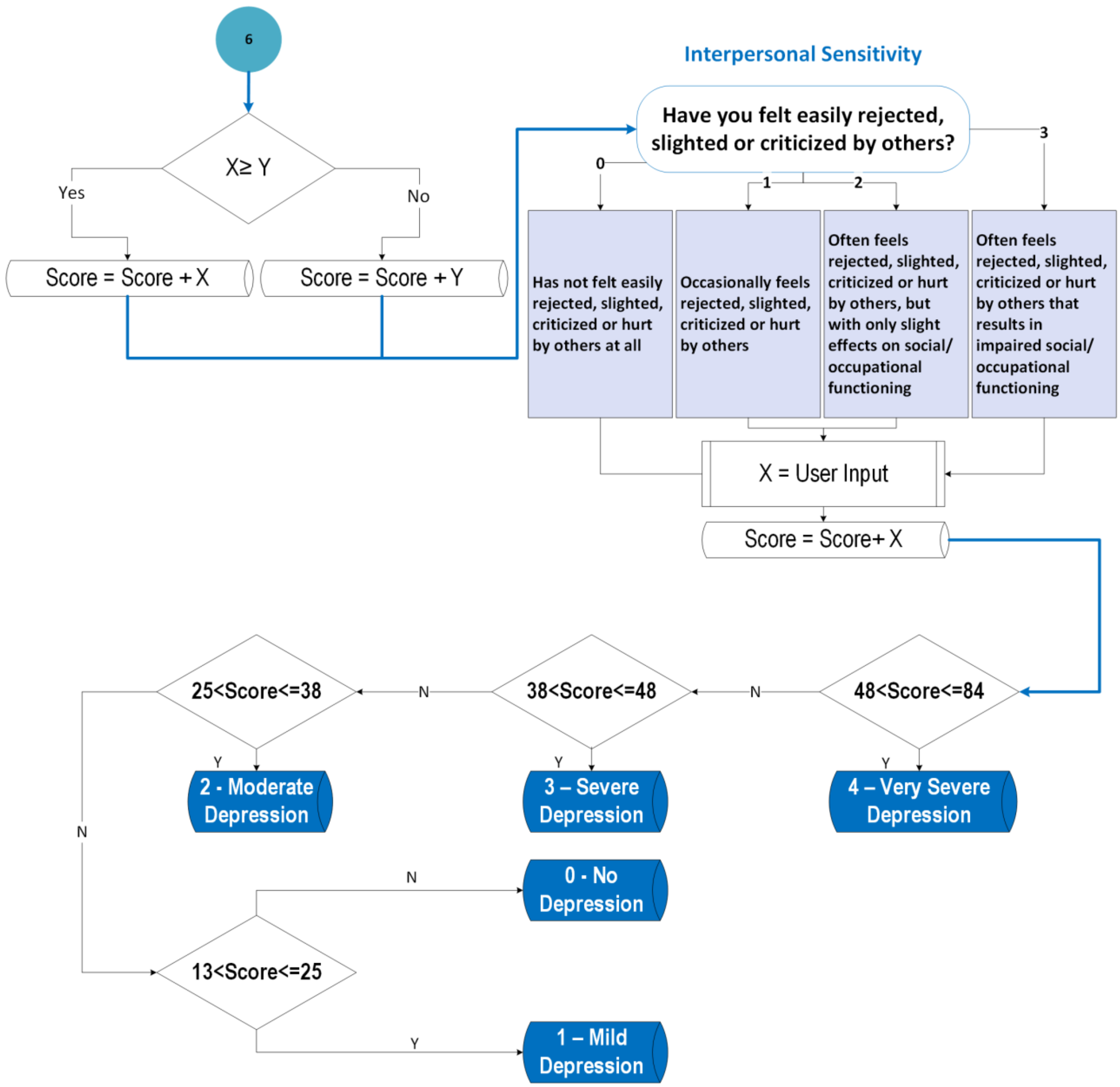


Figure 2 - Conceptual flowchart of Depressive Symptoms

The sequence commences with a question: “How have you been feeling since last week?”. Then, according to the response, a value between 0 to 3 allocates to the score of this specific question. The range of the scores are graded according to 0 for unlikely and 3 for very severe. The score would be saved in a parameter and the trend continues to the next questions. Then, the same scoring method will be applied from 0 to 3 range. When all the 27 questions are answered by the participant for each symptom, the overall score will be calculated. If the overall score is less than 13, there would be no depression for the participant. If the score is between 13 and 25, mild depression will be generated. The range between 25 and 38 demonstrates moderate depression. Moreover, the score range between 38 and 48 generates severe depression. Lastly, if the score ranges from 48 to 84 very severe depression will be demonstrated. This scoring system which is applied in the Figure 2, the flow chart on the bottom of it, is based on IDS-SR scoring system.

3.2.2 DEPRA Contemporary Platform

For this study, we have implemented the DEPRA chatbot with the aid of Dialogflow contemporary platform. Dialogflow is a NLP platform. This environment provides a robust basis to design, integrate and connect a user interface into a public accessed tool such as a mobile application, chatbot, a web page and more. Both text messaging and voice recognition can be utilized in Dialogflow and it has the capacity to analyse various types of inputs. The method of response by Dialogflow can also be through text messaging or vocal interaction. There are two versions of services through this platform, CX and ES. Dialogflow CX is designed for complex agents while Dialogflow ES is dealing with small and basic agents. In order to use and apply the Dialogflow ES or CX, there is a set of documentations provided. These documentations, which are similar for both of these versions, are called editions, support and resources. Throughout this study, we used the Dialogflow ES version.

Most of the chatbots available today are designed for therapeutic purposes. However, DEPRA chatbot is designed with the focus on depression detection at early stages. Moreover, the participants of this research who interact with the DEPRA chatbot are not aware of their depression level at first. Although, through mass screening, DEPRA will find out if any individual is suffering from depression. Then, the level of the depression will be revealed at the end of the interaction. If they have depression at any level, they will be advised to visit a health care professional for further assistance. DEPRA is an exceptional tool to be used in mass screening for early depression detection. DEPRA is designed based on a structured early detection depression standard interview guideline that is similar to a professional psychiatrist. This standard interview guide is SIGH-D and IDS-C. In order to make this guide more practical we have limited the number of questions asked into a fewer number of enquiries in the DEPRA chatbot design. However, the basics are all followed according to this guide and it is applied as a map. The intention of applying the structured guide is to impersonate a psychiatrist's triage session. By this approach, the DEPRA chatbot closes the gap of requiring more health care professionals at the times of emergencies when there are not sufficient professionals available.

The concepts of Dialogflow ES are agents, intents, entities, contexts, follow-up intents, Dialogflow console, integrations and fulfillment. We are briefly introducing each concept and explaining the reason that they are playing a significant role in the design and implementation processes.

A Dialogflow agent is a digital assessor that interacts between the users and the system which can be a chatbot. The agent is a module with the potential to comprehend and analyse human beings natural language. Dialogflow receives the text or vocal messages from the user and converts the message into a format that is understandable by the designed application. This application is designed and implemented through Dialogflow to serve our requirements in any environments that we need them for assessment. The agent in Dialogflow is similar to an agent that has been trained in a call centre. This training is meant to address the predictable conversation flow and the training should not be entirely explicit.

Regarding the Dialogflow intent, it is applied to divide the user's intention so the user will wait for the next round of conversation. Many intents are defined for a specific agent to form the comprehensive conversation structure. As the user interacts with the system by entering a text message or by voice, also known as end-user expression, Dialogflow comes to action and matches the end-user expression with the most probable intent in the agent. This process is called intent classification. Figure 3 displays an overall view of the intents defined for DEPRA chatbot. We did not include all the intents in this view and it only shows the very first intents which are defined and used such as welcome and end intents.

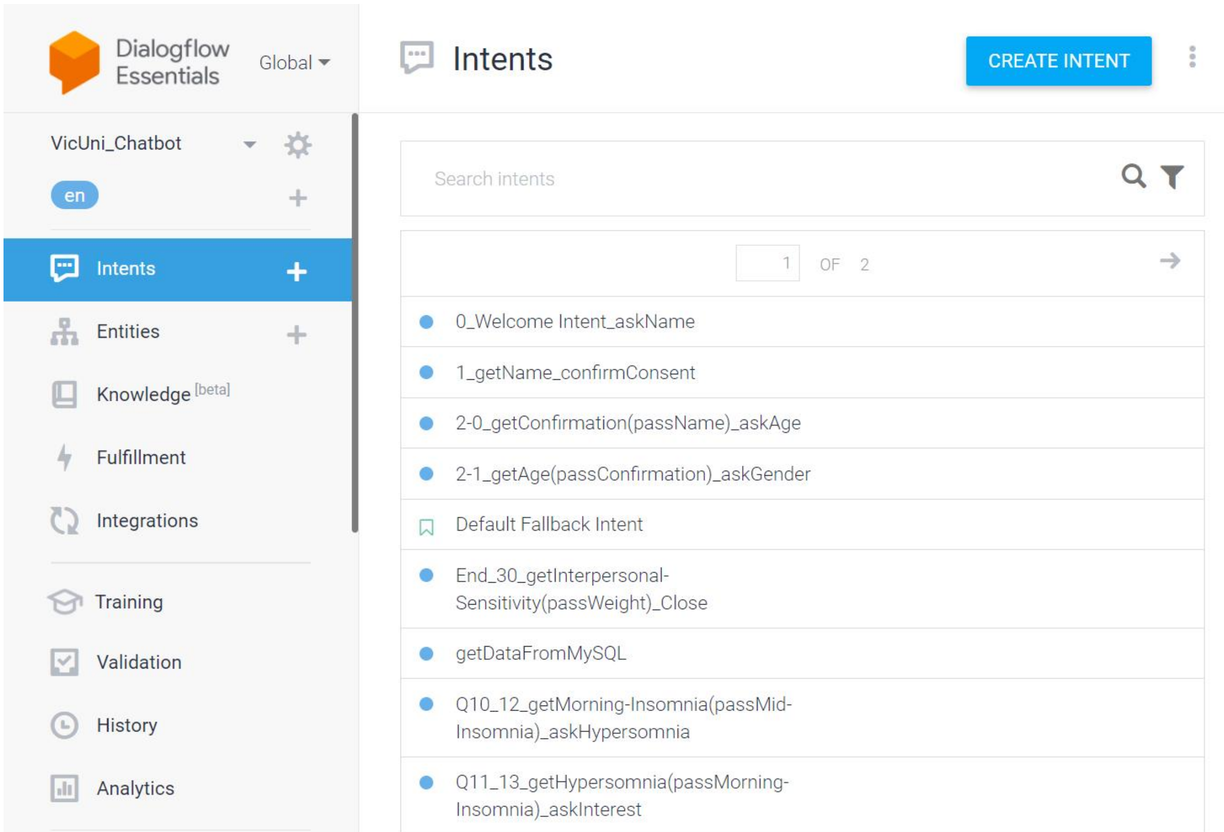


Figure 3 - Dialogflow Intents Window

An initial and primitive intent pertains to training phrases, action, parameters and responses. To implement the conversation in Dialogflow, DEPRA defines total of 38 intents including HDRS questionnaires, welcome messages, personal identifications, confirmation of consent form, database transmissions and other system intents. Table 2 includes the intents breakdown on how Dialogflow is utilized for DEPRA chatbot implementation.

Table 2 - DEPRA Chatbot Dialogflow Intents Breakdown

Type	Number of Intents
HDRS Questionnaires	27
Welcome, age and gender	3

Name and Consent form confirmation	1
Fallback	1
Final intent	1
Database & MySQL Transactions	5
TOTAL	38

Training phrases include the terms and phrases predicted by the system designer. These training phrases share the possible responses that a user might choose them in the interaction with the system. In case of matching the end-user expression with any of these training phrases, Dialogflow will call the relevant intent with matches to the expression in the best way. The internal ML mechanism designed in Dialogflow makes it possible that the designer of the system to avoid all training phrases that can be predicted. It will extend the list of training responses to collect the probable matches. Figure 4 shows a visual display of the training phrases window on Dialogflow.

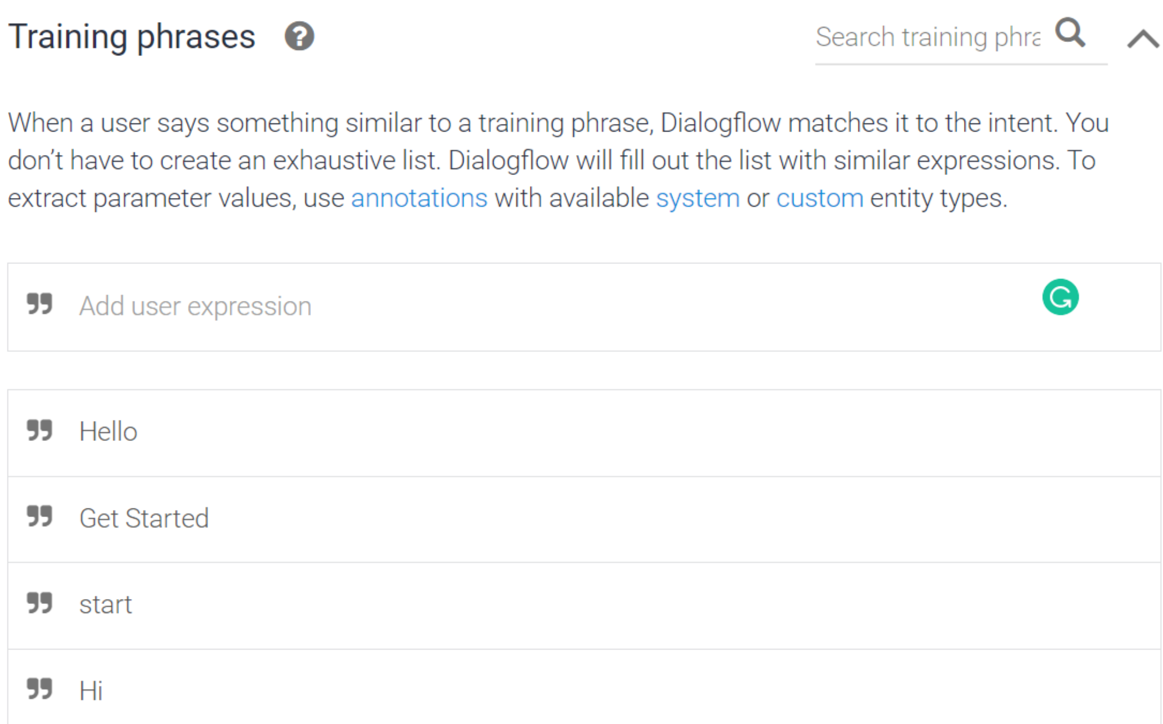


Figure 4 - Dialogflow Training Phrases Section

For each intent, we can define an action. As the intent matches, the action related to it would be called by Dialogflow. An action can be utilized to catch other actions defined in the system. Moreover, parameters are extracted values in the end-user expression. They are targeted when an intent performs in the runtime process. There is a value called entity type which is based on the fact that each parameter includes a type. The entity type indicates the way the data has been derived. Parameters are considered as structured data and they can lead to logic and generating of meaningful responses. Figure 5 is a sample of action and parameters section in Dialogflow.

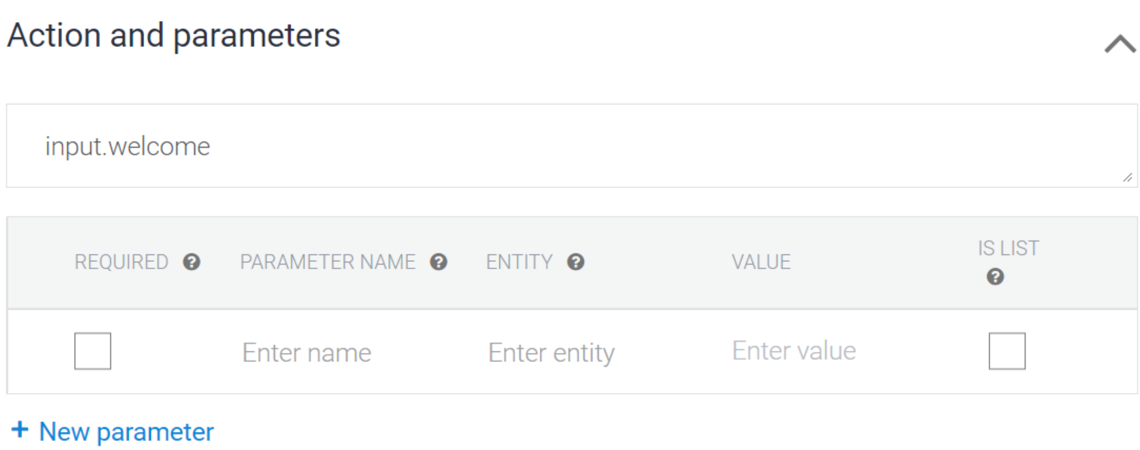


Figure 5 - Dialogflow Action and Parameters Section

Responses are in the format of text, voice or visual which are returned to users. They might trigger a request for more information, ask for termination of the interaction or answer to the users. Figure 6 is a demonstration of how responses are displayed in Dialogflow.

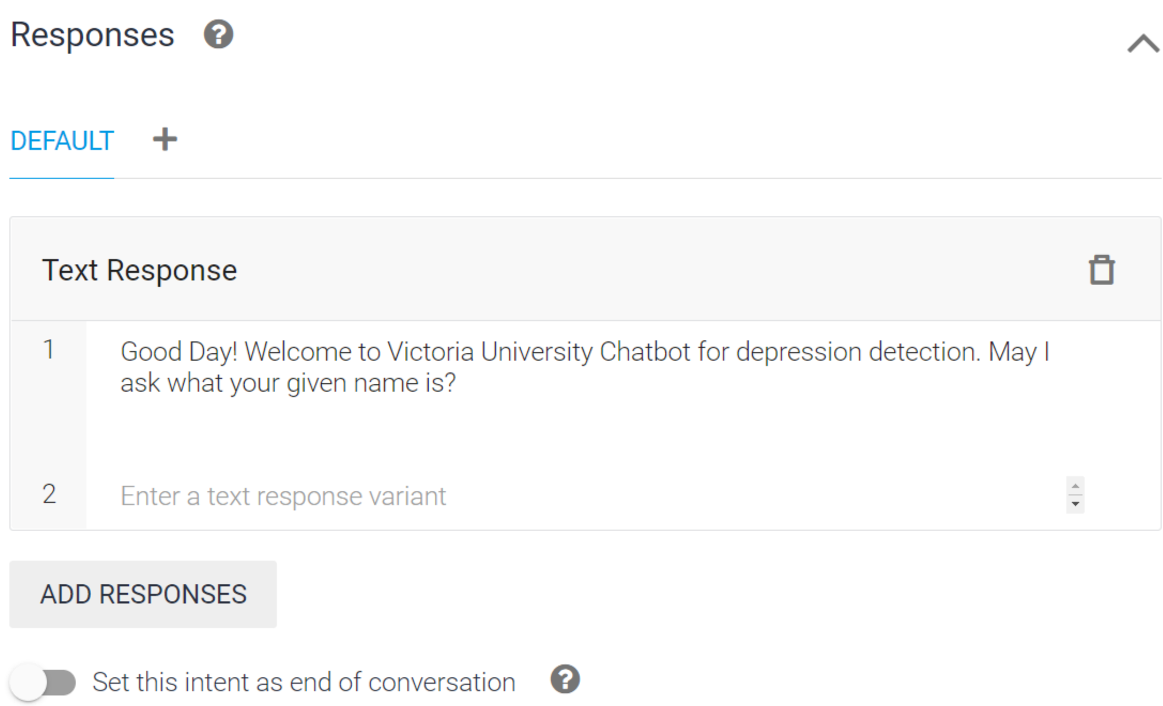


Figure 6 - Dialogflow Responses Section

Regarding the entities, Dialogflow includes two types of entities, system entities and custom entities. System entities are predefined and they can be applied in any system such as dates, times, email addresses and so forth. Custom entities, as the name suggests, can be defined and allocated to the system by the designer team. For instance, we can define a food entity that matches all kinds of food such as fast food, pizza, homemade and so forth. Figure 7 partially shows the entities that are defined for DEPRA chatbot system.

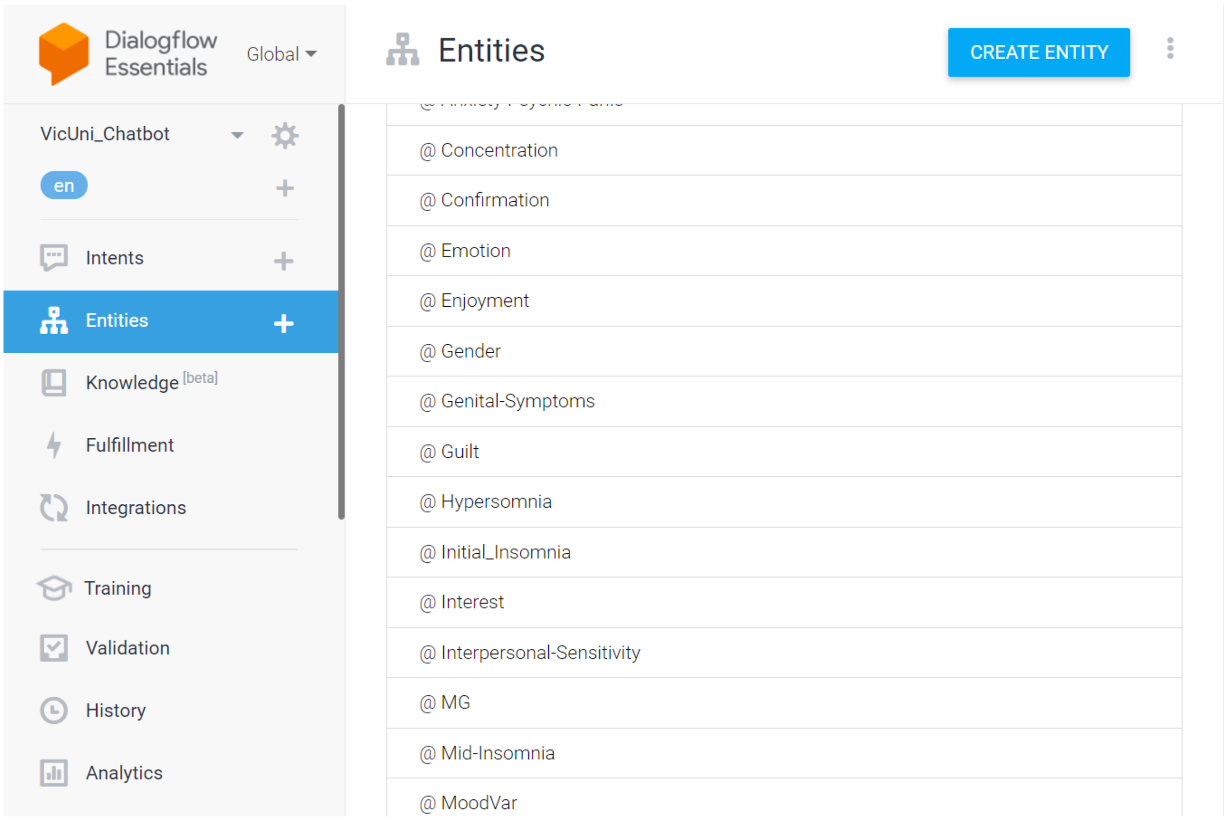


Figure 7 - Dialogflow Entites defined for DEPRA

As for the contexts, Dialogflow is similar to natural language context and structure. If for instance, a user mentions "we are the champions", we need the context to make sure what the "we" refers to. So, to address the end-user expression, Dialogflow requires a context to match with a corresponding intent. Basically, the contexts in Dialogflow are defined to control the flow of the interaction. Input and output contexts are used to control this flow by string names allocated to each of them separately. In this process, when an intent is matched any predefined output contexts, the relevant contexts to that intent will be activated. As long as a context is active, the matching intents with input contexts may also be activated. Figure 8 shows the dialogbox for entering the input and output contexts for the getName_confirmConsent intent in DEPRA chatbot system.

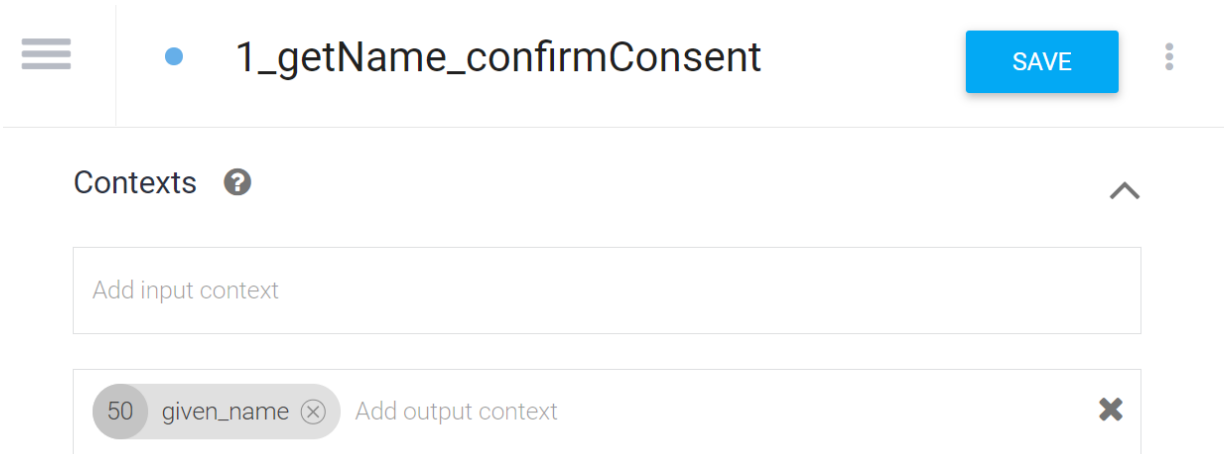


Figure 8 - Dialogflow input and output Contexts

There is a concept of Dialogflow which deals with contexts set up in a way that two intents can pass the contexts automatically, this is called follow-up intents. It is the child of its parent. When the parent intent matches the previous status, the follow-up intent would be activated. Follow-up intents can be designed in a hierarchical order. Dialogflow supports a wide range of common end-user expressions such as "yes", "no" or "cancel", however, the designer can define and construct its own follow-up intents to manage these sorts of responses.

The web user interface implemented by Dialogflow is called Dialogflow console. This is the area where the designers can create, manage and test their agents. As the name suggests, it can be applied for the managerial purposes of the desired systems. Google Cloud Platform (GCP) console and Dialogflow console are not the same concepts. GCP console is allocated to GCP-specific Dialogflow settings such as billing and other resources while Dialogflow console is allocated to design, build, manage and implement agents. For more professional design of agents the Dialogflow API is recommended, however, Dialogflow console can be utilized for any basic and initial management of agents. Figure 9 displays an overview of Dialogflow console which has been applied for DEPRA chatbot project.

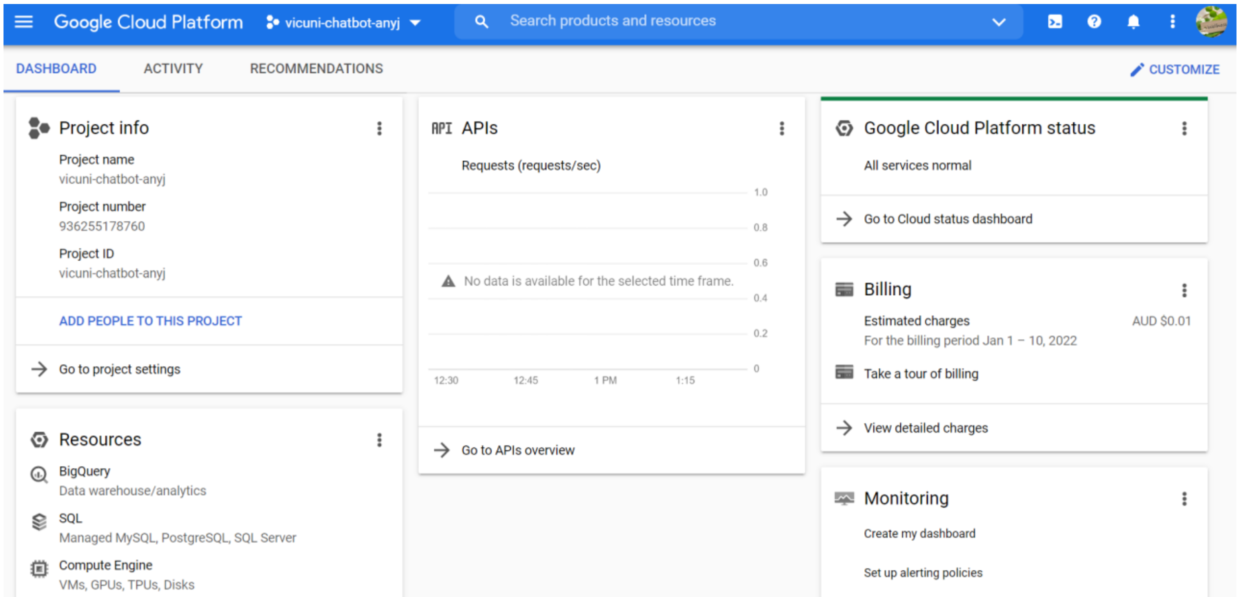


Figure 9 - Overview of Dialogflow Console

For DEPRA chatbot, we have applied integrations with Facebook Messenger. However, the concept of integrations of Dialogflow includes a wide range of environments and applications such as Google Assistant, Slack, Telegram, Web Demo, Line and so forth. The advantage of handling the integrations by Dialogflow is that the designer can focus on managing the agent rather than get involved to connect and integrate the system with an environment. Figure 10 provides an overview of the integrations section on Dialogflow including the option to connect into Facebook Messenger. Figure 11 shows the DEPRA chatbot integration with Facebook Messenger.

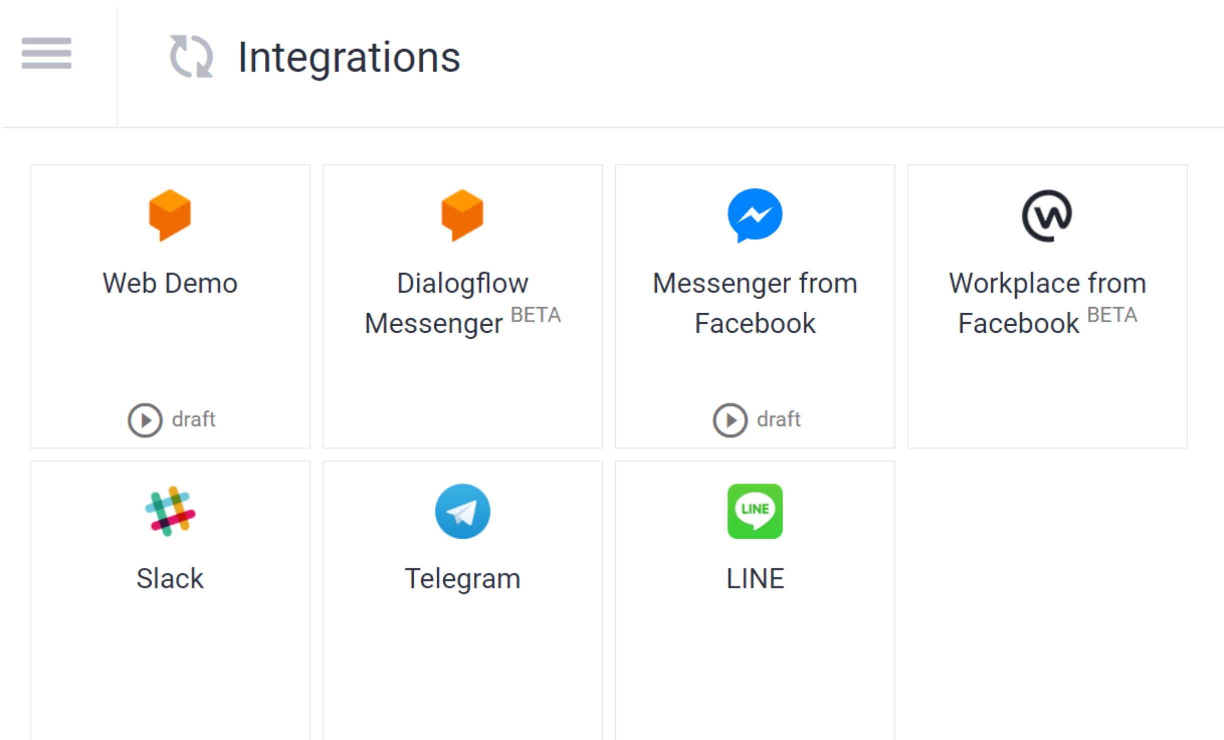


Figure 10 - Overview of the Integrations Section on Dialogflow

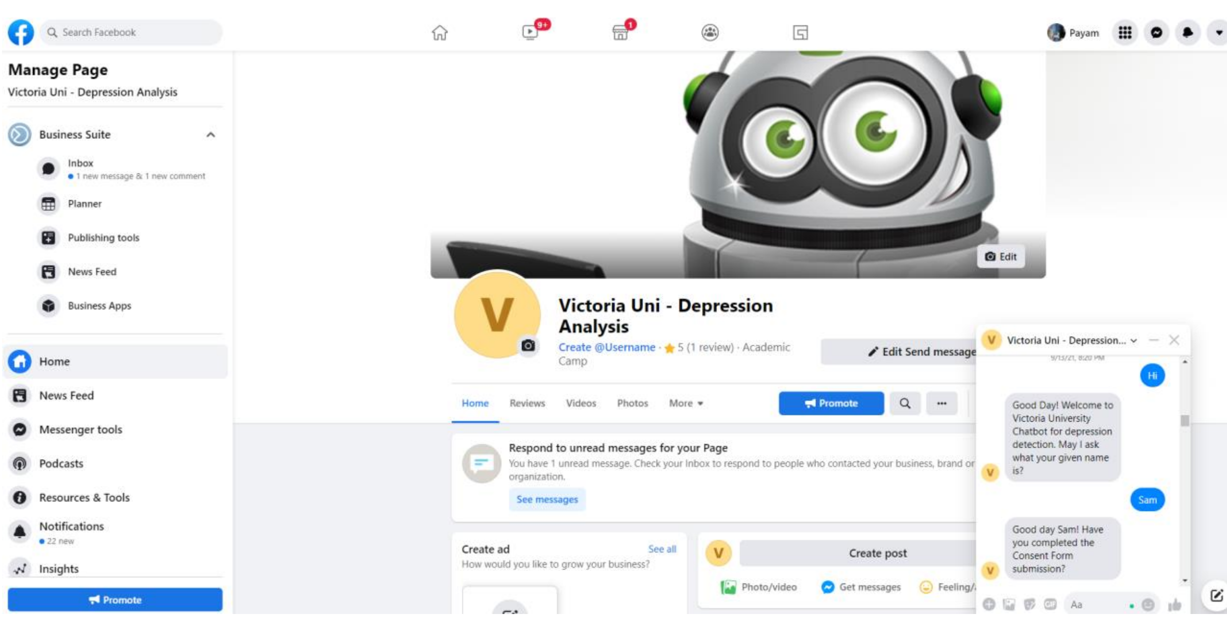


Figure 11 - DEPRA Chatbot Integration with Facebook Messenger

The default setup of Dialogflow is based on the agent communicating with intent by a static response. By applying fulfillment, dynamic responses can be added to the system. Fulfillment is the coding background of the designed bot or any conversational agents that creates an environment to generate codes that are essential for the system to make the project work in the way that is intended by the designer. Fulfillment includes the concept of webhook, which is disabled by default, and an inline editor powered by Google Cloud Functions with the Node.js set of coding. We have designed functions for DEPRA chatbot project and the code controls various aspects of the design such as the limitation of the age group who can participate in the survey, the establishment of a connection between the chatbot and the database and the way the agents are sorted for a sequential trend between responses by the users. Figure 12 displays the fulfillment section on Dialogflow and the index.js file content.

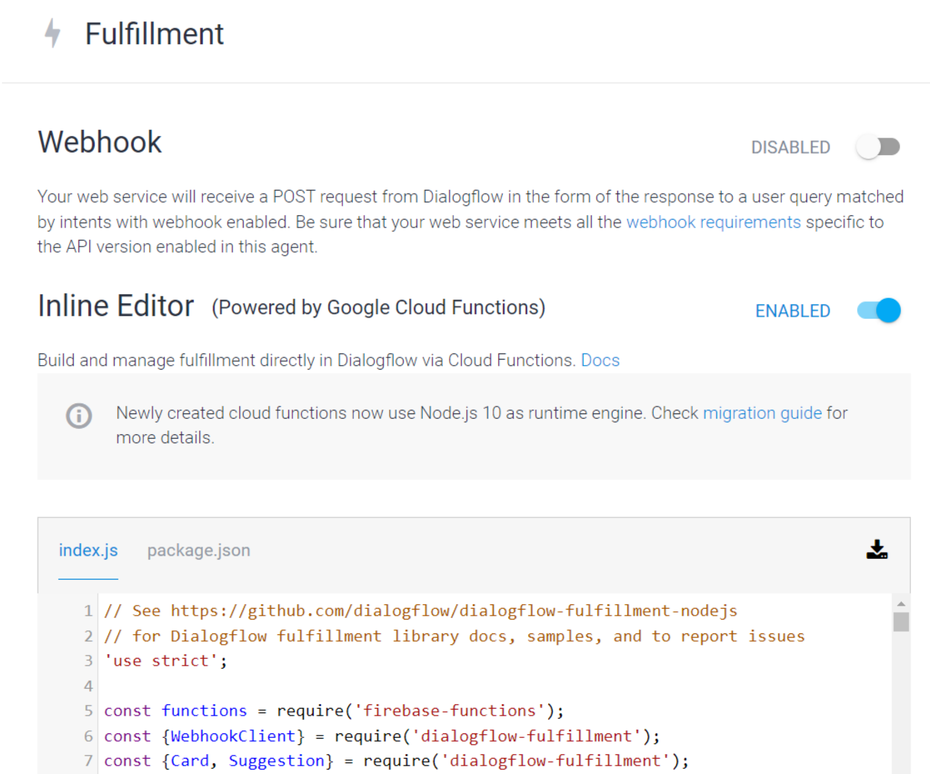


Figure 12 - Fulfillment Section on Dialogflow

This environment and its features are used and we created several functions to support our design. For example, a function has been designed to control and check the age group who are eligible to participate. If the user is over 80 years old or they are below 18 years old, a message will be generated and let the user know that they are not allowed to interact with the chatbot. One of the main functions that we have in fulfillment is the function that enters a record into the database as the user reaches the final question. When the user reaches the final stage, a message with greetings for participation as well as a link to the ranking form will be displayed. There would be a record regarding the successful insertion of the data into the database. Following code represents a function in fulfillment that performs the age checking of the participants.

```
function checkAge(agent) {
  const age = agent.parameters.age.amount;
    if (age > 99) {
            agent.add(`I am affraid! `+ age +` doesnot seem right. How old are you?`);
    } else if (81 < age && age < 100){
            agent.add(`Sorry you cannot participate! Only users under 80 can participate.`);
    } else if (age < 18){
            agent.add(`Sorry you cannot participate! Only users over 18 can participate.`);
    } else {
            agent.add(`What is your gender?`);
    }
}
```

There are 7 functions designed for this research. Here is the list of functions defined for this research in fulfillment:

1. function connectToDatabase()
2. function insertIntoDatabase_tbl1(connection, data_tbl1)
3. function insertIntoDatabase_tbl2(connection, data_tbl2)
4. function insertIntoDatabase_tbl3(connection, data_tbl3)
5. function insertIntoDatabase_tbl4(connection, data_tbl4)
6. function checkAge(agent)
7. function handleWritetoMySQL(agent)

The entire DEPRA chatbot fulfillment Node.js code is as per Appendix A.

3.2.3 Database

The structure of our database design is based on two facts. First of all, we required a central AWS RDS database with the capability of being connected through generating the fulfillment code. Secondly, in order to design the required tables in a client, we utilized MySQL Workbench which is a robust client environment for projects like DEPRA chatbot. Furthermore, we had to create a connection between the AWS RDS and the MySQL Workbench client. As soon as a record is entered into one or all of the tables which are discussed below, the AWS RDS database is updated through the connection defined for it access.

For the design phase of creating the tables, we have prepared a schema, vicreseach_Schema, and within this section we have defined our tables. There are four tables for this study, namely, Personal_Information, Response, ManualRanking and NLPRanking. Figure 13 represents Schema/E-R diagram of the central database on MySQL Workbench.

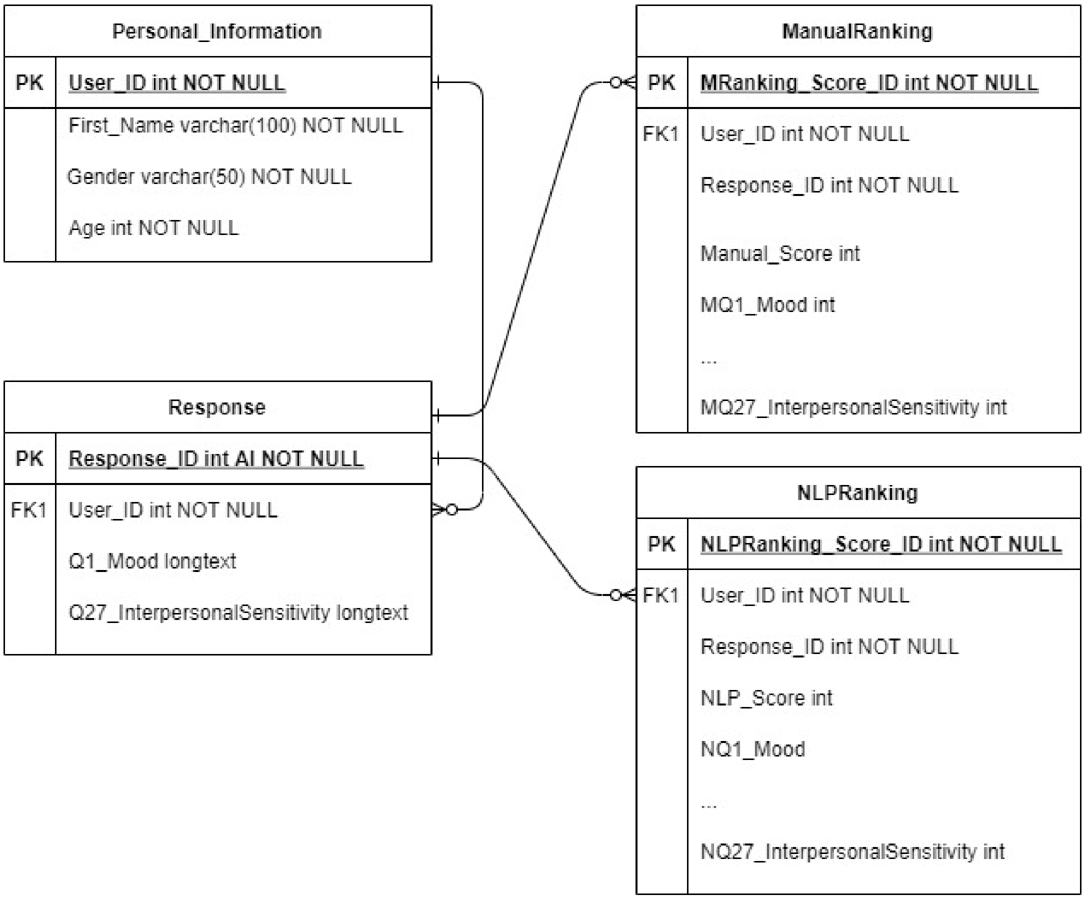


Figure 13 - Schema/E-R diagram of the central database

Personal_Information table includes the basic details of all the participants. Its fields are User_ID (primary key), First_Name, Gender and Age. This table is connected to Response table through a one-to-many relationship. This table is a very essential and critical table as it contains the details of the personal information of the participants. Response table includes Response_ID (primary key), User-ID (foreign key), Q1 to Q27 questions. This table establishes relationships with all other three tables. The foreign keys are the shared entities to spread the relationships. The relation between Response and Personal_Information tables is many-to-one. ManualRanking is designed to include the manual scoring details related to each participant in the database. Its fields are MRanking_Score_ID (primary key), User_ID (foreign key), Manual_Score and MQ1 to MQ27 which are manual score belonging to each question. ManualRanking table is connected to Response table only based on a many-to-one relationship. Finally, NLPRanking table is designed to include the Natural Processing Language scores within its records. Very similar to the structure and design of ManualRanking, its fields are NLPRanking_Score_ID (primary key), User_ID (foreign key), Response_ID, NLP_Score and NQ1 to NQ27 which are NLP score belonging to each question. It also establishes a many-to-one relationship with the Response table only.

Regarding AWS Educate account, we have created an account on Amazon Web Services (AWS) website. The credit for applying this database into our research was $100. By which we had about 5 months of credit to run our database on AWS RDS platform. Figure 14 shows the window related to AWS Services and in particular RDS service.

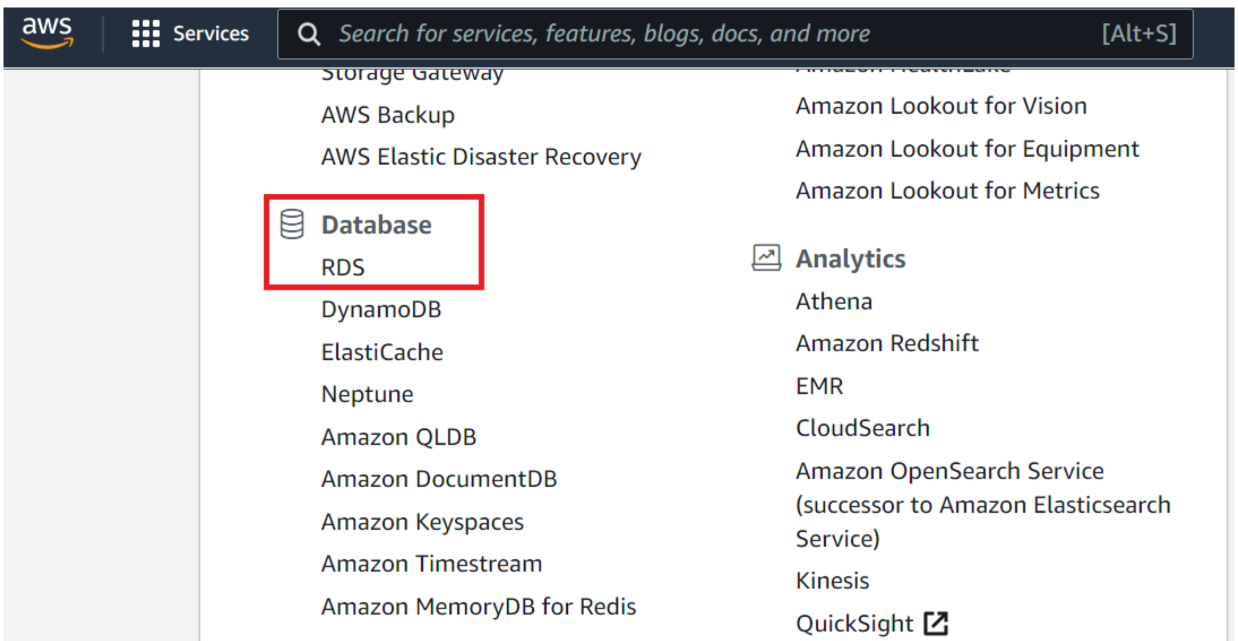


Figure 14 - AWS RDS Service for Database

By the end of the allowed period when the $100 credit was exhausted and AWS Educate changed the policy of supporting our research on the previous policy circumstances, we switched to AWS Academy with access to RDS. We have imported the database on MySQL Workbench, including all tables and their schema, under a new MySQL Connections. Figure 15 demonstrates the navigator display of the schemas and tables designed in MySQL Workbench client.

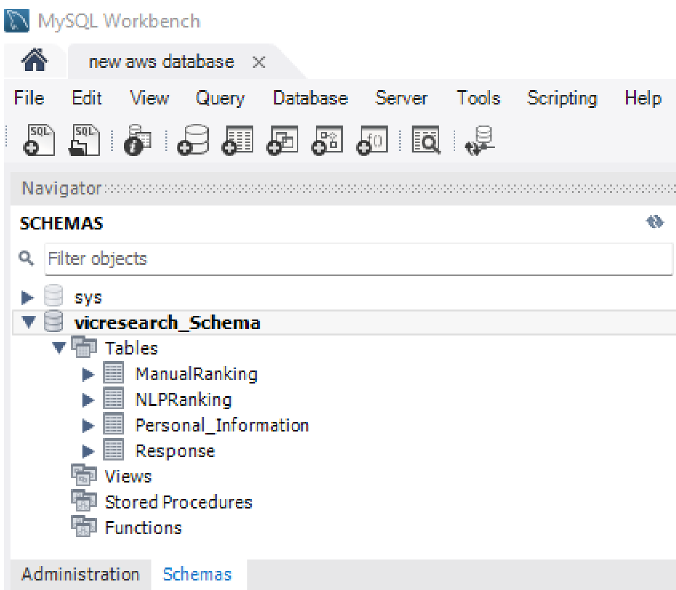


Figure 15 - Display of the Schemas Designed in MySQL Workbench Client

During a transaction, when the participant commences the communication and interaction with the chatbot, they are prompted by a first name. If the participant completed the consent form the sequence continues. If not, they are prompted to complete the consent form before continuation of the interaction. Then, they are checked to be in the right range of participation. Users who are under 18 or above 80 are not accepted for tis research. If in the right age range of participation, the age is entered by the participant. Gender is the next entry asked from the participant. At this stage, the actual data collection commences. The questions 1 to 27 are posed to the participant and they can enter any text messages at any length. The sequence of the questions is significant as they are interrelated to each other. As soon as the participant reaches the last question, a record of all temporary saved data is generated. At this time, the records are saved into all four tables in MySQL Workbench. As mentioned before, the four tables established relationships with each other. The first name, age and gender are saved into Personal_Information table.

Answers to the questions are stored into Response table. We manually score the responses and store them into ManualRanking table. In case we generate the autoscoring, the value related to this is stored into NLPRanking table. The connection which is defined at the design time of the tables come into action and AWS RDS database is also updated.

3.3 Experimental Design

The participants of DEPRA chatbot survey showed a wide range of depression levels. Although the records show a remarkable range of participants experience no symptoms and they are healthy, the level of participants who have mild and severe symptoms are also noticeable. The participants with very severe symptoms are at a high range too. According to the records, participants with moderate symptoms also form a portion of users who are struggling with the depression. The reason behind these statistics can be that most of the participants tried not to open up and reveal their genuine feelings and emotions which were asked by the questions. We have received feedback such as this chatbot sometimes asks about very personal questions that one cannot honestly reply to them. Even after signing the consent form, there were some uncertainties about privacy and how the responses are going to be analysed. By encouraging more participants, we can hopefully achieve more accurate statistical results. Besides, there are remarkable opportunities to attract more participants who can share their true emotions and feelings. As the DEPRA chatbot is published on Facebook Messenger there are a wide range of organic participants who interact with the chatbot based on their requirements and concerns about their mental health. The focus of this research is to prepare trust and mutual understanding especially for the organic participants who are desperate to get help through the social media. They look on the Facebook messenger to find a virtual assistance to establish a connection and receive medical advice related to their critical mental conditions. They would like to be evaluated by the chatbot and receive a score to be classified as healthy, mild depression, moderate depression, severe depression or very severe depression. According to the result of participation they can decide whether they should visit a healthcare professional, a psychiatrist, or they are feeling healthy, and no further action is required. There are some steps that we can take to assure the interaction with organic samples directs to a smooth route. The range of names which are defined for the chatbot are limited at this stage. Most of the English names with some Arabic, Persian and African names are accepted. There are participants who lose their hope when they cannot enter their name at the very first stage of the interaction. Expansion of such name repository can be a rationale move to facilitate the way the mutual conversation takes place. The obligation defined by ethics committee at Victoria University to limit the participants to reside in Australia is also a boundary that needs revision. We receive requests from countries around the world. Some countries such as Syria that are experiencing war in the country can be a suitable candidate to check if the population is suffering from depression at early stages. This defines humanitarian applications to the DEPRA chatbot and its usage.

Due to the popularity of DEPRA chatbot and its user-friendliness that was the result of the design phase, a considerable range of participants voluntarily interacted with the chatbot. Regarding the range of participants, this research includes a variety of genders and age groups. Moreover, the collected data and responses are stored in an AWS RDS database under secure circumstances. Only the administrator of the research group has the authority to access the data or analyse the data. During the collection phases, the participants agreed to reflect their ideas through the questions on DEPRA and the several contacts even took a step further and encouraged their friends and family who were eligible to be a part of the experiment to participate. At the beginning of the research, a closed group was considered. The HDR students and staff at Victoria University were invited to interact with the chatbot. This closed group assisted with the possible training responses that any potential participants might come across and share the same responses that we could expect from the participants. As the participants follow the questions and they answer by reaching the final question they will receive a greeting message as well as a link to the evaluation form where they have the opportunity to fulfil the user experience survey.

3.3.1 Ethics Approval

For this research, the Victoria University ethics approval number HRE20-184 has been approved and allocated. This includes a range of participants who are residing in Australia at the time of data collection and aged between 18 and 80 years old. The participants voluntarily interacted with the DEPRA chatbot to share their emotions and feelings through a non-clinical trial. The requirement for the ethics approval process is that the privacy and confidentiality of participants data to be maintained throughout the research. The policy of this study indicates that only the researcher and the research team have the authority to access the data during the data collection and any stages of the research. The collected data is stored in an R Drive on one of the Victoria University workstations which is secured and has limited access to the research team. Appendix B includes the link to the ethics approval document applied to this research.

3.3.2 User Consent Form

Within DEPRA chatbot interaction, one of the most important forms to be completed by any individual participant is the consent form. When the participant commences the conversation, it is posed to share their given name. Then, a message asks if the participant has completed the form. In case they have a negative response and did not complete the form, a link to the consent form will be displayed on the screen. When the participant completes the consent form, they can return to the conversation and by indicating "confirmed" on the research questionnaire the process continues.

The user consent form commences with a brief guide to the participants by stating the online format of the survey and an email address in case there are any enquiries or doubts to complete the consent form. This section is followed by information to participants and we invited the participants to interact with the chatbot by mentioning the title of this research as a reference and guide. After sharing the significance of the participation of any individuals, the introduction of the survey to pertain 27 questions related to mood, sleep habits, appetite and general health will be mentioned. The only potential risk associated with the research is discussed and the mitigation of plans to avoid such breach of data is mentioned. In this section, the participants declare they are at least 18 years old and voluntarily give their consent to participate in the survey. They also confirm that the relevant details of risks and safeguards associated with this research have been discussed by the research team. The participants also accept that they will be ready to interact with a chatbot online to swap questions and answers, the data collection will be completed by the end of the session, data will be stored in an R drive of Victoria University for analysis and the success or failure of the outcome will be reported to them. It is also crucial to confirm with the potential participants that they can withdraw at any stage of the process and this will not jeopardise them in any way. As per the ethics approval requirement, the participants will be confirmed that their private information is confidential. By the end of the consent form, the contact details of the researcher and ethics secretary at Victoria University human research ethics committee is listed for the participants to share any verdicts, complaints or comments. The next section will ask for personal information such as first name and suburb. Finally, the consent form electronic signature will be collected from the participants by asking if they agree with the terms and conditions or not.

Appendix C represents the consent form applied to the participants of this study.

3.3.3 User Rating Form

To access and evaluate the quality and preciseness of the survey, after the successful completion of 27 questions designed in the conversation, a message is shared with the participants. There is the appreciation of interacting with the chatbot and a link to a user rating form.

In the user rating form, we ask the participants to share how they find the survey in general. There are six questions in the survey. The first five questions are scaled from 1, strongly disagree, to 5, strongly agree. The first question related to the ease of understanding and responding to the questions of the chatbot. 41.4% (12 out of 29) agreed that DEPRA's questions were easy to understand and respond to. 34.5% (10 out of 29) of the participants strongly agreed that DEPRA's questions were easy to understand. Only 24.1% (7 out of 29) of the participants felt neutral, neither agreeing or disagreeing that DEPRA's questions were easy to understand. No participants indicated that they disagreed with the statement. The second question related to the time spent on this survey in comparison with a real psychiatrist session. The same number of participants 34.5% (10 out of 29) either agreed with or felt neutral to the question about the time required to participate in the DEPRA conversation being less than the time required to see a psychiatrist in person. 27.6% (8 out of 29) strongly agreed that the time was managed better and only 3.4% (1 out of 20) strongly disagreed. The third question asks the participants to indicate their preference for the use of text messaging rather than talking to a psychiatrist. The same number of participants 31% (9 out of 29) either strongly agreed or felt neutral regarding their preference for text messaging rather than talking to a medical professional. The remainder of the participants 27.6% (8 out of 29) agreed that they preferred to send text messages rather than engage in a conversation with a psychiatrist. No participants strongly disagreed with the idea of text messaging compared to attending a real psychiatrist session, however, 10.3% (3 out of 29) disagreed. The fourth question asks the participants whether they felt that the sequence of questions directed the participant to reveal their level of depression. Most of the participants 37.9% (11 out of 29) were neutral, 31% (9 out of 29) agreed that sequence of the questions led the participant to reveal their level of depression and 24.1% (7 out of 29) strongly agreed. Only 6.9% (2 out of 29) disagreed with the idea that the sequence of questions led the participant to reveal their level of depression and no participants strongly disagreed. The last question asks the participants whether they are likely to recommend this survey to their friends or families. 34.5% (10 out of 29) strongly agreed that they would invite their friends and families to participate in the survey and 31% (9 out of 29) agreed. 27.6% (8 out of 29) of the participants were neutral and the same number of participants 3.4% (1 out of 29) disagreed and strongly disagreed. The responses to the rating form reflect the high satisfaction rate of the users.

Finally, the last question is an open-end question seeking the comments and suggestions of the participants. Appendix D shares a link to the user rating form.

3.3.4 User Experience

In this section, we are going to describe the user experience and the journey that they take when they interact with DEPRA chatbot integrated with Facebook Messenger. The chatbot link was shared with the potential participants. The method was to send the link in an email or ask the participants to look up the DEPRA Facebook page through the Internet. Then, when the participants click on the link, they will be directed to the DEPRA Facebook page. On the page, the right-hand side contains an edit send message button. By clicking the link, they can commence the interaction with the chatbot. In the opened window, they can enter a command such as “hi”, “get started” or simply “hello”. After a greeting message to welcome the participants into the experiment, the chatbot will seek the given name of the user. This is a very critical point of the conversation as the chatbot will ask about the consent form and whether it is completed. If the participant already completed the form, with a simple confirmation as yes they will be directed to the next stage. In case they did not complete it, a message will be generated with the relevant link to the consent form to be revealed and shared with the participant. They also will be consulted to type in “confirmed” as soon as they complete and return to the conversation. After the consent form completion, the next question from the participants would be their age. If they are under 18 or over 80 years old they would be advised that they are not eligible to participate. In any other cases, they will continue the conversation. The next step would be to ask for the gender of the participants. They can be male, female or neutral, that is, prefer not to share. So, the next step will be jumping to the questions. It starts with question 1 and goes all the way through to question 27. As soon as the participants respond to the last question, a message would be generated that appreciates the participants' interaction and leaves a link to the rating form to complete and ask about the survey quality.

3.4 DEPRA Chatbot Implementation

As the first step in the implementation phase of building the virtual agent, the concept of intent in the Dialogflow platform was addressed. For each question, we required an intent to be defined. Dialogflow contexts are similar to natural language contexts. In our case, input and output contexts were used to manage the flow of conversation and to pass the value of parameters in the code. We use inline editor to develop and deploy the code and the service within Dialogflow platform. Furthermore, one of the significant features of the DEPRA chatbot is the open-ended way of responding to the questions. With the input and output contexts, we can receive or pass the parameters to the other intents respectively. This can be crucial in the design phase as we should be able to have control over the parameters. Following the implementation and deployment, we integrated Facebook Messenger with Dialogflow. To complete the integration the Facebook Messenger platform used to create a messaging app and to configure the integration between the Facebook app and the Dialogflow to send messages to the end-user through the Facebook Messenger API. Facebook Messenger was chosen as the proxy due to its popularity around the world and including the most population of active users compared to other platforms such as WhatsApp, We Chat and so forth. When a participant interacts with DEPRA by sending a message on Facebook, a mutual conversation takes place. As the participant reaches the last question, that is question number 27, the data stored throughout the conversation transfers into the AWS RDS database. Then, the participants have the opportunity to rank the DEPRA and its interaction method to collect the data. The guideline applied guarantees that the symptoms of depression have been addressed as the participant makes progress on the chatbot conversation.

3.5 DEPRA Features

DEPRA chatbot and its features have been summarized in Table 3.

Table 3 - Features of Early Depression Detection DEPRA chatbot

Early Depression Detection Chatbot	Open-Ended Response	Real-time Database	Experiment Duration	Scoring System	Guideline Used
DEPRA	Yes	AWS and My SQL Workbench	1 session (30 minutes)	IDS-SR and Q-IDS-SR	Hamilton Depression Scale (SIGH-D) and Inventory of Depressive Symptomatology (IDS-C)

To our knowledge, DEPRA is the first early depression detection chatbot which is designed to assist the medical science to create virtual assistance to the health care professionals and psychiatrists. It can be used to collect the data from the patients when due to lack of professionals or in crises such as Covid-19 access to medical support is limited. DEPRA can guide the patients to take a wise step to deal with their requirements and receive medical advice when the medical professionals are not accessible. Communication with DEPRA as an open-ended assessor, the 24/7 availability, and the approach to detect depression at early stages prepare a basis to trust the chatbot and benefit from its application in critical times. Moreover, it takes approximately half an hour to commence and finalize the conversation. With this approach, the hassle of follow-ups and sharing the wellbeing of the participants at several sessions and within different time frames can be limited. With only 27 questions posed to participants by DEPRA, we noticed that the participants lose interest when they reach the last 10 questions. They prefer to leave a short response, sometimes even the responses that do not make sense, and they rush to finish the conversation. It can be concluded that they will not be ready for a set of conversations based on various time frames. So, it can be driven that the shorter the conversation sessions the better results and data would be collected. Also, the DEPRA chatbot is based on the Hamilton interview guideline and it guarantees the set of questions are medically approved by professionals. The scoring system is also in place so the participants are informed about their health and the level of depression if they are suffering from it. IDS-SR and QIDS-SR are the utilized scoring systems used throughout this study. Manual scoring is in place and the next step in this study is to share the result of automatic scoring with the participants. When they reach the final question, the score will be shared with them within the final greeting message.

CHAPTER 4 NON-CLINICAL TRIAL USING DEPRA

4.1 Introduction

Chapter 4 addresses the different stages of non-clinical trial using DEPRA including reviewing the participants demography, the user experiences analysis and manual scoring to determine the participants' depression severity.

In this section, we will discuss the analysis and results that are generated as a consequence of conducting the research. We have considered participants' profile and the analytical values concerning age, gender and ethnicities. We also considered participants' depressive variables with regard to two scoring methods, IDS-SR and QIDS-SR, and their effect on calculating the overall scores. Moreover, we discuss the score analysis and a comprehensive list of participants corresponding to the scoring methods applied to the list of participants. Finally, we will discuss and analyse user satisfaction results and responses collected from participants.

4.2 Data Collection

Table 4 shows the demography of the entire sample (N=50) and the breakdown of individual groups. The average age of the participants is 37.4 years, and more than half are male 54% (27/50). The majority of the participants are Middle Eastern 62% (31/50) followed by Asian 28% (14/50). Caucasians and Indians, etc., represnt 10% (5/50) of theparticipants.

In the first group (VU Staff and HDR Students), 50% (8/16) of participants are females, 32% (5/16) are males, and 18% (3/16) prefer not to disclose their gender. The average age of the first group is 22.2 years. The majority of this group is Asian 63% (10/16). The remainder of the first group of participants are Middle Eastern 25% (4/16) and others 12% (2/16).

The second group (Facebook Friends) consists of 42% (12/28) females, 50% (14/28) males, and 8% (2/28) who prefer not to disclose their gender. The average age of this group is 38.6 years. The majority of this group is Middle Eastern 71% (20/28) with 22% (6/28) Asian, and others 7% (2/28).

The third group has the highest rate of dropout, which is to be expected. People found the chatbot on the page on their own, and from their curiosity, they decided to interact. After reading the consent form, many shows their unwillingness to continue. The organic group consists of 66% (4/6) females, 17% (1/6) males, and 17% (1/6) who prefer not to disclose their gender. The average age of this group is 37.2 years. The majority is Asian 66% (4/6). The remainder of this group of participants is Middle Eastern 17% (1/6) and others 17% (1/6).

Table 4 - Demographic Variables of Participants

Demographic Variables	Group 1 (VU Staff and HDR Students)	Group 2 (Facebook Friends)	Group 3 (Organic)	Entire Sample (N=50)
Gender, n (%)				
Female	8 (50%)	12 (42%)	4 (66%)	22 (44%)
Male	5 (32%)	14 (50%)	1 (17%)	27 (54%)
Not Disclosed	3 (18%)	2 (8%)	1 (17%)	1 (2%)
Ethnicity, n (%)				
Asian	10 (63%)	6 (22%)	4 (66%)	14 (28%)
Middle Eastern	4 (25%)	20 (71%)	1 (17%)	31 (62%)
Other	2 (12%)	2 (7%)	1 (17%)	5 (10%)
Age Group, n (%)				
20-29	3 (19%)	5 (18%)	1 (17%)	18 (36%)
30-39	3 (19%)	6 (22%)	3 (49%)	12 (24%)
40-49	8 (50%)	16 (57%)	1 (17%)	15 (30%)
50+	2 (12%)	1 (3%)	1 (17%)	5 (10%)

4.2.1 Participants

The participants of DEPRA chatbot survey showed a wide range of depression levels. Rarely the participants experienced severe depression and most of the range was in the moderate range. The reason behind these statistics can be that even after signing the consent form, there were some uncertainties about privacy and how the responses are going to be analysed. A solution to receive more genuine responses and participations can be to invite more contacts to interact with the DEPRA chatbot. A policy is to improve the quality of questions by reviewing them and modifying or adding more questions. This can be based on feedback that we have received from the participants who completed the ranking form

One of the key factors that we would like to focus on to attract more participants is the families and friends of the participants who already participated. They can discuss the questions with their referrals, participants who already experienced the process. With a clearer image in mind, they can contribute to the data collection at its next stage. They can be assured that the privacy and confidentiality of their information is secured. By having this ease of mind, they can respond to the questions and be confident that the responses are not going to be used against them at any stages. The only use of this interaction is for research purposes and they also have a chance to check their depression levels to find out if they are healthy or require to visit a health care specialist for further discussion.

Furthermore, the organic participants are a crucial potential to attend the data collection phase. They are showing a remarkable interest to participate and be evaluated regarding their health status. It is usually the matter of a follow up to ask them to return to the interaction and share their responses with the DEPRA chatbot. The range of names that can be accepted by the chatbot is limited at the moment. We can establish a more comprehensive name acceptance by the bot to remove the constraints of being rejected to continue the process with the chatbot. For instance, African names are not fully supported by the current version of the bot. So, when the participant enters their name, in some cases, the system cannot recognize it as a real name and asks the user to enter a valid name. This is a drawback when they would like to commence the conversation. Also, most of the organic participants refuse to submit the consent form. They might feel uncomfortable to sign the electronic version of the form which is a significant form and a prerequisite of the interaction. We should find a way to encourage the participants to trust the research and sign in the form as the first priority.

4.2.2 Participants' Recruitment

Our participants' recruitment policy strictly follows the guideline defined in our ethics approval. As per the ethics guideline, all participants lived in Australia during data collection, aged between 18 and 80 years old and were interested in sharing their moods and emotions in a scientific trial. This research analyses the first 50 participants which are collected until 30th October 2021 (cut-off date for this experiment). We used three different channels for recruitment - the first group was university students and university professional staff either full-time, part-time or sessional. We emailed a broad group of potential participants with the webhook link to contact and invite the users to allocate time and participate in this research. As per the plan, we required the academic group to assist in achieving two tasks. First, we recruited 10 participants (N=10) to share their ideas and responses in a closed group. They were crucial to design the utterances on Dialogflow as the platform for the DEPRA chatbot. Secondly, we asked the same group to participate in data collection to assist with the research furthermore. To invite more participants, we emailed 75 university staff and HDR students. We managed to receive 24 complete records in our AWS central database.

The second plan was to collect data from friends or relatives who were Australian residents at the time of the data collection. Facebook friends were recruited for this research. In this phase, we targeted various backgrounds such as professions, genders and age groups. Since we published the DEPRA chatbot on Facebook Messenger, the qualified participants initially signed the consent form and then participated in the data collection process. We received 16 complete records from the second group. The 40 participants recruited following the first two processes are inorganic as we targeted these participants. The third group is exciting and organic and is still growing at the time of drafting this thesis. A Facebook page was published, and more organic participants interacted with this Facebook page to address the requirements of the experiment. The organic participants had to respond to the set of questions and complete the consent form and rating form. The participants were from various age groups, located in Australia and were between the age range that we defined for this research. The organic population who assisted in collecting data were Facebook users keen on opening up and sharing their experiences with the DEPRA chatbot through our Facebook page. This group found the page on their own and felt the necessity to interact. We registered ten records from this group, tallying the total to 50, excluding duplication and incomplete conversation until the cut-off date. For all three groups, the exclusion criteria include non-resident of Australia, outside of age group, unsigned consent form, and incomplete conversation. Figure 16 summarizes the DEPRA participant recruitment flow in a hierarchical chart.

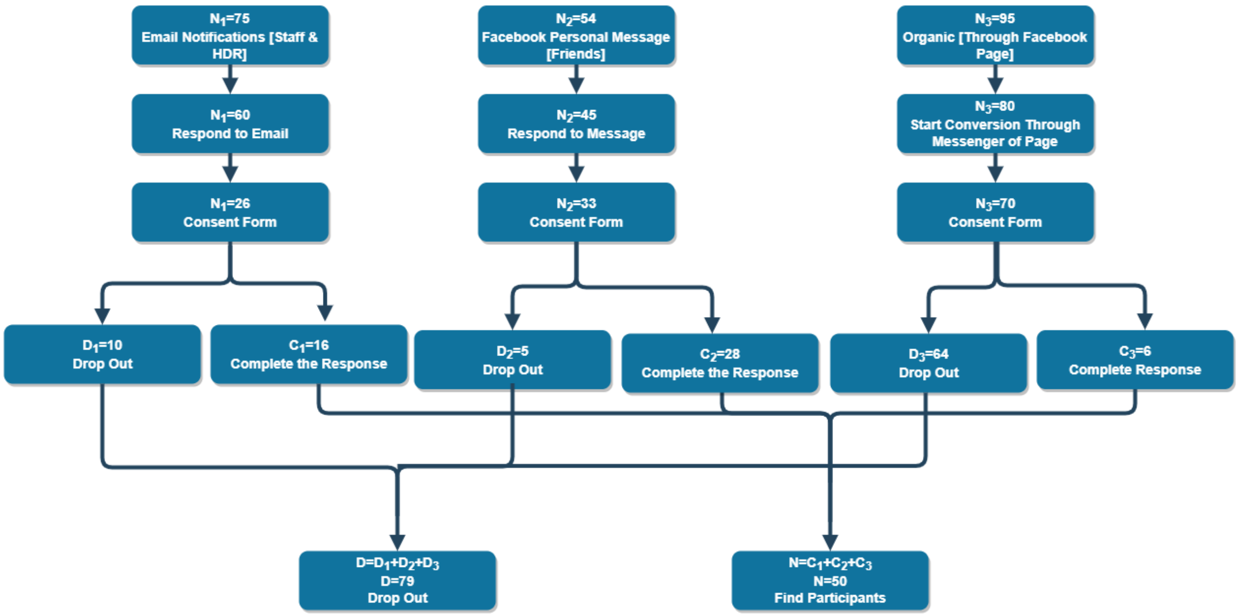


Figure 16 - DEPRA Participant Recruitment Flow

4.2.3 Participants' Depressive Variables

In this study, we have applied two different methods for manual scoring. Namely, IDS-SR and QIDS-SR [45]. In IDS-SR method all the 17 categories are considered including mood, reactivity of mood, outlook (future), guilt, insomnia, suicidal ideation, anxiety, involvement, concentration, psychomotor, somatic, symptoms, general, pleasure, weight, interpersonal sensitivity, and genital symptoms. However, in QIDS-SR the categories are divided into 9 groups of emotions. They are cited as sleep (sleep onset insomnia, mild-nocturnal insomnia, early morning insomnia and hypersomnia), mood, weight (appetite decreased, appetite increased, weight decrease, weight increase), concentration, guilt, suicidal ideation, interest, fatigue, psychomotor changes (psychomotor slowing, psychomotor agitation). The equations, eq 1 and eq 2, which are applied to calculate the IDS-SR and QIDS-SR methods are as the following:

$$IDS - SR\ Total\ Score = \sum_{i=1}^{27} MQ_i \qquad (1)$$

Where MQi is the manual score of the ith question.

$$\begin{aligned}\boldsymbol{QIDS-SR\ Total\ Score} &= \mathrm{MQ22}_{\mathrm{SomaticEnergy}} + \mathrm{MQ1}_{\mathrm{Mood}} + \mathrm{MQ5}_{\mathrm{Guilt}} + \mathrm{MQ6}_{\mathrm{Suicidal}} + \mathrm{MQ11\_Interest} + \mathrm{MQ13\_Concentration} \\ &+ \mathrm{MAX}(\mathrm{MQ7\_InitialInsomnia}, \mathrm{MQ8\_MildInsomnia}, \mathrm{MQ9\_MorningInsomnia}, \mathrm{MQ10\_Hypersomnia}) \\ &+ \mathrm{MAX}(\mathrm{MQ21\_Appetite}, \mathrm{MQ26\_Weight}) + \mathrm{MAX}(\mathrm{MQ14\_PsychomotorSlowing}, \mathrm{MQ15\_Agitation})\end{aligned} \tag{2}$$

To summarize, for the IDS-SR total score we have added up all the manual scores of each psychiatric question so as you can see in the equation, all the 27 manual values added up and the summation displays the overall value for this IDS-SR method.

Regarding the QIDS-SR, there are four categories. Frist of all, the maximum value under three groups will be calculated. Maximum value of insomnia would be between initial insomnia, mild insomnia, morning insomnia and hypersomnia. The next step is the maximum value between appetite and weight. Last group of symptoms includes psychomotor slowing and agitation. By having the maximum values, we just add up manual scoring of the semantic energy, guilt, suicidal, interest and concentration. Summation of these groups will produce the QIDS-SR total score.

The categories related to IDS-SR and QIDS-SR and the corresponding questions have been summarized into Table 5.

Table 5 - Symptoms Categories and Relevant Questions of Scoring Systems

IDS-SR Symptoms	
Category	**Questions**
Mood	1,3,4
Outlook	2
Guilt	5
Suicidal	6
Insomnia	7,8,9,10
Interest	11,12
Concentration	13
Psychomotor Slowing	14
Agitation/Anxiety	15,16,17
Panic	18
Arousal	19
Gastro	20
Appetite	21

QIDS-SR Symptoms	
Category	**Questions**
Mood	1
Guilt	5
Suicidal	6
Interest	11
Concentration	13
Somatic Energy	22
Insomnia	7,8,9,10
Appetite/Weight	21,26
Slowing/Agitation	14,15

Somatic	22,23,24
Genital	25
Weight	26
Interpersonal Sensitivity	27

4.3 Analysis and Interpretation

In this section, we discuss how the collected data is managed and prepared to be regarded as clean data. Data pre-processing, DEPRA chatbot training and categorization of the participants according to their gender, age and profession is in place. Finally, the sequence of responses is summarized into a table.

4.3.1 Data Pre-processing

After the data collection phase, we have pre-processed the dataset to remove the irrelevant records entered by the participants. Data cleaning was completed based on standard procedures. Although, we had integrated data validation into the DEPRA design and implementation to minimize the inconsistency in data. Like any dataset, our dataset also contained invalid, missing and outliner data. First we screened our dataset for errors or inconsistency, removed the invalid entries and developed codes to detect outliners and map to valid data. They include the special characters such as exclamation marks, question marks and so forth. Also, the incorrect responses which are irrelevant to the questions are removed or modified to make them more meaningful. In some cases, the participants lost their interest when they reached the questions 20 out of 27. For instance, they pressed the alphabetic characters on their keyboards, and they created irrelevant responses. According to the way they have responded to the set of questions, we decided to remove their records or keep them when they have complied with most of the questions by comprehensive and meaningful responses. The reason behind such approach is to prepare records that can be processed in the manual and automatic scoring phases. In NLP this process is called cleaning and tokenizing. After this stage, the sentences that are clean and without any hurdles in the way will be used to calculate the score for the responses from the participants.

4.3.2 DEPRA Chatbot Training

Training DEPRA based on DialogFlow with adequate utterances for an acceptable and accurate association between intent and response is one of the essential design requirements. On this note, we conduct a closed group survey, where participants (10) are university staff and students, who are considered as the first level of connection to the research. The research on its initial formation was focusing on development and implementation of DEPRA chatbot. The closed group survey regarded as the assistants to share and to generate the potential responses that could train our chatbot. The reason behind utilising the group from university staff and HDR students is that they were aware of the process of the research, and they could contribute for training the chatbot. Given the COVID-19 constraint, any other potential participants could put the research timeline at risk to achieve our study critical path. They are aware of the outcomes and methodology of this project. Figure 17 shows the statistics of closed group with regard to age, gender and profession.

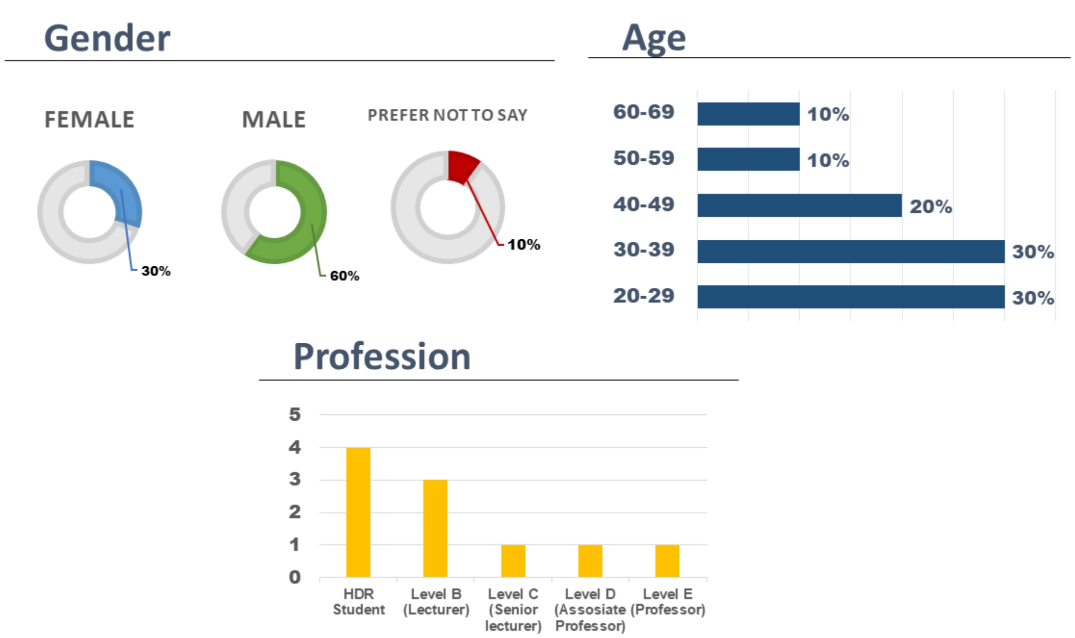


Figure 17 – Closed Group Statistics based on Gender, Age and Profession

About 60% of the closed group participants were males, followed by 30% females and only 10% of the participants preferred not to share their genders. Age group of 20 to 29 and 30 to 39 formed 30% of the participation. Participants between 40 and 49 were 20% while the elderlies have the lowest level of participation with only 10%. The result confirms that the young generation takes the experience more seriously compared to the aged groups of 50 years old or over.

We use Google Forms as a simplified platform to conduct this closed group survey. Appendix E shows the atomic data of all questions from ten (10) participants. The focus group survey aims to extract a requisite number of keywords to generate various lengths of utterances.

By the end of closed group participation, the proposed utterances which were suggested by the participants in this group were analysed. The responses to some common questions were almost identical, however, the closed group participants shared very interesting and challenging responses that could be applied to the DEPRA chatbot design. We basically applied these range of responses in the design and implementation of the chatbot.

Table 6 shows the semantic of the responses w.r.t extracted keywords, sentiment score and length (min/max) of replies for each question. We use the excerpted keywords to generate one hundred (100) utterances for answers to each question (except those with the numerical number), sentiment score to understand the association between the response and respondent's state, and length to bring variation within the synthetic utterances.

Table 6 - Sequence of Responses

Seq	Data Type	Keyword	Sentiment Score	Response Length Min/Max (char)
Q1	Str	Not bad, good, Normal, Pumped Up, Under pressure, Not confident, Disappointed, Down, Frustration, Challenging	Negative 94.6%	4/166
Q2	Str	Fluctuating, sometime good, sometime bad, bit down, excited, opportunities, nervous, a bit anxious, so nervous, all over the palce,	Positive 58%	15/125

		uncomfortable, down,struggling, out of shape, on the edge, lack of sleep, a lot of pressure,fed up, no chance to relax		
Q3	Str	Hopeful, Good, Positive, bit frustrated, Confident, Concentrate on present activities, negative feelings, unwanted challenges, no hopes, difficult times, Crisis, more wars, lose home countries, problems, mystery	Positive 75.3%	4/201
Q4	Str	No, Not really, put myself down, Judged outcomes, challenged my wife, Yes, Let a colleague down, Dwon, out of control, Satisfied with performance, dedicate to team	Negative 69.7%	2/232
Q5	Str	Lost few relatives, Sad, Loss, Fine, Yes, Positive, Passed away, not feel the same, Unfortunately, so many losses, the same feeling, crying, Grieving, Pressure, Lose business, crisis, feeling down, Getting worse, Passed away, tragic remembrance, Lost job, Irritated, Hopeless, Lay off, The same feeling, Overwhelmed	Positive 70.5%	3/259
Q6	Str	Not realy, No, Under control, Felt down, Down, Yes, Worse, Not bothered with depression, Can not breathe, Negative thoughts, Could not monitor, Depressed	Negative 95.6%	2/176
Q7	Str	No, Too busy, Self harm, End my life, Yes, Meaningless, Dead, Tough time, Newver think about self harm, Value life, Enjoy, Vary harsh shock	Positive 83.8%	2/168
Q8	Str	No, Yes, Loud music, Mind so busy, Not feel sleepy, No difficulties to sleep, Stop crying, Normal day, Buzzer went off, Technical issue, No interruptions	Positive 91.90%	2/142
Q9	Str	Yes, No, Woke up to drink water, Tough, Manege to sleep, Not go back to sleep, Scared, Nightmare, Fix issues	Negative 98.80%	2/130
Q10	Str	Doing nothing, Reading novel, Watching TV, Family, Staying with daughter, Assisting kids, Gardening, Discussion, Doing laundry, Think about a business, Studying, House chores, Sleeping, Listening to music, Preparing meals, Taking care of nephew, Talking to family overseas	Negative 91.90%	6/118
Q11	Str/Int	N/A	Neutral 77.20%	N/A
Q12	Str/Int	N/A	Neutral 76.20%	N/A
Q13	Str	No, Yes, Playground, BBQ party, enjoy, Essential tasks, No fun, Not really, No time for fun, Zoom conference call	Negative 58%	2/130
Q14	Str	Poor, Not bad, Normal, Focused, Tend to lose concentration, In place at work place, Terrible, Acceptable level, Could not concentrate, Focus on studies, Act as on-call staff, Pretty fine	Negative 91.70%	4/100
Q15	Str	Yes, No, Little bit, Slowed down in the morning, Get better, Deal with skills, Slow down in thinking, No problems, Up and running	Positive 78.30%	2/144
Q16	Str	Not exactly, Little bit, No, Yes, on edge, Anxious, nervous, Perform well, Crisis, Overwhelmed	Negative 85.60%	2/128
Q17	Str	No, Yes, Tense, Deal with tasks, Irritable, New task, People with corona	Negative 69.30%	2/101
Q18	Str	No, Yes, Restless, Pressure of life, Fidgety, Solve problems, Normal life, Infected patients	Negative 93.10%	2/83
Q19	Str	No, Not at all, Yes, Extremely uncomfortable, frightened, Uncomfortable, Imagined, Argue about instructions, Number of cases jumped	Negative 95.90%	2/155
Q20	Str	No, Nothing, Palpitations, Blurred vision, Chest pain, Increased sweating, Hot and cold flashes, Dyspnea, No such symptoms	Negative 99.90%	2/60

Q21	Str	No, Constipation, no symptoms, Diarrhea,No such symptoms, None	Negative 100%	2/120
Q22	Str	Poor, Tired, Normal, Really low, Block, Low, Lowest possible level, Average, Pretty average	Negative 98.10%	4/67
Q23	Str	Same, Normal, Good, No, Increased, Unchanged	Positive 66.70%	2/53
Q24	Str	Same, Normal, No, Decreased, Unchanged	Negative 85.40%	2/71
Q25	Str	No, Bone pain, Headaches, Body pain, Swollen fingers, Blurred eyes, Body is killing, Stomatch ache, Sore arm, Sore neck	Positive 70.60%	2/82
Q26	Str	No, No such feelings, Carrying my body, Yes, Feeling to be weighted down, Just ok	Neutral 52.80%	2/79
Q27	Str	As usual, Good, Normal, The same level, Pretty off	Negative 75.90%	5/79

4.3.3 *Depression Score Results*

According to Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS) [46] the medical range for IDS-SR and QIDS-SR methods are as per Table 7.

Table 7 - Comparison of Total Scores with regard to Depression Levels

IDS-SR	QIDS-SR	Depression Level
Range between 0 to 13	Range between 0 to 5	No Depression, Healthy
Range between 14 to 25	Range between 6 to 10	Mild
Range between 26 to 38	Range between 11 to 15	Moderate
Range between 39 to 48	Range between 16 to 20	Severe
Range between 49 to 84	Range between 21 to 27	Very Severe

Regarding the IDS-SR scoring system, it is observed that around 30% (15 out of 50) of the participants were experiencing no depression and they were healthy. As a result, no further follow up session with the medical professions or a psychiatrist were recommended. The patients with very severe symptoms and moderate symptoms were almost at the same level with 20% (10 out of 50) and 22% (11 out of 50) respectively. Finally, the participants with severe symptoms and mild symptoms experienced the same level of depression with 14% (7 out of 50). Figure 18 displays the findings and how the results are distributed.

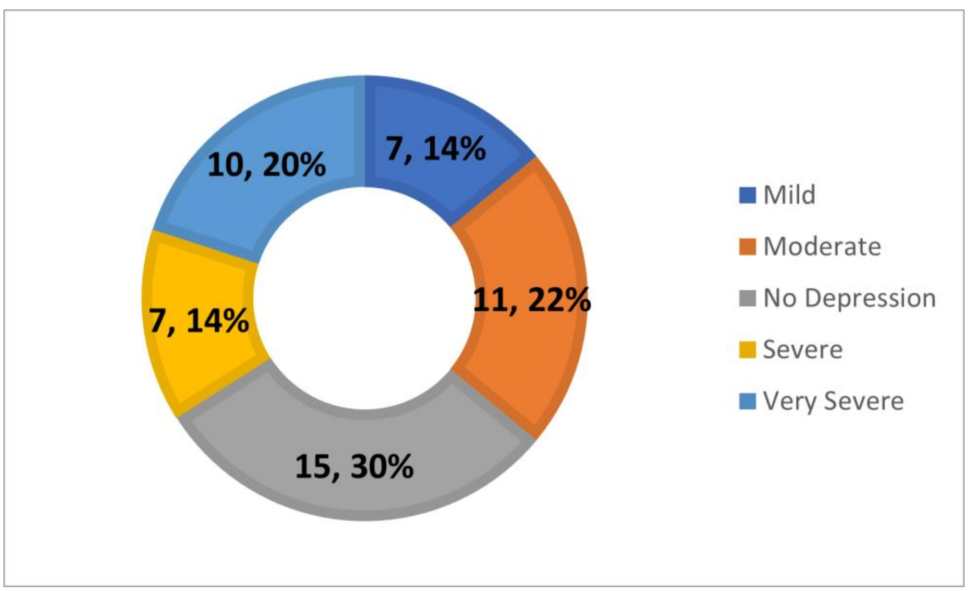


Figure 18 - IDS-SR Depression Level Statistics

Regarding the QIDS-SR scoring system, the graph shows that around 32% (16 out of 50) of the participants were completely healthy. So, no further action was required to be taken by the participants in this group. The patients with very severe symptoms formed 22% (11 out of 50). The participants in this group required immediate medical advice and they were suggested to contact and visit a medical profession as soon as possible to follow up their health condition. The next group includes participants with severe symptoms and mild symptoms. They formed the results at the same level, they are both reporting 18% (9 out of 50). For this group, the patients with severe symptoms are advised to receive medical consultation when it is possible for them to avoid the condition to escalate and have it under control. The participants with mild symptoms are not required to visit a health care professional and they can read more about their symptoms and be aware in case the symptoms might increase. They are basically safe and sound at this stage of the research. Lastly, the moderate symptoms of the participants reported as 10% (5 out of 50). Figure 19 displays the findings and how the results are distributed to reflect the results.

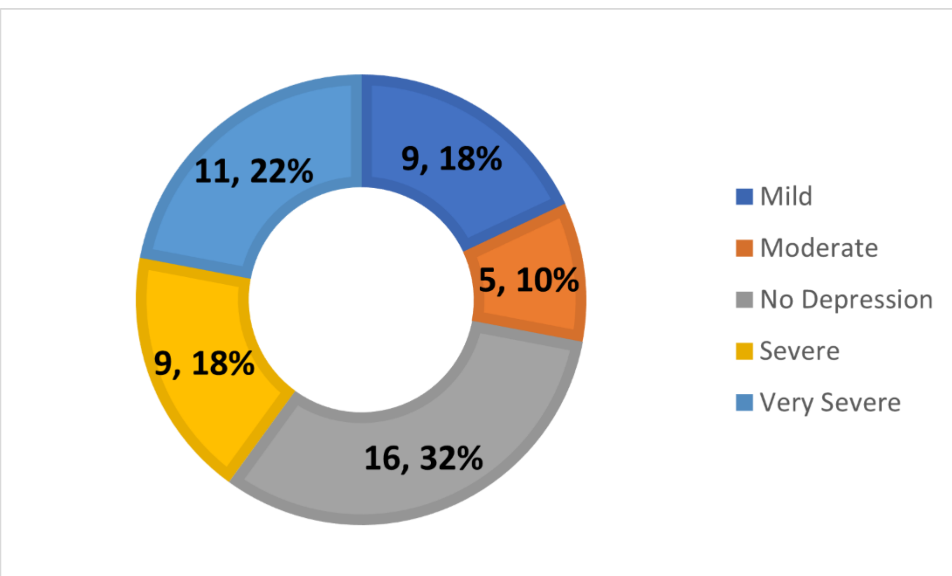


Figure 19 - QIDS-SR Depression Level Statistics

For the range of 50 participants in the dataset that we have collected the data so far, here is a comparison between the ID-SR against QIDS-SR in Figure 20.

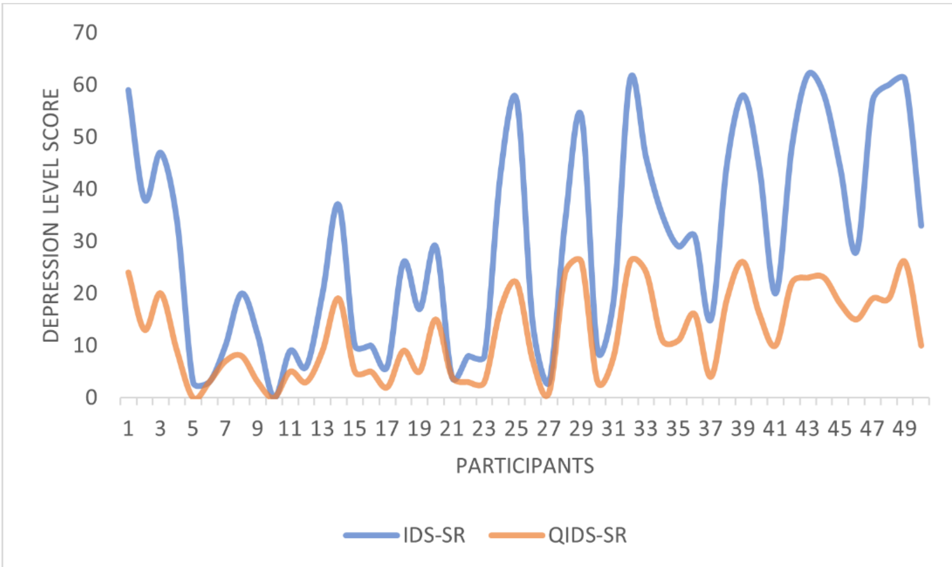


Figure 20 - IDS-SR vs QIDS-SR

As the graph reflects the overall trend of these two scoring systems are identical. However, there are minor differences to be noticed. For example, the number of participants with very severe depression for IDS-SR system was 10 out of 50 while QIDS-SR experienced 11 out of 50 participants suffering from very severe depression. By applying both scoring systems and comparing them against each other, the overall analysis reflects a more accurate and self-explanatory outcome. Appendix F includes a link to a comprehensive list of participants and their responses as well as the IDS-SR and QIDS-SR overall score related to the first 50 participants in DEPRA chatbot interaction.

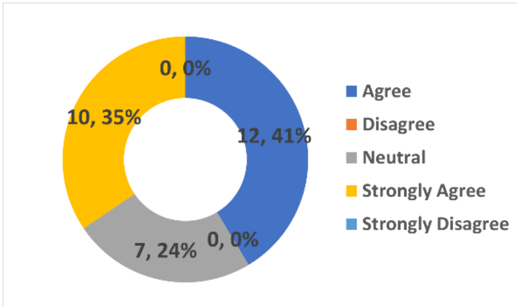
4.3.4 User Satisfaction and Engagement

In this section, we intent to describe the set of questions which were posed to the participants at the end of the data collection phase. A link was proposed to the participants as soon as they finalized answering to the DEPRA chatbot session. There are five linear scale questions that encourage the participant to choose from a range of 1 (very disagree) to 5 (extremely agree) scale and one open-ended question in this rating assessment. Overall, we received and recorded 29 participants rating form submissions from the survey.

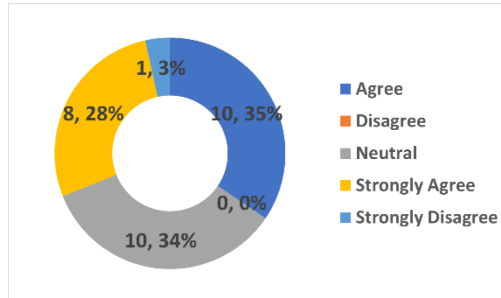
The first question focused on the rating of the easiness to comprehend and response of the chatbot. 41.4% (12 out of 29) were agree that the DEPRA questions were easy to understand and respond. 34.5% (10 out of 29) of participants felt they were extremely agreeing with the idea of easiness of the questions. Only 24.1% (7 out of 29) were feeling neutral, they could not agree or disagree with the easiness of the DEPRA questions. There were no participants with disagree idea of easiness. Figure 21a reflects the participants' verdict over the question 1.

The second question posed the time frame spent on this survey in comparison with a real psychiatrist session. The ratio for the participants feeling neutral or agree to the idea of a better time frame of the survey participation and conversation with DEPRA compared with a real psychiatrist session was at the same level of 34.5% (10 out of 29). 27.6% (8 out of 29) were thinking extremely agree that the time frame was managed better. There was the low level of very disagree of the time frame with only 3.4% (1 out of 29). Figure 21b reflects the participants' verdict over the question 2.

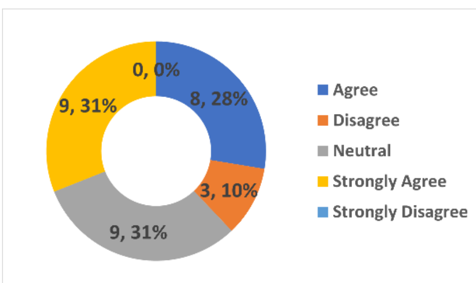
The third question targets the preference by the participants to use text messaging rather than talking to a psychiatrist. Participants who are extremely agreeing or neutral portion of the participants who preferred text messaging rather than talking to a medical professional formed the same level of 31% (9 out of 29). The remainder of the participants who were agreeing with text messaging rather than an actual conversational session were at the 27.6% (8 out of 29), slightly lower than the previous group of participants. There were no participants who felt very disagree with text messaging compared to a real psychiatrist session, however, around 10.3% (3 out of 29) were disagree. Figure 21c reflects the participants' verdict over the question 3.



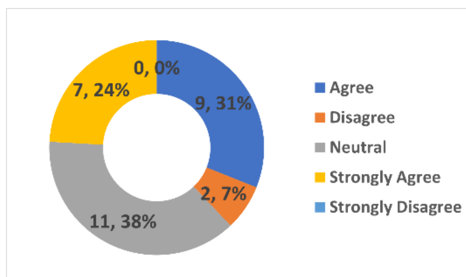
(a). Participants' Verdict over Question 1



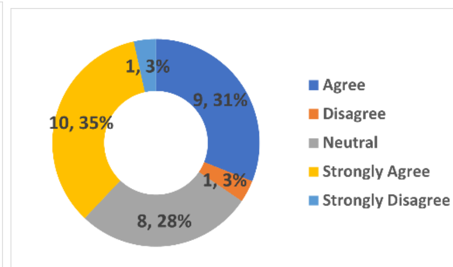
(b). Participants' Verdict over Question 2



(c). Participants' Verdict over Question 3



(d). Participants' Verdict over Question 4



(e). Participants' Verdict over Question 5

Figure 21 - Depression Level Statistics According to Participants Interactions

The fourth question reflects the fact that if the sequence of questions directs the participant into revealing of level of depression. Most of the participants were neutral with the idea of sequence of questions that leads to revealing the depression level with 37.9% (11 out of 29). 31% (9 out of 29) were agree with the idea of the sequence of questions will result in depression level assessment while 24.1% (7 out of 29) were extremely agreeing with it. Only 6.9% (2 out of 29) were disagree with the idea that the sequence of question lead into the depression level assessment and there was no participant who was very disagree with the idea. Figure 21d reflects the participants' verdict over the question 4.

The last question concentrate on the possibility of recommendation of this survey to the participants' friends or families. 34.5% (10 out of 29) extremely agree with the idea to invite the friends and families to participate in the survey. Also, the participants who were agreeing with the idea of recommending the survey to close friends and families rated as 31% (9 out of 29). Neutral participants had the level of 27.6% (8 out of 29) of rating the question of close friends and families. Finally, the disagree and very disagree participants evaluated the same level of values as 3.4% (1 out of 29). Figure 21e reflects the participants' verdict over the question 5.

The open-ended question also reflected the comments of the participants, and they mentioned some valuable comments such as "Following the pandemic, having access to a digital health system seems a necessity".

Table 8 summarizes the 5 questions rated by 29 participants and how the average satisfaction rate corresponds to each participant's verdict.

In summary, the overall satisfaction rate stands on 3.95 out of 5. That is 79\% which suggests the user satisfaction and engagement is at a promising rate. As it can be observed from Table 8, most of the participants rated all the questions as an average of 4 out of 5, that is 80%. Although this rate is promising there are still chances to improve by designing the questions to increase the satisfaction rate of the participants and they feel more connected to questions, and reflect a higher rate of engagement with any set of questions. As the results suggested there are high satisfaction rates reported by the participants in relevance to the responses. The questions were easy to comprehend and respond to, not as time-consuming as a real psychiatrist session, text messaging has a higher degree of preference compared to talking in a consultation session. Also, most of the participants suggested that the sequence of questions direct them to share the severity of depression.

Table 8 - User Satisfaction and Engagement

Participants	*Q 1*	*Q 2*	*Q 3*	*Q 4*	*Q 5*	*Average Satisfaction Rate*
Participant 1	3	4	4	3	3	3.4
Participant 2	3	3	3	3	4	3.2
Participant 3	3	3	3	4	1	2.8
Participant 4	3	3	3	3	3	3
Participant 5	5	4	4	3	5	4.2
Participant 6	5	5	4	4	4	4.4
Participant 7	5	5	5	5	5	5
Participant 8	4	4	3	3	4	3.6
Participant 9	5	5	3	4	5	4.4
Participant 10	4	3	2	3	3	3
Participant 11	4	3	5	4	4	4
Participant 12	5	4	3	3	5	4
Participant 13	5	5	5	5	5	5
Participant 14	5	4	4	3	5	4.2
Participant 15	4	4	2	3	3	3.2
Participant 16	4	3	3	4	4	3.6
Participant 17	3	3	3	3	3	3
Participant 18	4	4	3	4	2	3.4
Participant 19	5	5	4	4	5	4.6
Participant 20	5	5	5	4	4	4.6
Participant 21	3	5	4	5	3	4
Participant 22	4	3	5	5	3	4
Participant 23	4	4	5	3	4	4
Participant 24	4	5	5	2	4	4
Participant 25	3	3	4	5	5	4
Participant 26	4	4	2	2	3	3
Participant 27	4	4	4	4	4	4
Participant 28	5	1	5	5	5	4.2
Participant 29	4	3	5	5	5	4.4

CHAPTER 5 AUTOMATIC SCORING

5.1 Introduction

This chapter addresses the NLP techniques to perform Sentiment Analysis and calculate depression scores automatically (Autoscoring). Then, the experiments with multiple state-of-the-art ML architectures are conducted to acquire the highest performance model. The ML algorithms and the theory behind them are briefly explained in this chapter. This includes discussion about the data pre-processing, the feature extraction, classification, and ML model development to automate the severity scoring. Finally, the results of the experiments reviewed and the performance of the models compared for accuracy.

5.2 AutoScoring Model

Autoscoring model is discussed in this section. As per the Figure 22 each question out of the total of 27 questions, passes through three main stages: data pre-processing, feature extraction and classification. We discuss each stage of auto scoring in the following context.

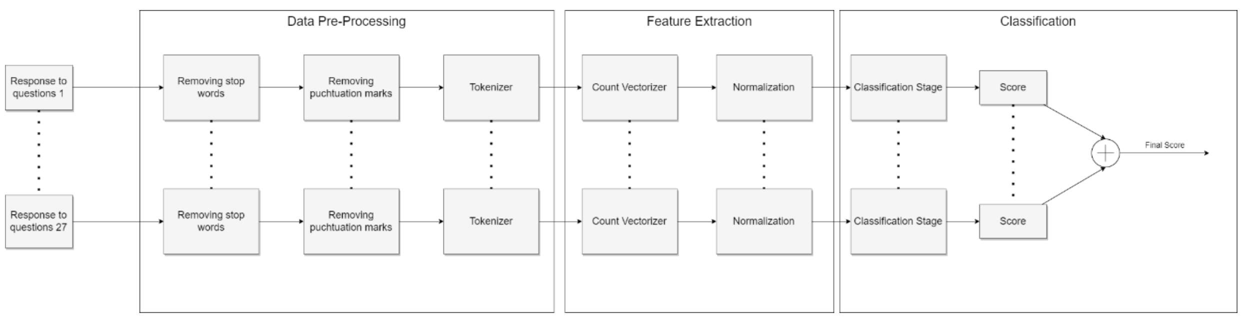


Figure 22 – Auto Scoring Block Diagram

5.2.1 Pre-Processing Stage

When building ML models, there are two types of data to work with, categorical and numeric continuous data. The categorical data is made up of categories of a finite set of values with subcategories of ordinal and nominal that is typically used to express categories or classes. The continuous data takes on numeric values which can be divided further into two subcategories, ratio scale and interval scale.

To represent numeric data, you are able to calculate mean, standard deviation and correlation while to analyse the categorical data you can tabulate data and count frequency using histograms or represent percentages using pie charts. Other forms of data including text and images must be transformed to one of these forms before we can build a ML model. When predicating numeric data, we are using regression models and when predicting categorical data, classification ML models are used.

In this research, the data is a mix of numerical data of participants' demographic segmentation of age, gender and education level as well as the participants' responses to DEPRA chatbot that falls under the nominal categorical data. Therefore, in order to build a ML model to predicate the existence of the depression and to determine the severity of the disorder extracted from the responses to the bot questionnaires, we had to encode the text responses to numerical values that was completed as part of pre-processing stage using the NLP to classify responses into positive, neural, negative and strong negative.

Our dataset was a multiclass classification task which is unlike binary classification that has only the notion of 0 or 1, the outcomes classified in 5 categories 0 (no depression), 1 (mild depression), 2 (moderate depression) and 3 (severe depression) or 4 (very severe depression) classes. Figure 23 is a visual representation of a sample scatter plot of multiclass classification dataset.

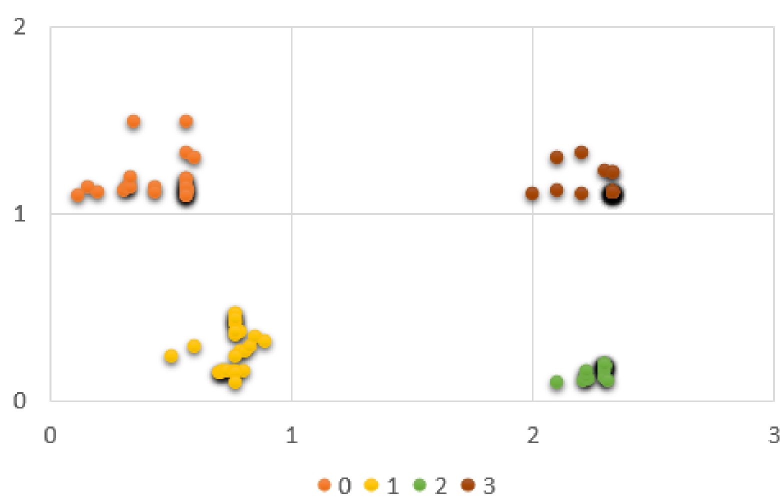


Figure 23 – Scatter Plot Sample of Multiclass Classification Dataset

During the data pre-processing stage using Jupyter Notebook which is an open source web-based interactive development platform, started with importing the required Python libraries and modules. Then, loaded the collected raw data from data collection phase in the dataset format. The dataset contained 50 observations, 60 variables and 1 target variable, the sentiment column indicating the severity of depression. The raw text data included the set of characters and punctuation sings were then examined and cleaned to build text classifier.

The following pre-processing steps completed: i) removing all the stop words such as 'the', 'is', 'at', etc. which do not have any value and the punctuation marks, such as exclamations points, question marks, commas, semi colons, periods, parentheses, brackets, apostrophe, quotation marks, and ellipsis to prepare a clean set of data, ii) conversion to lower case, iii) stemming, the goal of stemming is to reduce the inflection in words to their root forms which helps in decreasing the size of the vocabulary space, iv) tokenization, to split the text into smaller units also known as Tokens, and v) generate visual representations of the cleaned data such as word cloud which is a technique for visualising frequent words where the size of the words represents their frequency. Figure 24 shows a sample of word cloud in Python.

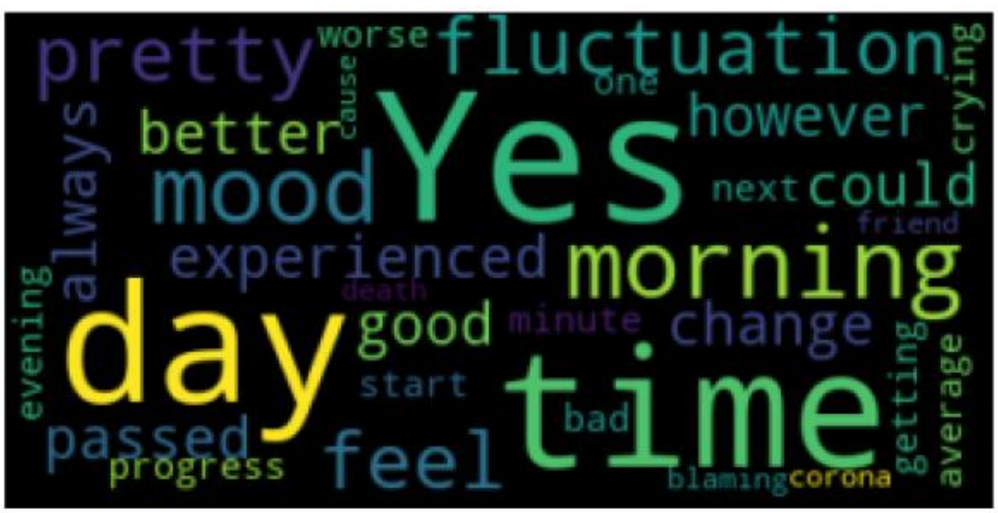


Figure 24 – Word Cloud in Python

5.2.2 Feature Extraction

The process of feature extraction includes count vectorizer and normalization steps. Count vectorizer and the code generated for it, directs us towards evaluation of each five ML algorithms accuracies assessments. Count vectorizer used to create a bag of words. This vectorizer is preferred to other available vectorizers as it converts the text data into machine-understandable format. As a result, the words are dealt with as vectors and they are converted into all lowercase letters. In theory, we have words connected to a numeric value, however, in practice, any words is represented by an index which is connected to a numeric value [47]. One of the main steps to apply count vectorizer is to define a method to print the output size. By importing the count vectorizer and allocation of a document to it, the next step is the creation of a vectorizer object. We can print the identified unique words in accordance with their indices. Then, we encode the document by transforming the document into a vector. The final step is to convert the vector into an array and print the output. Another step in feature extraction stage is the normalization of data. Data normalization is the process to detect and delete the duplicate data. First of all, we group the redundancies and secondly if they are not dependant, we remove one of them from the dataset. Normalization is basically designed to increase the usability of the data. The achievement of normalization are easiness of visualization, insights extraction efficiency and more robust updating the data [48]. Moreover, normalized databases are in logical size and they do not take much space, they reduce the hassles of too many disk space, and the performance are at a high level. There are three types of normalization: first, second and third normal forms.

Up to this point, we have processed the text, however, for building ML models we need to convert the processed text to word frequency vectors to be able to generate a matrix encoded in numeric values. There are multiple techniques to perform vectorization. We applied the count vectorizer method.

The cleaned data along with tokenizing then used for further analysis. We also split the dataset into two categories, train dataset and test dataset. The value and percentage of each set is 40% for testing and 60%

for training. The output of both train and test are matrix of token counts. The test matrix shape is a 24 to 55 dimensions and the test matrix shape is a 16 to 55 dimensions. Pre-processing is a critical stage in auto scoring that is regarded as the basic preparation of the collected data.

By applying counter vectorizer on the responses of each question, the number of feature in our datasets will be different for each question. Because the number of unique words is different for each question's responses. The following Table 9 shows the questions number and the corresponding number of features resulted from applying counter vectorizer.

Table 9 –Number of Features per Question for Count Vectorizer

Question No.	Number of features
1	55
2	62
3	72
4	124
5	94
6	91
7	117
8	116
9	39
10	17
11	97
12	74
13	53
14	74
15	83
16	105
17	110
18	115
19	34
20	18
21	67
22	49
23	33
24	29
25	57
26	33
27	73

By applying the components of feature extraction, count vectorizer and normalization, the dataset is ready to be used for building the text classification ML models in the next stage of auto score model [48] followed by the evaluation of the performance of each model on test data using an accuracy score function to compare the accuracy.

5.2.3 Classification Stage

At classification stage, a common work of ML algorithms is evaluated. ML algorithms can distinguish the objects and the objects can be categorized. This process helps to set a large data into manageable and more organized values. For each question we performed the classification and at the next stage we calculated the score for each question. At the final step, we have summated the separate scores of each question and the overall score is calculated [49]. The final score is analysed to distinguish the category that the participant belongs to it. In other words, the participant can have no symptoms and being completely healthy, they can experience mild depression, they can have moderate depression, severe depression can also be allocated to a participant and finally some participants can suffer from very severe symptoms.

Popular algorithms that can be used for multi-class classification include:

- Decision Trees
- Naive Bayes
- Random Forest
- Logistic Regression
- Support Vector Machine

In our research, we build our models with Naive Bayes, Logistic Regression and Support Vector Machine (SVM) algorithms. To complete this list of algorithms, we also applied the Linear SVC ML algorithm to find the best algorithm in this research.

5.2.3.1 Logistic Regression

It is basically a technique considering ML which is designed by the field of statistics. Logistic regression is dealt with as a solution for problems with two values such as True/False, 0/1 and so forth [50]. Classifying and forecasting are important applications of Logistic Regression.

To mention the logistic regression there are other names to refer to this algorithm, such as log-odds, logistic function, and logit. There are some methods used to learn about coefficients and train this model to apply the required data. Within this study, we will discuss how to make predictions using this ML algorithm.

Sigmoid function or logistic function is applied in ecology, rapid movement and carrying capacity with the highest possible value in any conditions. The structure of the graph which is used at logistic regression is an S-shaped curve or the probability curve. Logistic Regression is to calculate the probability curve and determine how probabilities, p(y), can be changed by actions or influenced by causes (X_n).

In Figure 25 there is a visual description of the numbers from -5 to 5 transformed into the range of 0 and 1:

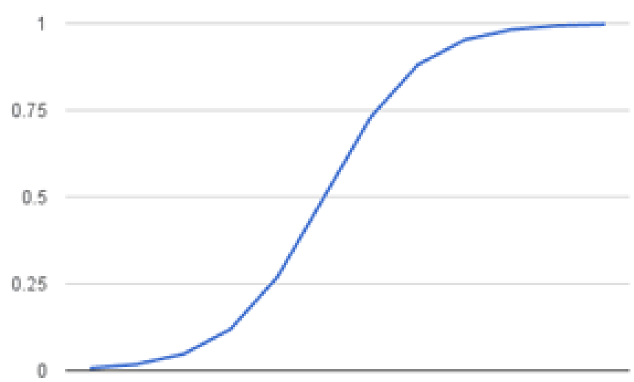


Figure 25 - Logistic Function

In this graph, any real number will be mapped into a value between 0 and 1. The equation, eq 3, that is applied is:

$1/(1 + e^{\wedge}(-value))$ (3)

where *e* is the Euler's number, it is also known as EXP() function in excel sheets. *Value* is the number that we would like to transform.

With this introduction and familiarity with logistic function let us consider the application of this function in logistic regression [51, 52].

5.2.3.1.1 Representation used for Logistic Regression

The equation which is applied for logistic regression is very similar in content with linear regression. Input values (x) linearly reflect coefficient values, displayed with Beta, and they generate output values (y). The main difference between logistic and linear is that the output matches a binary value not a numeric one. Here is a sample of logistic regression equation, eq 4:

$\boldsymbol{y} = \boldsymbol{e}^{\wedge}(\boldsymbol{b0} + \boldsymbol{b1} * 1\boldsymbol{x})/(\boldsymbol{1} + \boldsymbol{e}^{\wedge}(\boldsymbol{b0} + \boldsymbol{b1} * \boldsymbol{x}))$ (4)

where y is the target output, b0 is bias and b1 is coefficient factor for x. There is a real value which is learnt from training data which belongs to each column of the x. The beta values are stored in memory and they are in fact forming the model representation [50, 52].

5.2.3.1.2 Training the logistic regression model

Beta values denoted by b in the formula, will be the result of estimation on the training data. Maximum-Likelihood Estimation (MLE) is applied. MLE is a learning algorithm utilized on different ML algorithms. This learning algorithm makes assumptions over distribution of the data. Two classes are considered. The selected coefficients lead to designing a model that can predict either 1 (default class) or 0 (various classes). Intuition for logistic regression tries to locate beta values in order to lower the probabilities generated by the model with regard to accessible data (for instance, 1 in the data of primary class).

The math behind the formula of MLE is not discussed in this study. The point that we should mention is that minimization algorithm is applied to optimize the best values for beta values in the training data. In practice, the implementation of this purposed is best served by using efficient numerical optimization algorithms which are available. For instance, Quasi-newton method can be a remarkable solution. When we need to learn logistic, implementation from scratch with utilization of Gradient Descent Algorithm (GDA) can be a logical choice [50, 53]. The optimization method which is applied throughout this study by default is 'lbfgs'. This optimization method is popular for parameter estimation in ML.

5.2.3.1.3 Prepare Data for Logistic Regression

The concept of making assumptions designed by logistic regression over distributions and relationships in our data are similar to making assumptions in linear regression.

There have been a considerable study over definition of assumptions, probabilistic and statistical knowledge. It is recommended that apply the suggested method or guideline and examine various data preparation schemes.

The interpretation of results are of less importance to be considered when we should focus on making accurate predictions. To achieve this goal, we should disregard some assumptions as long as the model is powerful and performs in a way that it supposed to be working.

- **Binary Output Variable**: As we just discusses, the logistic regression is based on two classes with the default values of 0 or 1. They are the two binary values per class.
- **Remove Noise**: logistic regression assumes that the data is clean and there are no errors in the output variable y. All the tolerances and mismatches have been removed from the data which is trained.
- **Gaussian Distribution**: Logistic regression is a linear algorithm and the output is non-linear transform. Input and output variables are benefiting from linear relationships. A more accurate model is achieved through better exposure of the linear relationship. For instance, the application of log, root, Box-Cox and other univariate transforms can generate a better result.

- **Remove Correlated Inputs:** Similar to linear regression, the logistic regression model can overfit if various highly correlated inputs are considered for this model. By considering correlations between the inputs and removing the ununiformed inputs we can improve the outcome.
- **Fail to Converge:** To avoid this issue, we can avoid making the data very sparse. Too many highly correlated inputs in the data can also fail to converge. As an example, too many zeros in the input can create this issue [54].

5.2.3.2 Naïve Bayes

In the following section, two types of Naïve Bayes algorithms are explained. First of all, Bernoulli NB and its characteristics and equation is discussed. Secondly, Multinomial NB and its concept is explained with more details.

5.2.3.2.1 Bernoulli Naïve Bayes

The Multivariate Bernoulli model is based on binary variables [55]. The variables are independent Booleans which are playing the main role to describe and define the inputs. This model is applied in document classification, just similar to the multinomial model, and binary terms features are used instead of term based on frequencies. Imagine x_i is a Boolean; This x_i expresses absence from the dictionary, then the probability of a source is given through a class C_k is given by eq 5:

$$\rho(x|C_k) = \prod_{i=1}^{n} \rho_{k_i}^{x_i} \left(1 - \rho_{k_i}\right)^{(1-x_i)} \qquad (5)$$

where p_{ki} is the probability of class C_k generating the term x_i.

The usage of this model is in short texts and it is not recommended for long texts. The absence of terms is clearly explained in this model. In multinomial NB classifier frequency counts are truncated to one while in Bernoulli event model this will not be the case [56, 57].

5.2.3.2.2 Multinomial Naïve Bayes

In the concept of multinomial model, feature vectors, also known as samples, demonstrate the frequencies. These frequencies include certain events which are generated by a multinomial, that is, ($p_1, \ldots, p_n$). pi is the probability of i to occur. In multiclass case the term would be K. X = ($X_1, \ldots, X_n$) is considered as feature vector which is basically a histogram. In the histogram Xi is the number of times even i happened in an instance [58]. This is the method which is applied to document classification with the events to be classified as the possibility of a word to occur in a document. A histogram X and its occurrence possibility is given by the following equation, eq 6:

$$\rho(x|C_k) = \frac{(\sum_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \rho_{ki}^{x_i} \qquad (6)$$

Log-space expression of multinomial naïve Bayes will turn this model into linear classifier by eq 7:

$$\log \rho(C_k|x) \propto \log\left(\rho(C_k) \prod_{i=1}^{n} \rho_{ki}^{x_i}\right)$$

$$= \log \rho(C_k) + \sum_{i=1}^{n} x_i \cdot \log \rho_{ki}$$

$$= b + w_k^T x \qquad (7)$$

Where $\mathrm{b} = \log \rho(C_k)\ \ and\ w_{ki} = \log \rho_{ki}$

In case of a class and feature value do not occur at the same time in the training data, the frequency-based probability value would be at zero level. The reason behind this probability estimate is directly related to the fact that the number of occurrences of a feature's value is none [59].

There is a serious issue with this fact and that is the information of other probabilities will be turned to zero when there is a multiplication between the probabilities. Small-sample correction, known as Pseudocount, is a practical policy to incorporate. With this strategy, even the probability estimates like no probabilities would not be set exactly as zero. There would be a value very small but still it does not affect other probabilities in case of multiplications. This process in Naïve Bayes is known as Laplace smoothing in case the pseudocount is one value and lidstone smoothing in other cases [56, 60].

5.2.3.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised ML algorithm. There are two applications defined for SVM, classification challenges and regression challenges. However, the classification application is mostly used compared to regression challenges. In the SVM algorithm, each data item will be plotted into a n-dimensional space where n represents the number of features. The feature value would be mapped into the value of a coordinate. The next step will be classification performance which is handled by hyper plane [61]. This hyper plane is the drawing which explicitly differentiates two classes.

Support Vectors coordinate with the outstanding monitoring. Hyper plane and line are two features of SVM that explicitly distinguish and separate from each other in a visual representation.

As discussed so far, the role of identifying the relevant hyper plane is the key in SVM algorithm definition. The target is to identify the relevant hyper plane which distinguishes the starts and circles the best. The thumb rule is to choose the best hyper plane that clearly separates and distinguishes the classes. In this scenario, we can observe that hype plane B is doing the best separation, so we choose it as the selected hyper plain [62, 63].

5.2.3.3.1 Linear SVC vs SVM

SVM and Linear SVC are the same algorithms in ML. They are considered as different implementations variety over the same algorithm. Linear SVC is considering a library for linear applications, and it only covers a linear kernel. On the other hand, SVM is a wrapper over a library that can utilize a variety of kernels. There are pros and cons related to each implementation and Linear SVC with linear kernel support brings relatively quicker implementation and it also plot much better than the SVM [62, 64]. Linear SVC is one of the selected algorithms for the autoscoring which is applied in this research.

5.2.3.4 SGD Classifier

Stochastic Gradient Descent is a method of optimization for unrestrained optimization problems. In SGD, the true gradient and value of $E(w, b)$ is calculated based on single training instance at a specific time. Batch Gradient Descent implements an approach in contrast to SGD [65].

The SGD Classifier class designs a first-order SGD learning method. SGD algorithm evaluates the training instances and for each instance it updates the parameters by the following rule, eq 8:

$$w \leftarrow w - \eta \left[\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial L(w^T x_i + b, y_i)}{\partial w}\right] \qquad (8)$$

where η is the learning factor to define parameter space. b value is updated with the same method, however, it does not include regularization. b value adds decay to matrices.

The learning factor of η has the flexibility of being constant or slowly decaying. The default learning factor equations is as the following eq 9:

$$\eta^{(t)} = \frac{1}{\alpha(t_0 + t)} \qquad (9)$$

where t stands for time step. t_0 defines expected initial updates in parallel to expected size of the weights.

The equation for the defaults learning factor is designed in inverse scaling as per the following eq 10:

$$\eta^{(t)} = \frac{eta_0}{t^{power\_t}} \qquad (10)$$

where eta_0 and $power\_t$ are the hyperparameters selected by the similar values.

The algorithm continues to perform until the stopping indicator is approached. At this stage, the learning factor is divided by five and the algorithm keeps running. The only occasion that the algorithm stops completely is when the learning factor reaches on or below 1e-6. In order to access the weights and b parameter, there are two attributes defined in SGD Classifier. coef_ for w and intercept_ for b. The equation, eq 11, for Averaged SGD is: $coef\_ = \frac{1}{T}\sum_{t=0}^{T-1} w^{(t)}$, where T is updates overall value according to t_ attribute [66, 67]. SGD Classifier was one of the selected ML algorithms within one of the experiments that we had for this study. This algorithm can be applied to calculate the autoscoring and compare it with the manual scoring which was conducted manually by the research team.

5.3 Results and Discussion

The accuracies in percentage of the trained models for each five algorithms reflect the three experiments on various groups of participants whose responses were collected on three stages of the research lifecycle. The first experiment is based on initial data collection phase with a focus group size of 26 participants. The second experiment is based on 40 participants and the size of samples in the third which is also the last experiment includes 50 participants. The results are presented in this section. In order to calculate the accuracies of the applied ML algorithms, we have used two methods. In the first method, we have applied metrics.accuracy_score function on Linear Support Vector Machine (SVC) and Stochastic Gradient Descent (SGD) classifiers to calculate accuracy of the classification models and to quantify the quality of the prediction. In the second method, we have used confusing matrix to examine the three Logistic Regression, Bernoulli Naïve Bayes and Multinomial Naïve Bayes models.

Table 10 represents the confusion matrix. By achieving more True Negatives and True Positives a more accurate and precise model was generated.

Table 10 - Confusion Matrix

		Actual Class	
		P	**N**
Predicted Class	**P**	***True Positive***	***False Positive***
	N	***False Negative***	***True Negative***

Finally, the accuracy of the model was calculated. As per the following, the accuracy is based on the summation of True Positives and True negatives over Total Number of Classes.

$$Accuracy = \frac{(True\ Positives + True\ Negatives)}{(Total\ number\ of\ classes)} \quad (12)$$

Typically, the trend of the Python code commences by accessing the dataset. Then, training the ML models and evaluating the accuracy of the trained models is the next steps. In our case, the dataset consists of responses of focus group who have already interacted with the DEPRA chatbot. As the first experiment includes 26 participants, the accuracy levels for each question per the selected ML algorithm was calculated and compared. We have also plotted the Receiver Operating Characteristic Curve (ROC) curve that is a graph illustrating the performance of a classification model at all classification thresholds. Table 11 summarizes the percentages and the applicable algorithm for each question.

Table 11 - Five Algorithms and their Accuracies per each 27 Questions (26 Participants)

Question number	For 26 Participants					
	Logistic Regression	**Bernoulli Naïve Bayes**	**Multinomial Naïve Bayes**	**Linear SVC**	**SGD Classifier**	**Applicable Algorithm**
1	18%	17%	18%	67%	67%	Linear SVC
2	24%	22%	22%	67%	67%	Linear SVC
3	79%	68%	50%	83%	75%	Linear SVC
4	79%	68%	50%	83%	67%	Linear SVC
5	47%	25%	61%	92%	92%	Linear SVC
6	46%	46%	100%	92%	100%	SGD Classifier
7	44%	25%	42%	83%	75%	Linear SVC
8	50%	40%	58%	75%	67%	Linear SVC
9	67%	67%	58%	75%	83%	SGD Classifier
10	100%	100%	100%	100%	100%	Linear SVC
11	100%	100%	100%	100%	100%	Linear SVC
12	79%	35%	79%	83%	83%	Linear SVC
13	54%	22%	22%	83%	75%	Linear SVC
14	33%	46%	33%	92%	83%	Linear SVC
15	64%	28%	50%	92%	92%	Linear SVC
16	58%	58%	58%	58%	67%	SGD Classifier
17	83%	61%	83%	92%	75%	Linear SVC
18	92%	92%	83%	100%	92%	Linear SVC
19	67%	58%	67%	67%	67%	Linear SVC
20	100%	100%	95%	100%	92%	Linear SVC
21	92%	92%	92%	92%	75%	Linear SVC

22	75%	33%	75%	67%	83%	SGD Classifier
23	92%	83%	92%	100%	75%	Linear SVC
24	100%	92%	92%	100%	100%	Linear SVC
25	92%	92%	92%	92%	92%	Linear SVC
26	92%	92%	92%	83%	83%	N/A
27	92%	92%	92%	92%	75%	Linear SVC

The strategy for distinction of applicable algorithm for each question is the highest accuracy value for each question. If the values are identical for two or more different questions, we have chosen the accuracy that has a more repetition within all the other questions. For instance, Linear SVC and SGD Classifier both have the accuracy of 67% for question 1. The Linear SVC is selected more as the applicable algorithm for several questions. As a result, it is selected as the preferred algorithm for question 1. Table 12 and Figure 26 summarize the process of selecting the applicable algorithm for 26 participants.

Table 12 - ML Algorithms Count of Appearance (26 Participants)

ML Algorithms	Number of Questions used the algorithm
Linear SVC	22
SGD Classifier	4

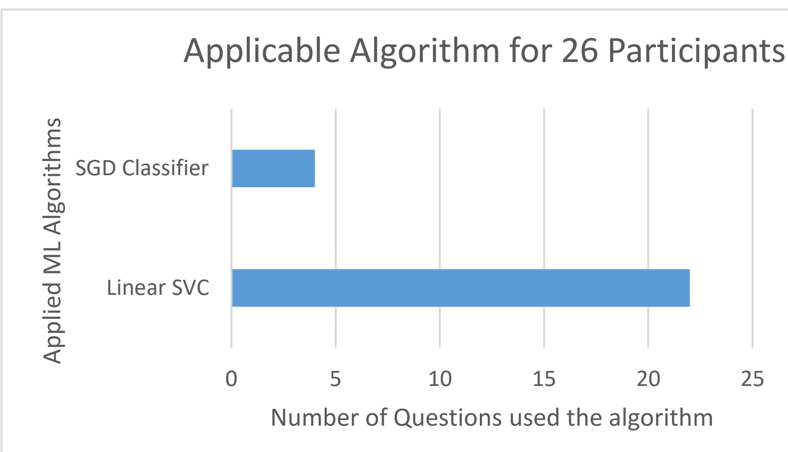


Figure 26 - Applicable Algorithm for DEPRA chatbot for 26 Participants

Moreover, the most applicable algorithm for 26 participants is Linear SVC. It was selected by 22 questions out of 27 possible questions in the dataset. There is only one more selected algorithm, SGD Classifier, by 4 questions out of 27. Other algorithms were not selected within this experiment. The average accuracy of Linear SVC within 26 participants is 88%.

The second experiment based on 40 participants. The same method applied for this experiment and the strategy to select the applicable algorithm was the same as the previous group of participants. For instance, for question 1, Logistic Regression and Bernoulli Naïve Bayes has the accuracy of 69%. Due to the fact that Logistic Regression reflects a higher value in the rest of the questions we have selected it as the applicable algorithm for this question. These statics are based on the responses from 40 participants and they are all summarized in Table 13.

Table 13 - Five Algorithms and their Accuracies per each 27 Questions (40 Participants)

Question number	For 40 Participants					
	Logistic Regression	**Bernoulli Naïve Bayes**	**Multinomial Naïve Bayes**	**Linear SVC**	**SGD Classifier**	**Applicable Algorithm**
1	69%	69%	56%	63%	56%	Logistic Regression
2	44%	38%	31%	31%	31%	Logistic Regression
3	56%	63%	44%	56%	44%	Bernoulli NB
4	75%	69%	69%	75%	69%	Logistic Regression
5	50%	44%	44%	50%	56%	SGD Classifier
6	88%	81%	81%	88%	81%	Logistic Regression
7	63%	44%	63%	81%	63%	Linear SVC
8	56%	31%	56%	56%	56%	Logistic Regression
9	69%	88%	81%	81%	88%	SGD Classifier
10	88%	88%	88%	94%	94%	SGD Classifier
11	50%	50%	56%	50%	50%	Multinomial NB
12	75%	69%	69%	69%	75%	SGD Classifier
13	69%	50%	63%	69%	69%	SGD Classifier
14	56%	56%	63%	56%	75%	SGD Classifier
15	81%	75%	94%	81%	75%	Multinomial NB
16	63%	69%	69%	63%	56%	Multinomial NB
17	50%	63%	69%	63%	63%	Multinomial NB
18	50%	44%	63%	50%	50%	Multinomial NB

19	88%	69%	75%	94%	94%	SGD Classifier
20	69%	69%	94%	69%	94%	SGD Classifier
21	81%	50%	69%	75%	69%	Logistic Regression
22	44%	50%	44%	44%	38%	Bernoulli NB
23	63%	50%	50%	63%	50%	Logistic Regression
24	75%	63%	75%	75%	75%	SGD Classifier
25	69%	75%	75%	69%	75%	SGD Classifier
26	75%	75%	75%	81%	81%	SGD Classifier
27	75%	75%	81%	75%	69%	Multinomial NB

This experiment evaluates the ML algorithms, such as Logistic Regression, Naïve Bayes (Bernoulli and Multinomial), Linear SVC and SGD Classifier. The accuracy of each ML algorithm is calculated. Table 14 and Figure 27 summarize the process of selecting the applicable algorithm for 40 participants.

Table 14 - ML Algorithms Count of Appearance (40 Participants)

ML Algorithms	Number of Questions used the algorithm
SGD Classifier	11
Logistic Regression	7
Multinomial NB	6
Bernoulli NB	2
Linear SVC	1

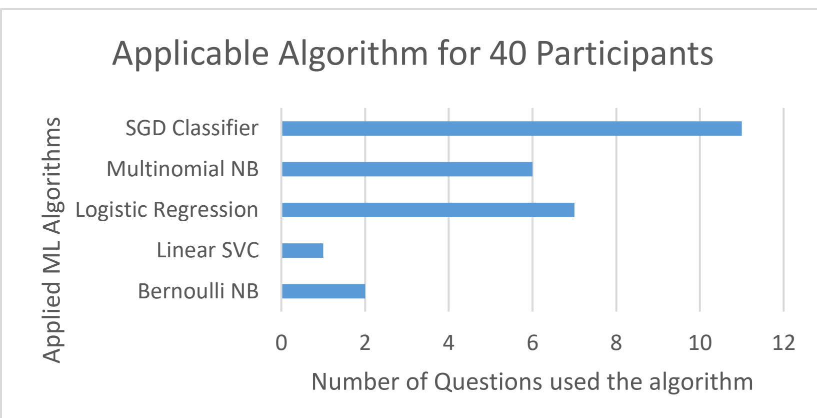


Figure 27 - Applicable Algorithm for DEPRA chatbot for 40 Participants

With this accuracy of classifiers and the application of Count Vectorizer as the feature extraction method, SGD Classifier has the highest level of application with 11 out of 27 questions accuracy rate. Logistic Regression and Multinomial Naïve Bayes followed the SGD Classifier with 7 and 6 out of 27 respectively. Bernoulli Naïve Bayes with 2 out of 27 questions followed by Linear SVC that has the lowest level of accuracy with only 1 out of 27 questions popularity. SGD Classifier appeared to be the most applicable algorithm with 80% accuracy rate.

The third experiment based on 50 participants. That is, the entire dataset details applied for this experiment. In case of the accuracies of a specific question are identical, we apply the same method as the previous two experiments to select the most applicable algorithm according to the registered values for each question. Table 15 summarizes the accuracies for the five algorithms according to 27 questions and 50 participants.

Table 15 - Five Algorithms and their Accuracies per each 27 Questions (50 Participants)

Question number	For 50 Participants					
	Logistic Regression	**Bernoulli Naïve Bayes**	**Multinomial Naïve Bayes**	**Linear SVC**	**SGD Classifier**	**Applicable Algorithm**
1	77%	77%	40%	80%	83%	SGD Classifier
2	56%	67%	55%	70%	65%	Linear SVC
3	45%	45%	78%	80%	80%	Linear SVC
4	65%	87%	80%	77%	77%	Bernoulli Naïve Bayes
5	60%	60%	60%	73%	70%	Linear SVC
6	100%	95%	95%	90%	90%	Logistic Regression
7	90%	90%	75%	70%	95%	SGD Classifier
8	57%	57%	65%	85%	77%	Linear SVC
9	90%	90%	100%	90%	95%	Multinomial Naïve Bayes
10	75%	75%	100%	100%	90%	Linear SVC
11	90%	90%	95%	95%	100%	SGD Classifier
12	83%	78%	78%	65%	78%	Logistic Regression
13	75%	84%	97%	90%	90%	Multinomial Naïve Bayes
14	88%	88%	90%	95%	95%	Linear SVC
15	100%	100%	100%	100%	100%	Linear SVC
16	90%	85%	78%	78%	80%	Logistic Regression
17	85%	85%	85%	90%	95%	SGD Classifier
18	90%	90%	75%	65%	77%	Logistic Regression
19	55%	55%	65%	75%	75%	Linear SVC
20	95%	95%	100%	85%	77%	Multinomial Naïve Bayes

21	75%	75%	66%	66%	80%	SGD Classifier
22	100%	100%	100%	100%	100%	Linear SVC
23	83%	83%	90%	90%	90%	Linear SVC
24	56%	56%	65%	70%	75%	SGD Classifier
25	77%	77%	83%	67%	67%	Multinomial Naïve Bayes
26	100%	100%	95%	90%	90%	Logistic Regression
27	56%	56%	60%	73%	78%	SGD Classifier

Table 16 - ML Algorithms Count of Appearance (50 Participants)

ML Algorithms	Number of Questions used the algorithm
Linear SVC	10
SGD Classifier	7
Logistic Regression	5
Multinomial Naïve Bayes	4
Bernoulli Naïve Bayes	1

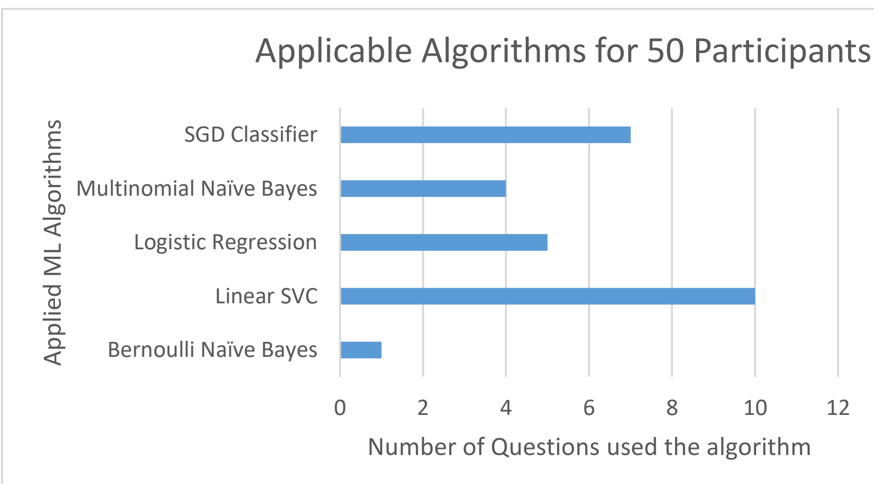


Figure 28 - Applicable Algorithm for DEPRA chatbot for 50 Participants

Table 16 and Figure 28 reflect the outcome of selection process to choose the most applicable algorithm. Linear SVC with 10 out of 27 questions is the most applicable algorithm. SGD Classifier is the next most applicable algorithm with 7 out of 27 questions. Logistic Regression and Multinomial Naïve Bayes ranked the third and fourth algorithms with 5 and 4 out of 27 questions, respectively. Bernoulli Naïve Bayes was only selected in one question, so it ranked 1 out of 27 questions. Linear SVC appeared to be the most applicable algorithm with 87% accuracy rate. As per the scoring system, the overall scoring was calculated and the participants received a verdict over their health. To evaluate the autoscoring feature of DEPRA, the accuracy of the ML algorithms implemented is calculated. Accordingly, manual scoring is applied and compared with the calculated depression scores.

CHAPTER 6 CONCLUSION

This study supports the fact that AI Botics offer an accessible mental health solution for early detection. Although Chatbots are not to replace any traditional examination diagnostic methods, they have potential to serve as a scalable tool to complement such methods. As the result of COVID-19 pandemic in 2020, access to digital health care systems is more critical than any time before. Digital assessors to assist the medical staff and to facilitate collecting the symptoms to ease the burden on the population who might be more vulnerable than others and in need of constant access to the professional consultations become a necessity. In this research, we have designed and implemented DEPRA chatbot on a contemporary platform, DialogFlow, based on the SIGH-D and IDS. In this research, 50 participants participated and successfully completed the chatbot conversation sessions and their responses are collected and analysed. To find the depressions level, two scoring systems are utilized, IDS-SR and QIDS-SR. The results of this study support and expand on previous studies and demonstrate that psychological AI has the promising role and ability to collect data from users participants to be examined further and with the use of ML models the users' responses can be analysed automatically to predict the existence of depression and its severity. In this research, SGD classifiers and SVC Linear based models proved to have more accuracy considering 27 questions extracted from SIGH-D rating scale examining users' depression symptoms including mood, reactivity of mood, future outlook, guilt, insomnia, suicidal ideation, anxiety, involvement, concentration, psychomotor, somatic, symptoms, general, pleasure, weight, interpersonal sensitivity and genital symptoms to assist medical science with the application of early detection of depression.

6.1 Limitations and Future Work

In this study we have focused on sizable focus group. However, in the future, we can increase the sample size which leads to dataset enlargement. That will improve the accuracy for text classification methods, and an efficient and desired output. Furthermore, future work could integrate the ML models into DEPRA chatbot to not only collect the users data but intelligently interacts with the users. Moreover, the idea of adding a statement after receiving a response from the participant and share a sympathy or to share a helpful comment with them will also be applied in the future implementations. The participants stated that they will feel more connected to the chatbot if a mutual understanding though a text message can relate them more to the bot. Data collection through organic participants who are concerned about their health, and they commence interaction with DEPRA chatbot through Facebook page but for some technical or personal issues they do not proceed to complete the survey should be addressed in future implementations. We can ask for ethical approval to be modified in some sections and give us the opportunity to have a more inclusive list of potential participants which leads to a more comprehensive dataset. The datasets can be divided into several phases. At phase one, total of 100 participants are targeted. They can be Facebook friends, organic participants, or even closed group. In other phases, we will invite more participants to interact with the chatbot and we can analyse the collected data from the previous phases.

Also, the autoscoring vs manual scoring requires to be discussed in more details in the next phase of this research. The methods to be used, the way they can be presented and the general comparison between the scoring systems should be reflected with more details.

Lastly, future work could also include the Explainable Artificial Intelligence (XAI), an emerging field of research to build trust by exploring the results of AI-Botics solutions and ML models on how the specific AI black-box decisions are made and providing explanations that can be understood by human experts.

APPENDICES

Appendix A – Fulfillment Code for DEPRA Chatbot Design

Microsoft Word Document, 25KB-Multimedia Appendix A

Appendix B – Ethics Application related to this study

Microsoft Edge Pdf Document, 293KB-Multimedia Appendix B

Appendix C – Consent Form for participants involved in research

Microsoft Word Document, 46KB-Multimedia Appendix C

Appendix D – User Rating Form

Google Forms, 98KB-Multimedia Appendix D

Appendix E – Atomic Data of Focused Group Responses

Microsoft Excel Worksheet, 53KB-Multimedia Appendix E

Appendix F – DEPRA Chatbot Scoring – 50 Participants

Microsoft Excel Worksheet, 69KB-Multimedia Appendix F

REFERENCES

[1] W. H. Organization, "Suicide: one person dies every 40 seconds," 9 September 2019. [Online]. Available: https://www.who.int/news/item/09-09-2019-suicide-one-person-dies-every-40-seconds.

[2] W. H. Organization, "Suicide," 17 June 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/suicide.

[3] W. H. Organization, "Depression," 13 September 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[4] B. D. Institute, "Facts & figures about mental health." [Online]. Available: https://www.blackdoginstitute.org.au/wp-content/uploads/2020/04/1-facts_figures.pdf.

[5] "What Is Depression?," *American Psychiatric Association,* 2021. [Online]. Available: https://www.psychiatry.org/patients-families/depression/what-is-depression.

[6] D.-G. o. t. W. H. O. Dr Tedros Adhanom Ghebreyesus, "The WHO Special Initiative for Mental Health (2019-2023): Universal Health Coverage for Mental Health," (in English), p. 4, 2019. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/310981/WHO-MSD-19.1-eng.pdf?sequence=1&isAllowed=y.

[7] A. A. Abd-Alrazaq, M. Alajlani, N. Ali, K. Denecke, B. M. Bewick, and M. Househ, "Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review," (in English), *J Med Internet Res,* Review vol. 23, no. 1, p. e17828, 2021, doi: 10.2196/17828.

[8] Z. Steel *et al.*, "The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013," *International journal of epidemiology,* vol. 43, no. 2, pp. 476-493, 2014.

[9] A. A. Abd-Alrazaq, B. M. Bewick, T. Farragher, and P. Gardner, "Factors that affect the use of electronic personal health records among patients: a systematic review," *International journal of medical informatics,* vol. 126, pp. 164-175, 2019.

[10] W. H. Organization, "Depression and other common mental disorders: global health estimates," World Health Organization, 2017.

[11] B. Given *et al.*, "Burden and depression among caregivers of patients with cancer at the end-of-life," in *Oncology nursing forum*, 2004, vol. 31, no. 6: NIH Public Access, p. 1105.

[12] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, "Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape," *The Canadian Journal of Psychiatry,* vol. 64, no. 7, pp. 456-464, 2019, doi: 10.1177/0706743719828977.

[13] A. B. o. Statistics, "Causes of Death, Australia," (in English), 2021. [Online]. Available: https://www.abs.gov.au/statistics/health/causes-death/causes-death-australia/latest-release.

[14] W. Y. B. L. e. al., ""Federated Learning in Mobile Edge Networks: A Comprehensive Survey", in IEEE Communications Surveys & Tutorials," vol. 22, no. no. 3, pp. 2031-2063, thirdquarter 2020, doi: 10.1109/COMST.2020.2986024.

[15] M. L. Tielman, M. A. Neerincx, R. Bidarra, B. Kybartas, and W.-P. Brinkman, "A Therapy System for Post-Traumatic Stress Disorder Using a Virtual Agent and Virtual Storytelling to Reconstruct Traumatic Memories," *Journal of Medical Systems,* vol. 41, no. 8, p. 125, 2017/07/11 2017, doi: 10.1007/s10916-017-0771-y.

[16] S. D'Alfonso, "AI in mental health," *Current Opinion in Psychology,* vol. 36, pp. 112-117, 2020.

[17] K. H. Ly *et al.*, "Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial," *PloS one,* vol. 10, no. 5, p. e0126559, 2015.

[18] T. Shinozaki, Y. Yamamoto, and S. Tsuruta, "Context-based counselor agent for software development ecosystem," *Computing,* vol. 97, no. 1, pp. 3-28, 2015/01/01 2015, doi: 10.1007/s00607-013-0352-y.

[19] B. Sharma, H. Puri, and D. Rawat, "Digital psychiatry-curbing depression using therapy chatbot and depression analysis," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018: IEEE, pp. 627-631.

[20] R. Fulmer, A. Joerin, B. Gentile, L. Lakerink, and M. Rauws, "Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial," *JMIR mental health,* vol. 5, no. 4, p. e9782, 2018.

[21] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM,* vol. 9, no. 1, pp. 36-45, 1966.

[22] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ‐9: validity of a brief depression severity measure," *Journal of general internal medicine,* vol. 16, no. 9, pp. 606-613, 2001.

[23] R. L. Spitzer, K. Kroenke, J. B. Williams, and B. Löwe, "A brief measure for assessing generalized anxiety disorder: the GAD-7," *Archives of internal medicine,* vol. 166, no. 10, pp. 1092-1097, 2006.

[24] C. Burton *et al.*, "Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression," *Journal of telemedicine and telecare,* vol. 22, no. 6, pp. 348-355, 2016.

[25] R. R. Morris, K. Kouddous, R. Kshirsagar, and S. M. Schueller, "Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions," *Journal of medical Internet research,* vol. 20, no. 6, p. e10148, 2018.

[26] R. R. Morris, S. M. Schueller, and R. W. Picard, "Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial," *Journal of medical Internet research,* vol. 17, no. 3, p. e4167, 2015.

[27] J. Carpenter *et al.*, "Seeing the "big" picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data," *Journal of medical Internet research,* vol. 18, no. 8, p. e5725, 2016.

[28] A. Ahmed *et al.*, "A review of mobile chatbot apps for anxiety and depression and their self-care features," *Computer Methods and Programs in Biomedicine Update,* vol. 1, p. 100012, 2021.

[29] P. M. Gardiner *et al.*, "Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial," *Patient education and counseling,* vol. 100, no. 9, pp. 1720-1729, 2017.

[30] T. W. Bickmore, K. Puskar, E. A. Schlenk, L. M. Pfeifer, and S. M. Sereika, "Maintaining reality: Relational agents for antipsychotic medication adherence," *Interacting with Computers,* vol. 22, no. 4, pp. 276-288, 2010.

[31] T. W. Bickmore, S. E. Mitchell, B. W. Jack, M. K. Paasche-Orlow, L. M. Pfeifer, and J. O'Donnell, "Response to a relational agent by hospital patients with depressive symptoms," *Interacting with computers,* vol. 22, no. 4, pp. 289-298, 2010.

[32] A. Podrazhansky, H. Zhang, M. Han, and S. He, "A Chatbot-based Mobile Application to Predict and Early-prevent Human Mental Illness," in *Proceedings of the 2020 ACM Southeast Conference*, 2020, pp. 311-312.

[33] T. Sharma, J. Parihar, and S. Singh, "Intelligent Chatbot for Prediction and Management of Stress," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021: IEEE, pp. 937-941.

[34] S. Greer, D. Ramo, Y.-J. Chang, M. Fu, J. Moskowitz, and J. Haritatos, "Use of the chatbot “vivibot” to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial," *JMIR mHealth and uHealth,* vol. 7, no. 10, p. e15018, 2019.

[35] J. T. Moskowitz *et al.*, "Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV," *Journal of consulting and clinical psychology,* vol. 85, no. 5, p. 409, 2017.

[36] N. Harilal, R. Shah, S. Sharma, and V. Bhutani, "CARO: an empathetic health conversational chatbot for people with major depression," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 2020, pp. 349-350.

[37] P. Kataria, K. Rode, A. Jain, P. Dwivedi, S. Bhingarkar, and M. India, "User adaptive chatbot for mitigating depression," *International Journal of Pure and Applied Mathematics,* vol. 118, no. 16, pp. 349-361, 2018.

[38] M. L. Tielman, M. A. Neerincx, R. Bidarra, B. Kybartas, and W.-P. Brinkman, "A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories," *Journal of medical systems,* vol. 41, no. 8, pp. 1-10, 2017.

[39] G. M. Lucas *et al.*, "Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers," *Frontiers in Robotics and AI,* vol. 4, p. 51, 2017.

[40] M. L. Tielman, M. A. Neerincx, M. Van Meggelen, I. Franken, and W.-P. Brinkman, "How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence," *Technology and Health Care,* vol. 25, no. 6, pp. 1081-1096, 2017.

[41] P. Kharel, K. Sharma, S. Dhimal, and S. Sharma, "Early detection of depression and treatment response prediction using machine learning: a review," in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 2019: IEEE, pp. 1-7.

[42] P. Philip *et al.*, "Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders," *Scientific reports,* vol. 7, no. 1, pp. 1-7, 2017.

[43] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial," (in English), *JMIR Ment Health,* Original Paper vol. 4, no. 2, p. e19, 2017, doi: 10.2196/mental.7785.

[44] D. S. W. Janet B.W. Williams, "STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON DEPRESSION RATING SCALE (SIGH-D)," pp. 1-16, 1988.

[45] "Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS)," 2021. [Online]. Available: http://ids-qids.org/.

[46] "Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS)." [Online]. Available: http://goodmedicine.org.uk/files/depression,%20assessment%20qids-sr16,%20background.pdf.

[47] Y. Sharma, "Understanding Count Vectorizer," no. The Startup, 2020. [Online]. Available: https://medium.com/swlh/understanding-count-vectorizer-5dd71530c1b.

[48] "Normalization," no. Techopedia, 2020. [Online]. Available: https://www.techopedia.com/definition/1221/normalization.

[49] S. learn, "Classification in Machine Learning: What it is & Classification Models," 2021. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/classification-in-machine-learning.

[50] J. Brownlee, "Logistic Regression for Machine Learning," April 1 2016. [Online]. Available: https://machinelearningmastery.com/logistic-regression-for-machine-learning/#:~:text=Logistic%20Regression%20for%20Machine%20Learning.%20Logistic%20regression%20is,discover%20the%20logistic%20regression%20algorithm%20for%20machine%20learning.

[51] G. Gasso, "Logistic regression," ed: INSA Rouen-ASI Departement Laboratory, 2019.

[52] Q. Wang *et al.*, "Overview of logistic regression model analysis and application," *Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine],* vol. 53, no. 9, pp. 955-960, 2019.

[53] F. C. Pampel, *Logistic regression: A primer*. Sage publications, 2020.

[54] R. Couronné, P. Probst, and A.-L. Boulesteix, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC bioinformatics,* vol. 19, no. 1, pp. 1-14, 2018.

[55] M. Singh, M. W. Bhatt, H. S. Bedi, and U. Mishra, "Performance of Bernoulli's naive bayes classifier in the detection of fake news," *Materials Today: Proceedings,* 2020.

[56] W. contributors, "Naive Bayes classifier, Wikipedia, The Free Encyclopedia," 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Naive_Bayes_classifier&oldid=1053686606.

[57] A. Askari, A. d'Aspremont, and L. El Ghaoui, "Naive feature selection: Sparsity in naive bayes," in *International Conference on Artificial Intelligence and Statistics*, 2020: PMLR, pp. 1813-1822.

[58] V. Tama and Y. Sibaroni, "Labeling analysis in the classification of product review sentiments by using multinomial Naive Bayes algorithm," in *Journal of Physics: Conference Series*, 2019, vol. 1192, no. 1: IOP Publishing, p. 012036.

[59] M. Abbas, K. A. Memon, A. A. Jamali, S. Memon, and A. Ahmed, "Multinomial Naive Bayes classification model for sentiment analysis," *IJCSNS Int. J. Comput. Sci. Netw. Secur,* vol. 19, no. 3, pp. 62-67, 2019.

[60] M. Artur, "Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features," *Procedia Computer Science,* vol. 190, pp. 564-570, 2021.
[61] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning*: Elsevier, 2020, pp. 101-121.
[62] J. Shaw, "What is the difference between SVC and SVM in scikit-learn?," 2019. [Online]. Available: https://intellipaat.com/community/19362/what-is-the-difference-between-svc-and-svm-in-scikit-learn#:~:text=They%20are%20just%20different%20implementations%20of%20the%20same,liblinear%20and%20only%20supports%20a%20linear%20kernel.%20So%3A.
[63] R. Gholami and N. Fakhari, "Support vector machine: principles, parameters, and applications," in *Handbook of Neural Computation*: Elsevier, 2017, pp. 515-535.
[64] X. Peng and D. Chen, "PTSVRs: Regression models via projection twin support vector machine," *Information Sciences,* vol. 435, pp. 1-14, 2018.
[65] S. Diab, "Optimizing stochastic gradient descent in text classification based on fine-tuning hyper-parameters approach. a case study on automatic classification of global terrorist attacks," *arXiv preprint arXiv:1902.06542,* 2019.
[66] s.-l. developers, "Stochastic Gradient Descent," 2021. [Online]. Available: https://scikit-learn.org/stable/modules/sgd.html.
[67] A. Yousaf *et al.*, "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)," *IEEE Access,* vol. 9, pp. 6286-6295, 2020.