

**VICTORIA UNIVERSITY**  
MELBOURNE AUSTRALIA

*Identifying self-disclosed anxiety on Twitter: A natural language processing approach*

This is the Published version of the following publication

Zarate, Daniel, Ball, Michelle, Prokofieva, Maria, Kostakos, Vassilis and Stavropoulos, Vasileios (2023) Identifying self-disclosed anxiety on Twitter: A natural language processing approach. *Psychiatry Research*, 330. ISSN 0165-1781

The publisher's official version can be found at  
<https://www.sciencedirect.com/science/article/pii/S0165178123005292>  
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/47474/>



# Identifying self-disclosed anxiety on Twitter: A natural language processing approach

Daniel Zarate<sup>a,\*</sup>, Michelle Ball<sup>b</sup>, Maria Prokofieva<sup>b</sup>, Vassilis Kostakos<sup>c</sup>,  
Vasileios Stavropoulos<sup>a,d</sup>

<sup>a</sup> College of Health and Biomedicine, Royal Melbourne Institute of Technology (RMIT), Australia

<sup>b</sup> Institute for Health and Sport, Victoria University, Melbourne, Australia

<sup>c</sup> University of Melbourne, Melbourne, Australia

<sup>d</sup> Department of Psychology, University of Athens, Athens, Greece

## ARTICLE INFO

### Keywords:

Anxiety  
Cyber-phenotype  
Digital footprint  
Natural language processing  
Sentiment analysis  
Twitter

## ABSTRACT

**Background:** Text analyses of social media posts are a promising source of mental health information. This study used natural language processing to explore distinct language patterns on Twitter related to self-reported anxiety diagnosis.

**Methods:** A total of 233,000 tweets made by 605 users (300 reporting anxiety diagnosis and 305 not) over six months were comparatively analysed, considering user behavior, Linguistic Inquiry Word Count (LIWC), and sentiment analysis. Twitter users with a self-disclosed diagnosis of anxiety were classified as 'anxious' to facilitate group comparisons.

**Results:** Supervised machine learning models showed a high prediction accuracy (Naïve Bayes 81.1 %, Random Forests 79.8 %, and LASSO-regression 79.4 %) in identifying Twitter users' self-disclosed diagnosis of anxiety. Additionally, a Latent Profile Analysis (LPA) identified four profiles characterized by high sentiment (31 % anxious participants), low sentiment (68 % anxious), self-immersed (80 % anxious), and normative behavior (38 % anxious).

**Conclusion:** The digital footprint of self-disclosed anxiety on Twitter posts presented a high frequency of words conveying either negative sentiment, a low frequency of positive sentiment, a reduced frequency of posting, and lengthier texts. These distinct patterns enabled highly accurate prediction of anxiety diagnosis. On this basis, appropriately resourced, awareness raising, online mental health campaigns are advocated.

## 1. Social media as a source of mental health information

The highly popular use of social media allows people to express their opinions, feelings, and thoughts to others via their instantly accessible personal devices (e.g., smartphones, computers, etc.; Zarate et al., 2022). Previous scholars suggest that the language chosen in one's social media posts has been featured as a valuable source of mental health information (Insel, 2018). Notably, the term 'cyberphenotype' has been introduced to describe how online behavior, including social media posts, may indirectly operate as a mental-health footprint (Zarate et al., 2022). Such diagnostic potential is maximized by the constant (*moment-to-moment*) and naturalistic (*in-situ*) flow of information that can be passively captured (i.e., without the user's conscious involvement; Zarate et al., 2022). These qualities can be seen as an important

adjunct offered by traditional clinical practice methods, projecting the decoding of mental health information in social-media participation as an imperative priority (Cuthbert and Insel, 2013; Insel, 2018).

Indeed, social networking sites (SNS) provide researchers with an ideal source of information, procuring access to individuals' uncensored voices and narratives (Ngai et al., 2015). This could be particularly important for assessing a person's experienced mental health, as the language adopted by a user to express themselves online may include spontaneous and uncensored expressions (Insel, 2018). Such qualities are reinforced by a sense of pseudo-anonymity, encouraging rich and meaningful information that may be otherwise elusive (Insel, 2018). Furthermore, over time observations of how a user conducts themselves on SNS (i.e., time of day, frequency, and content of posts/ interactions with others over lengthier periods of time; > 6 months) often portray

\* Corresponding author at: Institute for Health and Sport, Victoria University, PO Box 14428, Melbourne 8001, Australia.

E-mail address: [Daniel.zarate@live.vu.edu.au](mailto:Daniel.zarate@live.vu.edu.au) (D. Zarate).

<https://doi.org/10.1016/j.psychres.2023.115579>

Received 1 December 2022; Received in revised form 13 September 2023; Accepted 27 October 2023

Available online 3 November 2023

0165-1781/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

their relationships-narratives as well as their identity-narratives (i.e., how they experience their engagement with others, and how they view or engage with their self; Denzin, 2004; McAdams, 2010; Stavropoulos et al., 2020). In other words, self-expressions in SNS could either consciously (purposefully) or unconsciously (latently) encapsulate content that aligns with deeper and conscious self-appraisals, likely associated with the user's mental health experience (Denzin, 2004).

Previous studies have aimed to assess linguistic expressions in social networking sites to detect a range of psychological disorders, commonly including depression, suicide and schizophrenia, and to a lesser extent, eating disorders and anxiety (Chancellor and De Choudhury, 2020; Coppersmith et al., 2014, 2015). For example, researchers have identified significant associations between discernible language patterns (including intonation, word rate, fluency, grammatical form, and lexical selection) and mood (Larsen et al., 2020), symptoms of depression (Gkotsis et al., 2017) and psychosocial stressors (Mowery et al., 2017). However, while a vast proportion of such studies focused on assessing depression or suicide, only a minority examined exclusively symptoms of reported anxiety, suggesting that further research is imperative (Dutta et al., 2018; Ireland and Iserman, 2018; Saifullah et al., 2021; Shen and Rudzicz, 2017). To address these recommendations and to expand the available knowledge, the current project will focus on identifying reported anxiety using Twitter posts while adopting novel and advanced analyses and methodology.

## 2. Identifying reported anxiety online

Anxiety presentations involve worry and apprehension about one or more different conditions or stimuli, often marked by bodily symptoms of physical tension (e.g., accelerated breath and heartbeat; Tascher-eau-Dumouchel et al., 2022). Previous literature suggests that anxiety symptoms and their accompanying behaviors are usually experienced on a continuum (i.e., all individuals are anxious at varying rates, with a minority experiencing extremely high anxiety; Dutta et al., 2018). While non-problematic anxiety constitutes a realistic and healthy reaction to a perceived threat, the experience of excessive or disproportionate anxiety, to the extent that one's wellbeing is compromised, is regarded as the common denominator across several debilitating mental health conditions. These conditions may include generalized anxiety (i.e., symptoms occur across a variety of life domains), social anxiety (i.e., symptoms relate to how others could perceive the person), specific phobias (i.e., symptoms evolve around a specific object/condition), or panic attacks (i.e., symptoms are accompanied by episodes of elevated/acute fear and physical discomfort entailing palpitations, sweating etc.; American Psychiatric Association, 2013).

Elevated anxiety has been evidenced to interfere with cognition and behaviors (i.e., risk evaluation thoughts; and risk avoidance actions), as well as the language used, particularly when online (Settanni and Marengo, 2015; Sonnenschein et al., 2018). For example, more anxious individuals have been shown to frequently use linguistic expressions with higher negative affect, lower positive affect, increased self-criticism, lower self-efficacy expectations, experiential avoidance and tensed utterance (Berman et al., 2010; Joiner and Blalock, 1995; Rook et al., 2022; Settanni and Marengo, 2015; Smith and Jones, 2013; Sonnenschein et al., 2018; Woodgate et al., 2020).

Studies exploring digital traces of mental health disorders have used social media posts to identify reported anxiety (Zarate et al., 2022). Such studies followed a sequence of steps/stages involving (a) accessing, (b) analyzing and (c) predicting anxiety-related information (Chancellor and De Choudhury, 2020). Firstly, the creation of large user/content databases frequently requires the use of application programming interfaces (APIs) to retrieve and organize information in meaningful ways (i.e., relevant concepts or words and users/profiles associated with reported anxiety are accessed and collated; Zarate et al., 2022). Secondly, the dataset(s) are examined via natural language processing (NLP) techniques, which aim to quantify meaningful signals and patterns of

experienced/reported anxiety within a given a compilation of texts or corpus (Chowdhary, 2020). For example, NLP techniques may quantify lexical and semantic forms of experienced/reported anxiety in text via the inclusion of *n*-gram analysis (i.e., assessing successions of words, symbols or prefixes-tokens), Linguistic Inquiry Word Count (LIWC; e.g., the proportion of words falling under different linguistic, psychological and topical categories), and sentiment analysis (i.e., identification of emotional tone in a specific text), among others (S.C. Guntuku et al., 2017). To maximize the prediction power of such text analysis findings, recent studies have additionally employed machine learning algorithms (i.e., analysis methods that "learn" to progressively improve their accuracy via leveraging data accumulation and testing; Singh et al., 2016; Zarate et al., 2022). These algorithms enrich text-analysis prediction models, making them more accurate (Singh et al., 2016).

Not surprisingly, previous studies analysing social media posts to detect experienced anxiety have demonstrated promising accuracy and invited further research in the area (Coppersmith et al., 2014, 2015; Ive et al., 2018; Ireland and Iserman, 2018; Shen and Rudzicz, 2017). For example, Ireland and Iserman (2018) used a decision tree algorithm to detect anxiety-related posts on Reddit with a 92 % accuracy. Similarly, Ive et al. (2018) used hierarchical neural models to detect anxiety-related posts with 82 % accuracy. Finally, Coppersmith et al. (2015) used Twitter posts and character *n*-gram language models to identify anxiety from nine other distinct mental health presentations.

Despite such important steps, to the best of the authors' knowledge, no studies have aimed to describe the cyber-phenotype of experienced/reported anxiety via concurrently examining sentiment analysis, user behavior (i.e., frequency, text length, and time of posting), and LIWC evidence. Combining such methods could inform anxiety classification and *profiling* models (i.e., models where anxiety-associated social media content typologies could be simultaneously portrayed based on all these features, such that anxious and non-anxious individuals are correctly classified). Indeed, several studies have successfully used latent profile analyses informed by psychometric indicators (i.e., questionnaire scores) to distinguish different types of digital media users, such as gamers (Billieux et al., 2015; Kovacs et al., 2022). Aside from distinctly connecting different social media text and usage patterns with reported/experienced anxiety (i.e., variable-focused research), the advantage of following such an approach would enable the portrayal of anxious profiles of social media users more holistically and accurately (i.e., person-focused research; Stavropoulos et al., 2021).

## 3. The present study

To expand the available knowledge, the present study innovatively co-examined posts of Twitter users reporting self-disclosed anxiety diagnosis over six months, aiming to: (i) decode patterns of linguistic Twitter expressions by examining a combination of user behavior, LIWC and sentiment analysis associated with reporting of a self-disclosed anxiety diagnosis; and (ii) to collectively consider such patterns as latent profile analysis indicators to accurately describe different Twitter posting typologies and their links to self-disclosed anxiety. These methodologically innovative aims were enhanced by the comparative employment of several machine learning algorithms to maximize prediction power. Accordingly, the findings are expected to have significant practical contributions. Firstly, from a clinical perspective, the early and cost-efficient identification of people who suffer from anxiety represents the potential to optimize treatment outcomes through their timely engagement. Secondly, knowledge of one's anxious cyber-phenotype could help tailor minimally invasive, resource-saving online interventions to address anxiety in the broader community. Thus, the following hypotheses were proposed:

**Hypothesis 1.** Linguistic expressions on Twitter can accurately predict self-disclosed Anxious/non-anxious status.

**Hypothesis 2.** There will be distinguishable latent profiles to describe the sample considering their linguistic expression on Twitter, with profiles characterized by low sentiment significantly related to one's reported anxiety diagnosis.

## 4. Method

### 4.1. Data collection

Data collection commenced after obtaining approval from the Victoria University Research Ethics Committee. Following the method employed by Coppersmith et al. (2014, 2015), two authors applied for a scientific Twitter developer account to access the maximum extraction of 10,000,000 tweets per month. The Twitter developer credentials, in conjunction with the rtweet package for RStudio (Kearney, 2019), informed the Twitter API, which identified participants having stated the phrase "I have been diagnosed with Anxiety". The most recent six months of tweets for 300 users who posted such phrase were collated to inform our Anxiety group. The content of this phrase was carefully chosen to be considered reflective of a medical/health procedure and not just any subclinical experience of anxiety. Similarly, a control group of Twitter users who did not post such phrase on their timeline was sourced ( $n = 305$ ). This study did not collect demographic information, as per ethics approval received, to (a) preserve participant anonymity and (b) assess whether online user behavior/content could accurately predict a publicly claimed anxiety diagnosis. Importantly, the self-disclosure of anxiety diagnosis on Twitter does not guarantee such a diagnosis, nor can this method identify confounding effects/variables (e.g., anxious users not stating it online and non-anxious users providing false statements).

All data used in this study was publicly posted on Twitter between May and November 2021 and made available through Twitter's API. The collections of tweets used here include 233,000 tweets in English and do not include direct messages, retweets, or data marked as private by the author. The dataset is available on the following repository <https://doi.org/10.17026/dans-zfd-pu7r>.

### 4.2. Data processing

Data processing involved a series of steps to quantify user behavior and linguistic expression on Twitter, including user activity, Linguistic Inquiry Word Count (LIWC), and sentiment analysis. Following the method employed by Silge and Robinson (2022), Tidytext (Silge and Robinson, 2016), tidyverse (Wickham et al., 2019), and lubridate RStudio packages (Grolemund and Wickham, 2011) were used to compare characteristics of linguistic expression on Twitter between the 'anxious' and the 'non-anxious' groups. Average tweet length and time of posting (including time of day and day of the week) across anxious/non-anxious groups were compared. Specifically, tweets were (a) compiled into a 'corpus', (b) *tokenized* (separate tweets into single words), and (c) stop words (stop words are frequent words such as "the", "is", "of", and may not add value to lexical analyses) and symbols (such as URLs and #) were removed. A bag-of-words approach was then used to quantify the frequency of terms using the *term frequency\*inverse document frequency* function (*tf idf*; Silge and Robinson, 2022). The *tf-idf* approach is commonly employed in NLP to normalize term frequency across documents (in this case, tweets) and thus obtain a score of 'term-salience', with higher scores representing higher importance. Finally, *sentiment analysis* was conducted to classify and quantify emotional intent in tweets. Bing (Liu, 2015), nrc (Mohammad, 2021), and affinn (Nielsen, 2011) lexicons were used to classify words into sentiment categories and ascribe emotional valence to tweets. The nrc lexicon classifies 13,875 words into ten sentiments: anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise and trust. The bing lexicon classifies 6786 words into positive/negative sentiment, and the affinn lexicon ascribes emotional value to 2477 words ranging from -3

to +3.

### 4.3. Prediction and classification models

Prediction and classification models were created to identify the presence/absence of cyberphenotypical characteristics of anxiety in Twitter users. Specifically, to answer whether self-disclosed anxiety status can be accurately predicted based on linguistic expressions on Twitter ( $H_1$ ), the tidymodels RStudio package (Kuhn and Wickham, 2020) was used to build predictive models (null, Naïve Bayes, LASSO-regression, and Random Forests). These models used eight indicators (text length, tf-idf, hour of the day, day of the week, week-day/weekend, sentiment bing, sentiment nrc, sentiment affinn) to identify whether users belonged to the self-disclosed 'anxious' or 'non-anxious' group. A ten-fold stratified cross-validation resampling method was used to train predictive models, and assessment of model accuracy involved examination of confusion matrix and area under the curve (AUC) with higher AUC indices representing higher accuracy of class prediction (see Hvitfeldt and Silge, 2022 for an explanation of machine learning models).

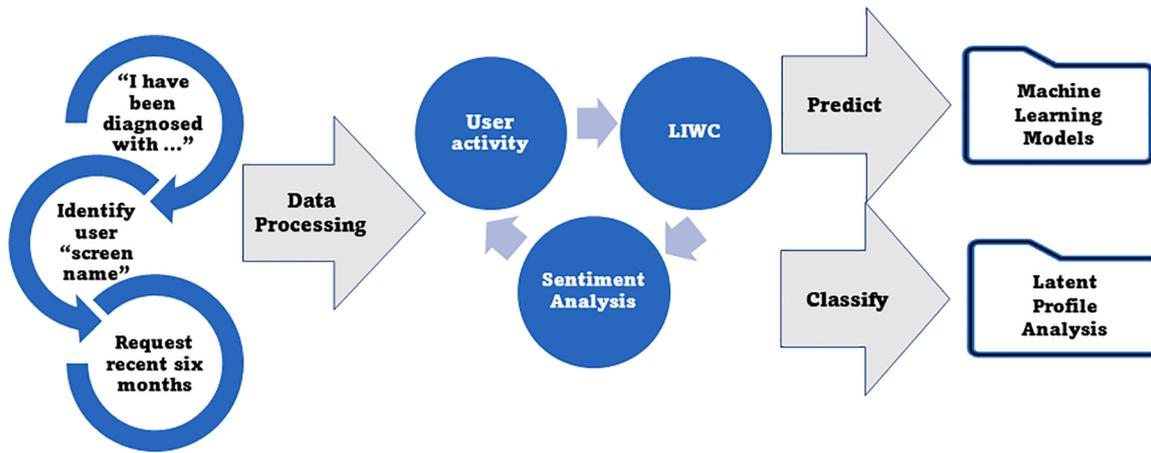
Machine learning models were chosen based on their performance and use in precision psychiatry (Bzdok and Meyer-Lindenberg, 2018). Considering that LASSO-regressions apply penalties to coefficients based on their magnitude, model indicators were standardized before fitting the model (Ahrens et al., 2020). Moreover, Naïve Bayes has been proposed as a low-variance classifier due to its limited complex structure (i.e., posterior probability distribution based on the class prior conditional probability; Webb, 2011) and it assumes independence of variables. Thus, before fitting the Naïve Bayes, highly correlated variables were excluded (i.e., sent value was excluded due to high correlation with sent bing [ $r = 0.975$ ] and sent nrc [ $r = 0.705$ ]) and a low variance filter (*low variance* <0.97) was applied (using the creditmodel package for RStudio [Fan, 2022]). Additionally, using the VIP package (Greenwell et al., 2022), a permutation-based variance importance test was conducted to identify important contributors to the models' accuracy. Subsequently, the recpselector package (Pawley, 2022) was used to implement a recursive feature selection and thus evaluate the loss in model accuracy by recursively removing variables with reduced importance.

Finally, to test  $H_2$ , the tidyLPA RStudio package (Rosenberg et al., 2019) was used to identify latent profiles (Latent Profile Analysis, LPA). The same model indicators used to inform the predictive models were used to inform a classification/profiling model. A maximum likelihood estimator (MLE) was used to estimate parameterization combinations, while information/classification criteria (Akaike information criterion [AIC], Bayesian information criterion [BIC] and standardized entropy [ $h$ ]) were used to determine the optimum number of latent profiles (see Kovacs et al., [2022] and Masyn [2013] for an explanation of LPA). Subsequently, a  $\chi^2$  test of independence was used to evaluate how (if at all) latent profiles relate to self-reported diagnoses of anxiety on Twitter. Fig. 1 presents the data flow process from data acquisition to prediction and classification models.

## 5. Results

### 5.1. Twitter user behavior and sentiment analysis

A series of statistical tests were conducted in RStudio (2020) to determine potential differences in user behavior across the self-reported anxious/non-anxious groups. Self-reported 'Anxious' users tweeted less frequently ( $M_{\text{anxious}}=321.55$ ,  $M_{\text{non-anxious}}=447.33$ ;  $\chi^2_{[1]}=6465$ ,  $p<.001$ , Cohen's  $W = 0.17$ ) and posted longer tweets than the non-anxious ( $Kruskal-Wallis \chi^2_{[1]}=47,050$ ,  $p<.001$ ;  $M_{\text{anxious}}=92.9$  /  $M_{\text{non-anxious}}=51$ ,  $\epsilon^2=0.20$ ). Additionally, self-reported 'anxious' users tweeted more frequently between 4pm and 11pm ( $F_{[1,44]}=18.67$ ,  $p<.001$ ,  $\eta^2=0.30$ ; 47 % of tweets between these times), compared to the 'non-anxious' self-reported group. There was a significant difference in



**Fig. 1.** Here we see the data flow process including data retrieval, data processing, and prediction/classification. The first step involved accessing a user database through the Twitter API. The second step involved data processing to obtain model indicators representing user activity (frequency, text length, and time of posts), sentiment analysis, and Linguistic Inquiry Word Count (LIWC). Finally, model indicators were used to predict (using Naïve Bayes, Random Forests, and LASSO regression) and classify (using Latent Profile Analysis) self-reported ‘anxious’ users.

tweeting frequency per day across both groups, with the ‘non-anxious’ self-reported users tweeting more frequently on Sundays and Tuesdays ( $\chi^2_{[6]}=693, p<.001$ ; Cohen’s  $W = 0.01$ ). See Supplementary Table 1 for a description of tweets grouped by self-reported ‘anxious’ and ‘non-anxious’ self-reported users, including tweet frequency, length, time of the day, day of the week, and text valence.

Sentiment analysis revealed significant differences across groups. Specifically, the Bing lexicon determined that self-reported ‘anxious’ users posted negative words more frequently (56 % negative, 44 % positive) than the ‘non-anxious’ self-reported users (33 % negative, 67 % positive;  $\chi^2_{[1]}=12,107, p<.001$ ; Cohen’s  $W = 0.24$ ). Similarly, the nrc lexicon determined that self-reported ‘anxious’ users tweeted terms expressing negative emotions more frequently, whereas the ‘non-anxious’ self-reported users employed positive emotions more frequently. For example, the self-reported ‘anxious’ group used anger-related words (e.g., abandon, abhorrent) 9.5 % of the time compared to 5.8 % in the ‘non-anxious’ self-reported users, and fear-related words (e.g., absence, badness) 10.4 % of the time compared to 5.3 % in the ‘non-anxious’ self-reported users (see Supplementary Table 2 for a comprehensive list of the most frequently used sentiment-related words by group). Finally, the sentiment distribution (i.e., emotional valence ascribed to tweets using the affinn lexicon) was significantly different across groups (*Kolmogorov-Smirnov*  $D = 0.26, p<.001$ ) with self-reported ‘anxious’ users showing lower sentiment value ( $M_{\text{sentiment}}=-0.30$ ) compared to the ‘non-anxious’ self-reported users ( $M_{\text{sentiment}}=1.00$ ). Supplementary Figure 1 illustrates Twitter user behavior and sentiment analysis discriminated by self-reported ‘anxious’ and ‘non-anxious’ self-reported users.

**5.2. Classification and profiling models**

Considering *Hypothesis 1*, supervised machine learning models were fitted to predict the classification of Twitter users into the self-reported ‘anxious’ and ‘non-anxious’ self-reported users. Specifically, eight model indicators (day of the week, weekday/weekend, hour of the day, Bing lexicon, nrc-lexicon, affinn-lexicon, text length, and tf-idf) were used to fit four machine learning models (null, LASSO-regression, naïve Bayes, and Random Forests). The null model was used to assess the randomness of classification without model indicators, and as expected, it correctly ‘guessed’ every second observation (51 % accuracy, 0.5 AUC), showing appropriate randomness of selection. As hypothesized, prediction models showed acceptable accuracy, with the naïve Bayes model producing the highest classification accuracy (81.1 %, 0.86 AUC), followed by the Random Forests (79.8 %, 0.90 AUC) and the LASSO-regression

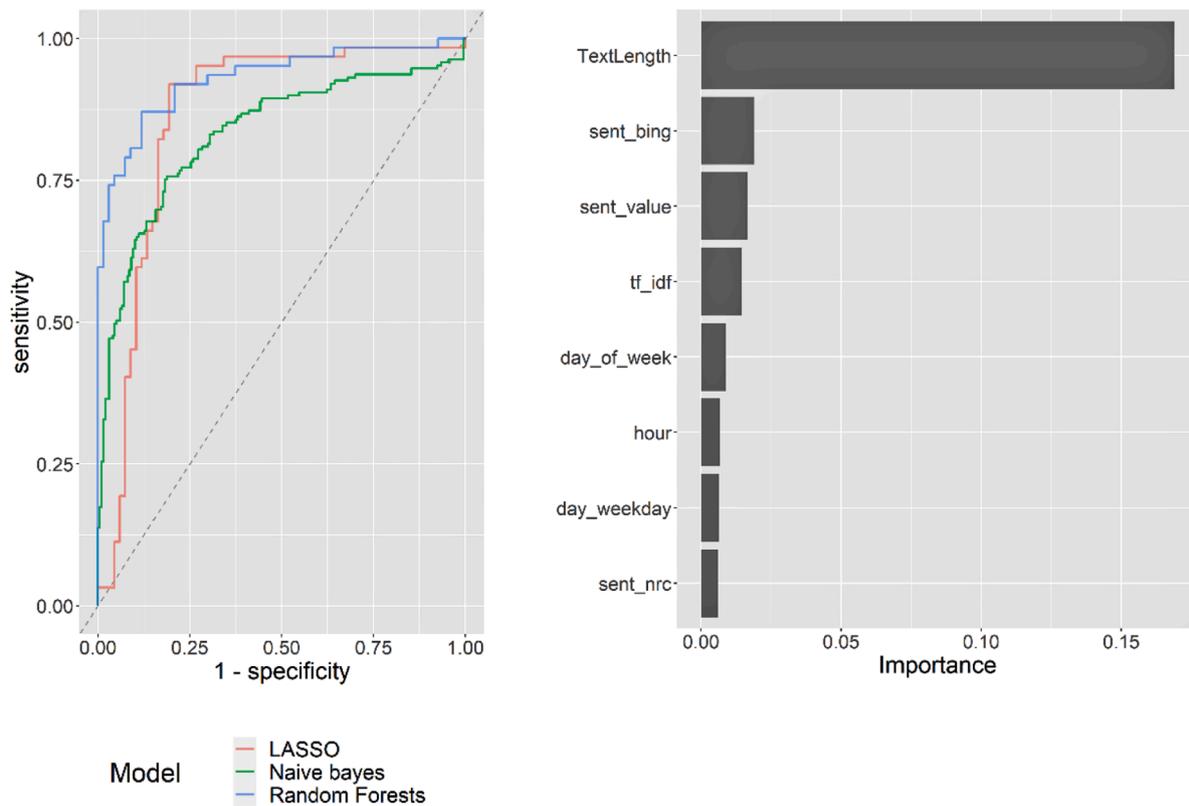
(79.4 %, AUC 0.90; Table 1 and Fig. 2 left panel). A permutation-based importance test identified text length, positive/negative words (Bing lexicon) and sentiment value as the most important predictors for Random Forests. Similarly, the most important predictors for LASSO-regression were tf-idf, text length and positive/negative words (Bing lexicon; Fig. 2, middle and right panel). A recursive feature elimination method retaining variables in the top 50th percentile in importance indicated that text length alone increased the Random Forests prediction accuracy (82.9 %, AUC 0.86). Tf-idf and text length were sufficient to maintain accuracy in the Naïve Bayes (80.6 %, AUC 0.84) and LASSO-regression models (79.8 %, AUC 0.89). Interestingly, removing tf-idf and keeping text length as the only predictor in the Naïve Bayes and LASSO-regression models did not converge on a solution.

Considering *Hypothesis 2*, a Latent Profile Analysis (LPA) was used to identify the optimum number of user profiles considering their linguistic expression on Twitter. Of the possible variance-covariance parameter combination, only the class-invariant diagonal parameterization (CIDP) model with equal variances and covariances fixed to zero converged on a solution. Specifically, this model assumes equal variability in model indicators for all latent profiles (equal variance) and no relationships across different profiles (covariance fixed to zero; for an explanation of possible variance-covariance parameterization models see Kovacs et al., 2022). As seen in Fig. 3 (left panel), increasing the number of latent profiles resulted in decreased model errors (AIC and BIC). However, the CIDP model, including four latent profiles, showed appropriate AIC, BIC, and the highest profile heterogeneity (*standardized entropy*  $h = 0.93$ ) and was therefore selected as the optimum fit (Supplementary Table 3

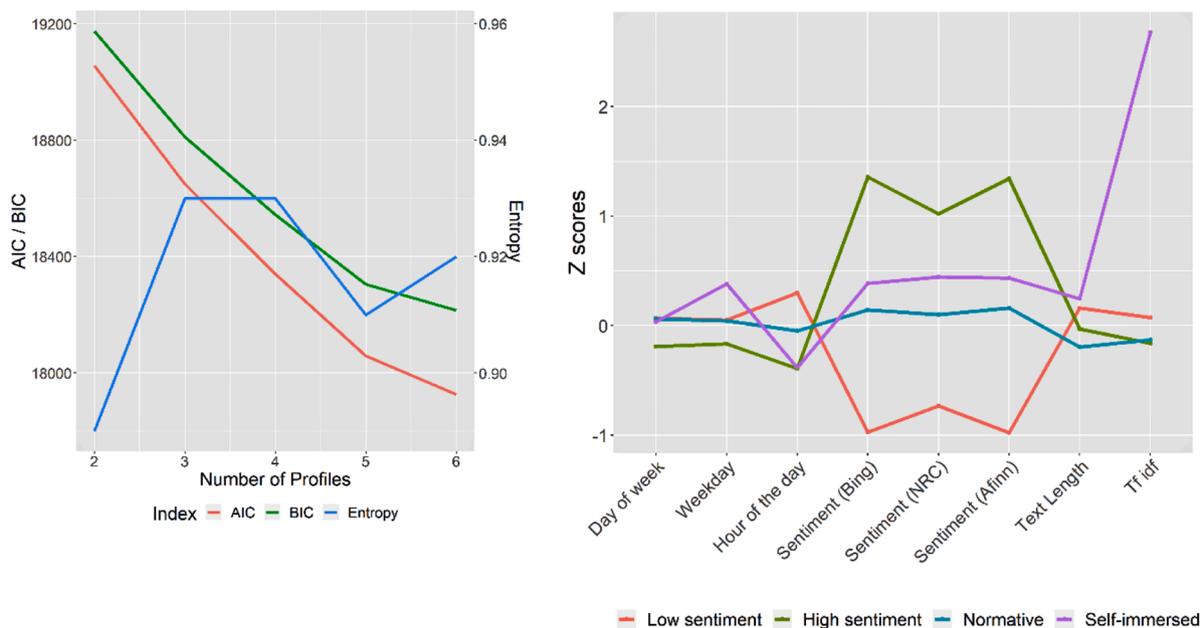
**Table 1**  
Machine learning models.

Model	Tuning	Accuracy	AUC	TP	TN	FP	FN
Null	N.A.	0.510	0.500	33	34	32	33
LASSO-regression	$\lambda = 0.014$	0.643	0.862	21	62	5	41
Naïve Bayes	N.A.	0.777	0.838	13.8	16.2	3.5	5.1
Random Forests	Mtry = 2, Min_n = 25	0.845	0.930	56	53	6	14

Note: Accuracy = number of correct classification divided by total observations. AUC = The area under the curve ranges from 0 to 1 with higher values indicating higher model performance. TP = True positives. TN = True negatives. FP = False positives. FN = False negatives.  $\lambda$  = Penalty term applied to LASSO regressions. Mtry = Number of random variables for recursive partitioning employed at each split (i.e., decision) to minimize node impurity; Min n = Minimum number of data points in a node required for the node to be split further.



**Fig. 2.** The left panel presents the area under the curve (AUC) for three machine learning classification models. As seen here, the Random Forest classification model provides the best performance. **The right panel** illustrates a permutation-based importance test. This examines the level of reliance of Random Forest classification on different model features. Permutation indicates the ability to randomly permute (or shuffling) the order of variables included in a model. Importance represents the difference in baseline model accuracy (before shuffling variables) and performance on the permuted dataset. Higher importance values indicate higher reliance on that variable to maintain model accuracy.



**Fig. 3.** Here the left plot helps visualize the identification of optimum number of latent profiles. The left vertical margin shows AIC and BIC values, the right vertical margin shows entropy values, and the horizontal axis shows different number of latent profiles. As seen here, higher number of profiles show lower AIC and BIC; however, entropy is optimized at 4 latent profiles. **The right plot** shows standardized mean scores (vertical axis) of each model indicator (horizontal axis) discriminated by 4 different latent profiles. The low and high sentiment profiles are characterized by sentiment values 1SD below and above the mean, respectively. The self-distancing profile is characterized by a different term frequency.

presents AIC, BIC, entropy, N-min and bootstrapped likelihood ratio test for model 1-CIDP with two to six latent profiles). Accordingly, the share of Twitter users in each latent profile was 40.8 % in Profile 1 ( $n = 210$ ), 25.2 % in Profile 2 ( $n = 130$ ), 32 % in Profile 3 ( $n = 165$ ), and 2 % in Profile 4 ( $n = 10$ ). Table 2 displays standardized mean scores discriminated by Twitter user latent profile.

Twitter user latent profiles were described considering mean standardized user behavior, the sentiment expressed in tweets, and the frequency of salient terms. As hypothesized, the four latent profiles showed different characteristics (Fig. 3-right panel). Specifically, Profile 1 was characterized by low sentiment value ( $-1$ SD sentiment *bing* and *afinn*, and  $-0.73$ SD sentiment *nrc*). Profile 2 showed high sentiment value ( $+1.35$ SD sentiment *bing* and *afinn*, and  $+1$ SD sentiment *nrc*). Profile 3 was defined by average model indicator values. Finally, Profile 4 was characterized by  $+2.7$ SD *tf-idf*. In this profile, individuals used the words ‘you’ ( $n = 583$ , 7 %) and ‘everyone’ ( $n = 0$ , 0 %) less frequently compared to the rest of the sample ( $n = 303,855$ , 16 %;  $n = 28,565$ , 1.5 %), thus named as the “self-distancing” Twitter users. Supplementary Figure 2 presents a comparison of the five most frequently used words between profile 4 and the rest of the sample. As hypothesized, latent profiles denoted by anxious traits (i.e., low sentiment and self-immersed language) were significantly more populated by Twitter users with a self-disclosed diagnosis of anxiety ( $\chi^2_{[3]}=52, p<.001$ ; Cohen’s  $W = 0.01$ ). Specifically, there were 139 self-reported ‘anxious’ users (66.2 %) in the low sentiment profile, 41 (31.5 %) in the high sentiment, 63 (38.2 %) in the normative, and 8 (80 %) in the self-distancing profile.

## 6. Discussion

This study sought to decode the cyber-phenotype of self-reported anxiety using linguistic expressions in social media and to identify different latent profiles of users disclosing such symptoms. To address these aims, a natural language processing approach was used to identify user behavior and patterns of linguistic content using a corpus of available tweets purposely accessed via the Twitter API. Overall, Twitter users with a self-disclosed diagnosis of anxiety tweeted less frequently, posted longer tweets, and used language conveying negative sentiments more frequently than those without a self-disclosed diagnosis of anxiety. Additionally, ML showed acceptable accuracy in predicting users’ group membership (‘anxious’/‘non-anxious’; Random Forests = 84 %, Naïve Bayes = 81 %, and LASSO-regression = 79 %). Finally, four distinct profiles of Twitter users were identified, describing users who expressed high sentiment (41 % of users), low sentiment (25 %), normative (32 %), and self-distancing (2 %) language, with a significantly higher proportion of ‘anxious’ users in the low sentiment and self-immersed profiles. Taken together, these results represent important implications for cost-efficiently and accurate identification of self-reported anxiety indications expressed on social media platforms.

### 6.1. Identifying self-reported anxiety in Twitter posts

Considering the first hypothesis addressed in this study, significant differences in linguistic expression and user behavior between Twitter users with a self-disclosed diagnosis of anxiety and those without such

self-disclosed diagnosis were evident. Specifically, self-disclosed ‘anxious’ users tweeted less frequently and posted longer tweets than non-self-disclosed “anxious” users. Dutta et al. (2018) observed a similar reduction in social interactions through online platforms between ‘anxious’ users and their strong online connections, indicating a fear of negative evaluations. Moreover, Berman et al. (2010) suggest that exaggerated beliefs about being evaluated negatively represent cognitive distortions (from a cognitive behavioral therapy [CBT] framework perspective) and may lead to experiential avoidance (from an acceptance and commitment therapy [ACT] perspective). In this context, ‘anxious’ Twitter users may show reduced social interaction due to learned maladaptive internal responses (e.g., inferiority, self-criticism, lack of self-compassion) that maintain and reproduce unhelpful patterns of behavior, giving rise to anxiety and fear of scrutiny, thus restricting their interactions with others via posts on Twitter (Wright et al., 2017). However, the authors acknowledge that self-disclosure of anxiety diagnosis via social media platforms, such as Twitter, could not validate the veracity or existence of such diagnosis, suggesting these observations to be interpreted with caution and inviting further research.

Interestingly, the results observed here indicate that while Twitter users with self-disclosed diagnoses of anxiety tweeted less frequently, they posted lengthier texts than the ‘non-anxious’ self-reporting group. This highlights the possibility of a dichotomous cognitive process in which anxious individuals either avoid posting on social media for fear of negative evaluations or may reversely post lengthier texts due to being overly concerned with minimizing errors and perfectionism (Gregersen and Horwitz, 2002). Ong and Twohig (2022) proposed that when worried, some people think about future communication mistakes excessively, aiming to prevent them via overly elaborated and lengthier messages. This is reinforced by literature suggesting that anxiety-induced cognitive biases may generate a “black and white” perspective of the world, eventuating perfectionistic engagement with their surroundings (Wright et al., 2017).

Another important difference between Twitter users with a self-disclosed diagnosis of anxiety and the control group resided in the sentiment valence embedded in their tweets. Interestingly, the words most frequently used were similar for both groups, with *good*, *love*, and *happy* being the most frequent positive words and *bad* and *hate* being the most frequent negative words. However, terms reflecting negative affect were more commonly posted by ‘anxious’ users across all three lexicons employed here (see Supplementary Tables 1 and 2 and Supplementary Figure 1). In line with Woodgate et al. (2020), this observation suggests that anxiety-affected individuals may frequently communicate their worry, lack of confidence, negative self-image, and emotional dysregulation.

### 6.2. Prediction and classification of self-disclosed anxiety in Twitter users

The above-discussed differences regarding the cyber-phenotypical characteristics between Twitter users with a self-disclosed diagnosis of anxiety and ‘non-anxious’ self-reporting users enabled accurate group membership classifications. Specifically, all supervised machine learning models showed good classification accuracy of self-reported

**Table 2**

Description of latent profiles including standardized scores of model indicators.

Profile	Text Length	Day of the week	Weekday / Weekend	Hour	Sent Bing	Sent NRC	Sent AFINN	tf-idf
Low sentiment (40.8 %)	0.16	0.07	0.05	0.29	-0.97	-0.73	-0.98	0.07
High sentiment (25.2 %)	-0.03	-0.19	-0.18	-0.39	1.36	1.02	1.34	-0.16
Normative (32 %)	-0.19	0.06	0.04	-0.05	0.14	0.10	0.16	-0.13
Self-distancing (2 %)	0.25	0.04	0.38	-0.38	0.39	0.44	0.43	2.68

Note. This table shows standardized (z scores) values for each model indicator discriminated by latent profiles. The ‘Low sentiment’ and ‘High sentiment’ profiles are characterized by  $\sim 1$ SD below/above the mean in word sentiment respectively. The ‘Self-distancing’ class is characterized by high *tf-idf* denoting a different word use distribution (i.e., the type of words used).

'anxious' users (Naïve Bayes 80.5 %, Random Forests 80 %, and LASSO-regression 84.6 %) and 'non-anxious' self-reported users (Naïve Bayes 81.5 %, Random Forests 79.7 %, and LASSO-regression 76.6 %). The most reliable variables for correct group membership classification were text length and tf-idf for the Naïve Bayes and LASSO-regression models and text length and sentiment analyses for the Random Forests. This indicates that these variables should be considered for the prediction/assessment of anxiety via linguistic expressions on social media platforms. Overall, and irrespective of the classification algorithm employed, the results suggest that the chosen model indicators provide sufficient information to detect and predict self-reported anxiety accurately. However, considering the limited available empirical evidence supporting this interpretation and the over-emphasis these models placed on text length to accurately predict self-reported 'anxious' Twitter users, further studies may aim to expand on this area.

Moreover, a latent profile analysis based on the indicators assessed (i.e., user behavior, LIWC and sentiment analysis) suggested that four distinct profiles denoted salient latent cyber-phenotypical characteristics of Twitter users in this sample. Specifically, 40.8 % of users in this sample showed low sentiment valence in their linguistic expressions via Twitter ( $-1SD$ ; Fig. 3 right panel). Most users in the low sentiment profile (66 %) disclosed a diagnosis of anxiety on Twitter, suggesting that 'anxious' users are more likely to use language that conveys negative affect (including anger, fear, disgust, and sadness; [Mohammad, 2021](#); [Woodgate et al., 2020](#)). Additionally, 32 % and 25 % of users in this sample showed normative (mean values) and high sentiment valence ( $+1SD$ ) in their linguistic expressions, respectively. Interestingly, 31 % of users who disclosed a diagnosis of anxiety on Twitter also showed high sentiment valence in their posts. This suggests that the cyber-phenotype of anxiety may be also co-informed by other elements beyond the low valence of texts (e.g., text length, frequency and time of posting, etc.), and a combination of these elements should be incorporated in models predicting anxiety based on an individual's social media activity.

Interestingly, 2 % of Twitter users posted the word *you* and *everyone* significantly less frequently and were thus categorized as the self-distancing group. Previous research suggested that self-focused attention, often due to experienced anxiety/distress, may result in self-immersed practices (i.e., diminished social interaction) and is associated with negative affect ([Mor and Winquist, 2002](#)). However, individuals showing self-distancing language did not exhibit above-average negative emotional valence or increased use of the first-person pronoun, suggesting an interesting combination of elements. Thus, the relatively low negative affect reported by those classified as belonging in the fourth profile could reflect their level of acceptance and embracing of their anxiety to the extent that they felt comfortable enough to announce it online. Considering the above sample limitations, further research may seek to explore this interpretation.

### 6.3. Conclusions, implications, and limitations

These results significantly contribute to understanding self-reported anxiety through one's linguistic expression on social media platforms, such as Twitter. Specifically, and in keeping with past literature, findings highlight that individuals reporting to suffer from anxiety may use language showing negative affect and reduced positive affect, often entailing statements related to a lack of confidence, fear, and worry ([Woodgate et al., 2020](#)). Furthermore, findings suggested a reduced frequency of posts from those disclosing an anxiety diagnosis. Interestingly, individuals enduring anxiety symptoms may likely engage in experiential avoidance and fear of negative evaluations, increasing their self-immersed practices, such as minimal posting on Twitter, while reducing their motivation for social interactions ([Berman et al., 2010](#)). This is particularly important considering the positive impact either offline (including social activities, hobbies, outdoor activities, sports, etc.) and online (e.g., social media communities) social support

networks have on mental health ([Li et al., 2021](#)). Moreover, the significantly lengthier posts of those disclosing an anxiety diagnosis could indicate their excessive concerns regarding minimizing errors and presenting as perfectly as possible ([Ong and Twhig, 2022](#)). Indeed, results suggested that the machine learning prediction models informed by the above differences show promise regarding the opportunity to automate reliable anxiety assessment based on an individual's Twitter activity.

These findings could be bearing significant epidemiological, assessment, prevention, and intervention implications. Firstly, from an epidemiological perspective, using highly naturalistic methods such as cyber-phenotyping would facilitate a more accurate estimation of anxiety prevalence and incidence rates, if concurrently cross-validated with formal diagnostic procedures. Specifically, considering that individuals suffering from anxiety might feel averse to voicing in person their psychological ailments due to stigma and lack of awareness, current statistics might not accurately represent the prevalence rates of such disorders. Secondly, from a clinical assessment perspective, this approach could represent an efficient and cost-effective strategy to help individuals suffering from mental health issues. For example, deploying social media mental health campaigns dedicated to predicting and detecting anxiety may facilitate accessing relevant information and resources to develop understanding and awareness, promoting action to address such presentations. Finally, from an intervention perspective, the knowledge of one's digital phenotype based on Twitter use could help tailor personalized applications to their recipients' profiles, maximizing the effectiveness of interventions. For example, one's cyber-phenotype profile may contribute to efficiently guiding the required intervention strategy (i.e., what works for whom approach). Taken together, this methodology has the potential to provide the basis for devising pioneering services designed to help individuals at risk of suffering from anxiety and potentially other comorbid psychological issues, if cross-validating with clinical interviews and/or reliable psychometric assessments.

Considering that this is the first study assessing the possibility of detecting self-reported anxiety-related symptoms via publicly available online content, the feasibility of deploying a service to detect and prevent the cyber-phenotype of anxiety is currently limited yet promising. Specifically, larger scales studies are needed to validate these preliminary analyses and thus confirm the ability to accurately detect an anxiety-cyber-phenotype by analyzing one's online behavior and content in conjunction with other formal diagnostic procedures. Nonetheless, a larger empirical base supporting the findings presented here will enable the deployment of strategic interventions (such as a dedicated API) to detect and prevent anxiety-related symptoms.

Despite these important contributions, the results reported here need to be interpreted in the context of several limitations. Firstly, considering this study used self-disclosed diagnoses of anxiety, the allocation of group membership did not follow rigorous clinical assessments, and thus results should be interpreted with caution. For instance, Munchausen syndrome/malingering cases via the Internet, where one falsely states to have been diagnosed with anxiety on Twitter to receive attention/sympathy comments from the online community cannot be excluded ([Feldman, 2000](#)). However, Munchausen syndrome prevalence rates have been reported to vary around 1 % ([Šileikytė and Viliūnienė, 2020](#)). Therefore, it is rather unlikely that our anxiety group does not represent individuals who have actually been diagnosed with anxiety. It is also possible that participants included in our anxious group may bear particular characteristics that differentiate them from the broader anxious population. For example, for one to publicly state "I have been diagnosed with anxiety" on Twitter, their aim to receive help and/or a distinct level of self-awareness/insight regarding their symptoms may be assumed. Secondly, the 'non-anxious' group randomly comprised Twitter users who did not post the phrase "I have been diagnosed with anxiety", and therefore the mental health status of this group is uncertain. Thirdly, the current sample represents a small proportion of Twitter users, and thus results presented here may need to be validated with

larger samples derived from other social media sources. This is important considering that text length was an important feature for prediction accuracy of machine learning models. Thus, further studies are required to validate these results. Fourthly, further studies may, alongside addressing the above limitations, also consider how (if at all) sentiment valence in social media posts changes at different milestones from receiving a psychiatric diagnosis.

Further consideration should be given to limitations regarding the lack of demographics, including age, gender, or educational status, as these would be useful to identify more specific differences in Twitter posting patterns of anxious users. However, their absence does not affect the validity of the current analyses, as they were not used as predictors. Only posting patterns such as sentiment, length of posts and posting time patterns were considered here, aiming to identify common expressions of anxiety irrespective of demographic differences. In addition, the cross-validation employed via bootstrapping procedures for our prediction neutralizes potential skewness or polarized results due to sample demographics (Berrar, 2018). Moreover, the sequence of the robust machine learning procedures employed to detect differences between the two groups increases the clarity of our results regarding common Twitter posting patterns, which are distinctive of anxious users (whilst acknowledging potential variability in line with broadly applied criteria; APA, 2013; Kuhn and Wickham, 2020). Similarly, considering that linguistic patterns and speech measures vary widely between different languages, cultures and contexts (Parola et al., 2022), the findings reported here need to be validated using diverse samples, including speakers of different languages and diverse demographic characteristics.

## Funding

Dr Vasileios Stavropoulos received funding by:

- The Victoria University, Early Career Researcher Fund ECR 2020, number 68,761,601.
- The Australian Research Council, Discovery Early Career Researcher Award, 2021, number DE210101107

## Authors' contribution

DZ contributed to the article's conceptualization, data curation, formal analysis, methodology, project administration, and writing of the original draft. VS contributed to the article's conceptualization, data curation, writing of the original draft, formal analysis, methodology, and writing of the original draft. MB, MP and VK contributed to writing, reviewing, and editing the final draft.

## Ethical standards – animal rights

All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

## Informed consent

Following guidelines outlined in the 2018 National Statement on the Ethical Conduct in Human Research, a waiver of consent was obtained to access publicly available social media content on Twitter.

## Confirmation statement

The authors confirm that this paper has not been either previously published or submitted simultaneously for publication elsewhere.

## Copyright

Authors assign copyright or license the publication rights in the present article.

## Supplementary materials

Additional material available at [https://github.com/Daniel28052/NLP\\_Anxiety](https://github.com/Daniel28052/NLP_Anxiety).

## Declaration of Competing Interest

The authors of the present study do not report any conflict of interest.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.psychres.2023.115579.

## References

- Ahrens, A., Hansen, C.B., Schaffer, M.E., 2020. Lassopack: model selection and prediction with regularized regression in Stata. *Stata J.* 20 (1) <https://doi.org/10.1177/1536867X20909697>.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Author. <https://doi.org/10.1176/appi.books.9780890425596>.
- Berman, N.C., Wheaton, M.G., McGrath, P., Abramowitz, J.S., 2010. Predicting anxiety: the role of experiential avoidance and anxiety sensitivity. *J. Anxiety Disord.* 24, 109–113. <https://doi.org/10.1016/j.janxdis.2009.09.005>.
- Berrar, D., 2018. Cross-validation. *Encyclopedia Bioinformatic. Comput. Biol.* 1, 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>.
- Billieux, J., Thorens, G., Khazaal, Y., Zullino, D., Achab, S., Van del Linden, M., 2015. Problematic involvement in online gamers: a cluster analytic approach. *Comput. Human Behav.* 43, 242–250. <https://doi.org/10.1016/j.chb.2014.10.055>.
- Bzdok, D., Meyer-Lindenberg, A., 2018. Machine learning for precision psychiatry: opportunities and challenges. *Biol. Psychiatr.: Cognit. Neurosci. Neuroimag.* 3, 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>.
- Chancellor, S., De Choudhury, M., 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Med.* 3 (43) <https://doi.org/10.1038/s41746-020-0233-7>.
- Chowdhary, K.R., 2020. *Fundamentals of Artificial Intelligence*. Springer. <https://www.springer.com/gp/book/9788132239703>.
- Coppersmith, G., Dredze, M., Harman, C., 2014. Quantifying mental health signals in Twitter. *Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signals to Clinical Reality*, pp. 51–60. <https://aclanthology.org/W14-3207.pdf>.
- Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., 2015. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. In: *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signals to Clinical Reality*, pp. 1–10. <https://aclanthology.org/W15-1201/>.
- Cuthbert, B.N., Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* 11 (126), 1–8. <https://doi.org/10.1186/1741-7015-11-126>.
- Denzin, N., 2004. Symbolic interactionism. In: Praneer, L. (Ed.), *Handbook of Research Methods in Health Social Sciences*. Springer, Singapore, pp. 151–172.
- Dutta, S., Ma, J., De Choudhury, M., 2018. Measuring the impact of anxiety on online social interactions. In: *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, pp. 584–587. <https://ojs.aaai.org/index.php/ICWSM/article/view/15081>.
- Fan, D. (2022). *Creditmodel: toolkit for credit modeling, analysis and visualization* [RStudio package]. <https://cran.r-project.org/web/packages/creditmodel/index.html>.
- Feldman, M.D. Munchausen by Internet: detecting factitious illness and crisis on the Internet. *South. Med. J.*, 93(7), 669–672. <https://doi.org/10.1097/00007611-200093070-00006>.
- Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T.J.P., Dobson, R.J.B., Dutta, R., 2017. Characterization of mental health conditions in social media using informed deep learning. *Sci. Rep.* 7 (45141) <https://doi.org/10.1038/srep45141>.
- Greenwell, B., Boehmke, B., & Gray, B. (2022). *Package 'VIP'* [RStudio package]. <http://cran.r-project.org/web/packages/vip/vip.pdf>.
- Gregersen, T., Horwitz, E.K., 2002. Language learning and perfectionism: anxious and non-anxious language learners' reactions to their own oral performance. *Modern Language J.* 86 (4), 562–570. <https://doi.org/10.1111/1540-4781.00161>.
- Grolemund, G., Wickham, H., 2011. Dates and times made easy with lubridate. *J. Stat. Softw.* 40 (3), 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C., 2017. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci* 18, 43–49. <https://doi.org/10.1016/j.cobeha.2017.07.005>.
- Hvitfeldt, E., Silge, J., 2022. *Supervised Machine Learning for Text Analysis in R*. CRC Press. <https://smltar.com/>.

- Insel, T.R., 2018. Digital phenotyping: a global tool for psychiatry. *World Psychiatry* 17 (3), 276. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6127813/>.
- Ireland, M.E., Iserman, M., 2018. Within and between-person differences in language used across anxiety support and neutral reddit communities. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 182–193. <https://aclanthology.org/W18-0620.pdf>.
- Ive, J., Gkotsis, G., Dutta, R., Stewart, R., Velupillai, S., 2018. Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 69–77. <https://aclanthology.org/W18-0607/>.
- Joiner Jr., T.E., Blalock, J.A., 1995. Gender differences in depression: the role of anxiety and generalized negative affect. *Sex Roles* 33 (1–2), 91–108. <https://doi.org/10.1007/BF01547937>.
- Kearney, M.W., 2019. Rtweet: collecting and analyzing Twitter data. *J. Open Source Software* 4 (42), 1829. <https://doi.org/10.21105/joss.01829>.
- Kovacs, J., Zarate, D., de Sena Collier, G., Tran, T.T.D., Stavropoulos, V., 2022. Disordered gaming: the role of a gamer's distress profile. *Can. J. Behav. Sci.* <https://doi.org/10.1037/cbs0000335>. Advance online publication.
- Kuhn, M., & Wickham, H. (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles* [RStudio package]. <https://www.tidymodels.org>.
- Larsen, M.E., Boonstra, T.W., Batterham, P.J., O'Dea, B., Paris, C., Christensen, H., 2020. We feel: mapping emotion on Twitter. *IEEE J. Biomed. Health Inf.* 19 (4), 1246–1252. <https://doi.org/10.1109/JBHI.2015.2403839>.
- Li, F., Luo, S., Mu, W., Li, Y., Ye, L., Zheng, X., Chen, X., 2021. Effects of sources of social support and resilience on the mental health of different age groups during the COVID-19 pandemic. *BMC Psychiatry* 21 (16), 1–14. <https://doi.org/10.1186/s12888-020-03012-1>.
- Liu, B., 2015. *Sentiment analysis: Mining sentiments, opinions, and Emotions*, 1st Ed. Cambridge University Press.
- Masyn, K.E., 2013. Latent class analysis and finite mixture modeling: statistical analysis. In: Little, T.D. (Ed.), *The Oxford handbook of Quantitative Methods*. Oxford University Press, pp. 551–611.
- McAdams, D.P., 2010. Narrative identity. In: Schwartz, S.J., Luyckx, K., Vignoles, V.L. (Eds.), *Handbook of Identity Theory and Research*. Springer, pp. 99–116. <https://link.springer.com/content/pdf/10.1007/978-1-4419-7988-9.pdf>.
- Mohammad, S., M., 2021. Sentiment analysis: automatically detecting valence, emotions and other affectual states from text. *Emotion Measur.* 2 <https://doi.org/10.48550/arXiv.2005.11882>.
- Mor, N., Winquist, J., 2002. Self-focused attention and negative affect: a meta-analysis. *Psychol. Bull.* 128 (4), 638–662. <https://doi.org/10.1037/0033-2909.128.4.638>.
- Mowery, D., Smith, H., Cheney, T., Stoddard, H., Coppersmith, G., Bryan, C., Conway, M., 2017. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. *J. Med. Internet Res.* 19 (2), e48. <https://doi.org/10.2196/jmir.6895>.
- Ngai, E.W.T., Tao, S.S.C., Moon, K.K.L., 2015. Social media research: theories, constructs and conceptual frameworks. *Int. J. Inf. Manage.* 35 (1), 33–44. <https://doi.org/10.1016/j.ijinfomgt.2014.09.004>.
- Nielsen, F.A., 2011. A new ANEW: evaluation of a word list of sentiment analysis in microblogs. In: Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages, pp. 93–98. <http://arxiv.org/abs/1103.2903>.
- Ong, C.W., Twohig, M.P., 2022. *The Anxious Perfectionist: How to Manage Perfectionism-Driven Anxiety Using Acceptance and Commitment Therapy*. New Harbinger Publications.
- Pawley, S. (2022). *Recipeselectors: extra recipes steps for supervised feature selection* [RStudio package]. <https://github.com/stevenpawley/recipeselectors>.
- Parola, A., Lin, J.M., Simonsen, A., Bliksted, V., Zhou, Y., Wang, H., Inoue, L., Koelkebeck, K., Fusaroli, R., 2022. Speech disturbances in schizophrenia: assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophr. Res.* <https://doi.org/10.1016/schres.2022.07.002>.
- Rook, L., Mazza, M.C., Lefter, J., Brazier, F., 2022. Toward linguistic recognition of generalized anxiety disorder. *Front. Digital Health* 4, 779039. <https://doi.org/10.3389/fgdth.2022.779039>.
- Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C., Schmidt, J., 2019. tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *J. Open Source Software* 3 (30), 978. <https://doi.org/10.21105/joss.00978>.
- RStudio Team (2020). *RStudio: integrated Development for R*. RStudio. <http://www.rstudio.com/>.
- Saifullah, S., Fauziah, Y., Aribowo, A.S., 2021. Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. *J. Informatika* 15 (10), 45–55. <https://doi.org/10.26555/jifo.v15i1.a20111>.
- Settanni, M., Marengo, D., 2015. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Front. Psychol.* 6, 1045.
- Shen, J.H., Rudzicz, F., 2017. Detecting anxiety on Reddit. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology, pp. 58–65. <https://aclanthology.org/W17-3107.pdf>.
- Šileikytė, A., Viliūnienė, R., 2020. The prevalence of Munchausen syndrome in Vilnius. *Ment. Health Fam. Med.* 16, 1033–1043. [https://www.researchgate.net/publication/348676833\\_The\\_Prevalence\\_of\\_Munchausen\\_Syndrome\\_in\\_Vilnius](https://www.researchgate.net/publication/348676833_The_Prevalence_of_Munchausen_Syndrome_in_Vilnius).
- Silge, J., Robinson, D., 2016. Tidytext: text mining and analysis using tidy data principles in R. *J. Open Source Software* 1 (3). <https://doi.org/10.21105/joss.00037>.
- Silge, J. & Robinson, D. (2022). *Text Mining with R: a Tidy Approach*. O'Reilly. <https://www.tidytextmining.com/index.html>.
- Singh, A., Thakur, N., Sharma, A., 2016. A review of supervised machine learning algorithms. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1310–1315. <https://ieeexplore.ieee.org/abstract/document/7724478>.
- Smith, A.R., Jones, D.M., 2013. *Handbook of Human Performance Volume 3: State and Trait*. Academic Press.
- Sonnenschein, A.R., Hofmann, S.G., Ziegelmayer, T., Lutz, W., 2018. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cogn. Behav. Ther.* 47 (4), 315–327.
- Stavropoulos, V., Gomez, R., Mueller, A., Yucel, M., & Griffiths, M. (2020). User-avatar bond profiles: how do they associate with disordered gaming?. *Addictive behaviors*, 103, 106245. <https://doi.org/10.1016/j.addbeh.2019.106245>.
- Stavropoulos, V., Motti-Stefanidi, F., Griffiths, M.D., 2021. Risks and opportunities for youth in the digital era: a cyber-developmental approach to mental health. *Eur. Psychol.* <https://doi.org/10.1027/1016-9040/a000451>.
- Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S.G., LeDoux, J.E., 2022. Putting the “mental” back in “mental disorders”: a perspective from research on fear and anxiety. *Mol. Psychiatry* 27 (3), 1322–1330. <https://doi.org/10.1038/s41380-021-01395-5>.
- Webb, G.I., 2011. Naïve Bayes. In: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning*. Springer. [https://doi.org/10.1007/978-0-387-30164-8\\_576](https://doi.org/10.1007/978-0-387-30164-8_576).
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., Francois, R., Grolemund, G., Hayes, A., Hery, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Muller, K., Ooms, J., Robinson, D., Seidel, D.P., Yutani, H., 2019. Welcome to the tidyverse. *J. Open Source Software* 4 (43), 1686. <https://doi.org/10.21105/joss.01686>.
- Woodgate, R.L., Tailor, K., Tennent, P., Wener, P., Altman, G., 2020. The experience of the self in Canadian youth living with anxiety: a qualitative study. *PLoS One* 15 (1), e0228193. <https://doi.org/10.1371/journal.pone.0228193>.
- Wright, J.H., Brown, G.K., Thase, M.E., Ramirez Basco, M., 2017. *Learning Cognitive-Behavior Therapy: An Illustrated Guide*, 2nd Ed. American Psychiatric Association Publishing.
- Zarate, D., Stavropoulos, V., Ball, M., de Sena Collier, G., Jacobson, N.C., 2022. Exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. *BMC Psychiatry* 22 (1), 421. <https://doi.org/10.1186/s12888-022-04013-y>.