ASM Sc. J., 21(2), 2026

https://doi.org/10.32802/asmscj.2026.0002

Detecting Depression in College Students Through Social Media Text Mining: A Predictive Modelling Approach

C.K. Ch'ng* and P.X. Chung

Decision Mathematics and Data Analytics, School of Quantitative Sciences, Universiti Utara Malaysia, Sintok, Kedah, 06010, Malaysia.

Depression is a rising mental health concern among college students, often manifesting through linguistic and behavioural cues on social media platforms such as Facebook. This study aims to analyse social media comments using text mining techniques to detect potential signs of depression. The research applies a structured methodology beginning with text preprocessing, word cloud visualisation, and content analysis to uncover patterns indicative of emotional distress. Hierarchical clustering is then used to group related depressive themes. Several predictive models, including Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and Naive Bayes, are developed and evaluated using performance metrics such as Area Under the Curve (AUC), Classification Accuracy (CA), Precision (Prec), F1-score, Recall, and Matthews Correlation Coefficient (MCC). Based on the evaluation results, logistic regression outperformed the other methods and was selected for deployment to predict future instances of depression from social media text. The findings provide valuable insights into students' digital expressions and support the development of early intervention tools for mental health monitoring.

Keywords: content analysis; depression detection; predictive model; social media; text mining

I. INTRODUCTION

Depression is a persistent mental health disorder that negatively affects emotional well-being, physical health, and daily functioning (World Health Organization, 2017). Beck's cognitive theory (1967) suggests that individuals with depression often have negative thoughts about themselves and their environment, reflected through the frequent use of first-person pronouns and negative language. A U-shaped pattern in depression prevalence has been identified, with higher levels observed in younger and older adults compared to middle-aged individuals (Mirowsky & Ross, 1992). Self-preoccupation and excessive self-criticism have been highlighted as common traits in depressive individuals (Pyszczynski & Greenberg, 1987).

Globally, more than 280 million people live with depression, with women having a significantly higher risk compared to men, and prevalence rising with age (Institute for Health Metrics and Evaluation, 2021; Woody *et al.*, 2017). Despite the availability of effective treatments, depression often goes undiagnosed and untreated, especially in low- and middle-income countries where more than 75 percent of affected individuals do not receive adequate care due to limited funding, lack of trained professionals, and mental health stigma (Evans-Lacko *et al.*, 2018; Friedrich, 2017). Suicide accounts for over 700,000 deaths each year and is the fourth leading cause of death among those aged 15 to 29 (Evans-Lacko *et al.*, 2018). To aid in diagnosis, tools like the Patient Health Questionnaire-9 (PHQ-9) are widely used and well-validated (El-Den *et al.*, 2018; Spitzer & Williams, 2001).

College students represent a particularly vulnerable population due to academic pressure, transitional life challenges, and social stressors. With the widespread use of social media, these platforms have become essential spaces where young adults share personal thoughts and emotional experiences. Facebook reported approximately 1.88 billion

*Corresponding author's e-mail: chee@uum.edu.my

daily active users worldwide during the first quarter of 2021, making it one of the largest social networking platforms for online communication (Meta, 2021). Furthermore, Facebook is one of the most frequently used social media platforms among university-aged populations, with usage rates exceeding 98% in some studies (Hainš *et al.*, 2020). The widespread use of Facebook and the large volume of user-generated textual content make it a valuable platform for analysing linguistic patterns associated with mental health and depression.

Previous research indicates that individuals experiencing depressive symptoms may exhibit distinct online behaviours, such as reduced interaction, mood-congruent posting, social withdrawal, or increased sharing of emotionally charged content (De Choudhury *et al.*, 2013; Kmetty & Bozsonyi, 2022; Musleh *et al.*, 2022). Experiences such as cyberbullying and social comparison further exacerbate these effects and may lead to negative emotional outcomes (Bono *et al.*, 2024; Normala & Azlini, 2018). These digital footprints can serve as meaningful indicators of psychological well-being.

Text mining is a valuable resource for analysing these patterns by extracting linguistic and affective features from large quantities of unstructured online text. Previous researchers have noted the utility of using text mining techniques to identify markers of depression such as self-referential language, and the presence of negative emotions (Musleh *et al.*, 2022). Text mining can also contribute to content analysis, which can provide quantitative and qualitative insights and offer systematic interpretations of user-generated text (Humble & Mozelius, 2022; Kulatunga *et al.*, 2007)

Despite growing interest in digital mental health research, studies that examine real-world social media data from college students and integrate text mining with predictive modelling remain limited. There is a need for research that can identify linguistic and behavioural indicators of depression using computational approaches grounded in psychological theory.

Therefore, this study seeks to analyse Facebook posts and examine college students using both text mining and content analysis techniques and identify linguistic and behavioural markers of depression. The results of this study may provide some information for early detection and ultimately contribute to development of data-informed mental health interventions for young adults.

A. Problem Statement

Depression affects approximately 264 million people globally, with some of the highest prevalence rates observed in countries such as Ukraine (6.3 percent) and the United States (5.9 percent) (Rickwood *et al.*, 2005). Despite advancements in healthcare systems, developed nations still face high rates of depression due to industrialised lifestyles. In Malaysia, depression rates have doubled, increasing from 2.3 percent in 2019 to 4.6 percent in 2023, according to the National Health and Morbidity Survey (Murugesan, 2024). This highlights an urgent need for improved mental health detection and intervention strategies.

Former Health Director of Malaysia, General Tan Sri Dr Noor Hisham Abdullah emphasised that the COVID-19 pandemic has exacerbated mental health issues, with untreated depression potentially leading to suicide (Bernama, 2021). Suicide remains a leading global cause of death, with over 700,000 deaths annually and depression as a major contributor (World Health Organization [WHO], 2021). In Malaysia alone, there were approximately 1,841 suicide cases in 2019, averaging five deaths per day (Lew *et al.*, 2022).

Social media platforms, particularly Facebook, are widely used among young adults and college students (Pempek *et al.*, 2009). Studies suggest that both active and passive engagement can be associated with increased depression levels (Barry *et al.*, 2017; Frison & Eggermont, 2016), although findings remain inconsistent (Banjanin *et al.*, 2015). The tragic case of South Korean K-pop star Sulli brought international attention to the intersection of depression, cyberbullying, and social media (Koreaboo, 2019; Monsters and Critics, 2019). Similarly, Malaysian suicide statistics remain concerning, with the country ranking second among Muslim-majority nations for suicide rates (Lew *et al.*, 2022).

Many college students use social media to express emotional distress, often subtly through indirect language or emojis (Chen *et al.*, 2020; Grabmeier, 2020). Such expressions are easily overlooked, especially among introverted individuals who may not display obvious symptoms (American

Psychiatric Association, 2013; Rickwood *et al*., 2005). Traditional face-to-face detection is further hindered by accessibility issues and stigma (Gulliver *et al*., 2010; World Health Organization [WHO], 2017). The rising demand for mental health services after COVID-19 has placed additional strain on counsellors, highlighting the need for scalable alternatives (Lakhan *et al*., 2020). As students increasingly share emotional content online, social media provides valuable opportunities for early detection (Narkbunnum & Wisaeng, 2022; Tavchioski *et al*., 2023).

However, existing research in this domain is limited, particularly in the use of sophisticated computational methods such as data mining and text mining to examine social media content (Isah *et al*., 2014; Karami *et al*., 2018; Nasralah *et al*., 2020). Moreover, many prior studies rely on structured surveys or focus on general populations, leaving college students on Facebook, a highly active yet vulnerable group, relatively underexplored. Another challenge lies in handling informal online language, including slang, mixed languages, and emojis, which reduces the robustness and generalisability of existing models. In addition, much of the current work emphasises traditional algorithms, while the potential of advanced approaches such as deep learning remains underutilised. Finally, ethical concerns related to privacy, consent, and stigma are often overlooked, raising further challenges for real-world implementation (Zhou & Mohd, 2025).

To address these gaps, this study employs text mining techniques to analyse Facebook posts from college students with the aim of detecting depression-related linguistic and behavioural patterns. The study identifies subtle cues in social media comments that may signal psychological distress, performs content and cluster analysis to uncover thematic groupings associated with depressive tendencies, and develops predictive models capable of estimating the likelihood of depression based on textual features. It further evaluates the performance and effectiveness of these models in detecting depression among college students. Through the integration of linguistic analysis, clustering techniques, and predictive modelling, this study seeks to advance more reliable and context-sensitive strategies for early mental health detection within digital environments.

B. Scope of the Study

This study focuses on the application of text mining techniques to analyse Facebook-based social media comments and posts for identifying depression-related linguistic patterns among college students. Facebook was selected because it remains one of the most widely used social networking platforms among university-aged populations and provides a rich source of user-generated textual data for analysing behavioural and emotional expressions. This population is notably active on social media and vulnerable to mental health challenges. The analysis is restricted to textual data and excludes other content types such as images, videos, or profile metadata.

The research employs Orange Data Mining Software for key analytical tasks, including preprocessing, visualisation, clustering, and model development. Supplementary tools such as Excel and Python are used for data preparation and additional analyses. The process involves constructing word clouds, performing content analysis, and applying clustering techniques to detect emotional themes and language patterns that distinguish depressed from non-depressed users.

The study also includes the development and evaluation of several machine learning models, including Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Naive Bayes. These models are assessed using metrics suitable for imbalanced datasets, such as Area Under the Receiver Operating Characteristic Curve (AUC), Classification Accuracy (CA), F1-score (F1), Precision (Prec), Recall (Sensitivity/Recall), and Matthews Correlation Coefficient (MCC). A final model is validated using an external dataset to test its generalisability.

Importantly, the study focuses on publicly available Facebook comment data from students and does not incorporate temporal, visual, or user-profile-based features. As communication styles and user demographics may differ across platforms such as X (formerly Twitter), Instagram, and TikTok, the findings of this study are intended to represent Facebook-based social media data and should not be generalised directly to other social media platforms without further validation. The findings aim to assist mental health professionals and educators in developing early intervention tools based on social media text analysis.

C. Significance of the Study

This research highlights the potential of social media, particularly Facebook, as a tool for detecting signs of depression among college students. Given the central role of social media in students' lives, their online expressions often mirror their emotional and psychological states. Recognising these digital cues presents an opportunity for early and timely identification of mental health issues.

By applying text mining techniques such as sentiment analysis, clustering, and linguistic pattern recognition, this study introduces a non-intrusive and data-driven approach to mental health monitoring. Unlike traditional assessment methods, which rely on self-reporting or clinical interviews, social media analysis allows for real-time and organic insights into users' mental well-being.

The significance of this study lies in its contribution to the development of early intervention strategies by mental health professionals, educational institutions, and policymakers. It also supports efforts to reduce the stigma surrounding mental illness by offering a less invasive method for identifying at-risk individuals. This approach may enhance mental health support frameworks, especially within academic settings.

II. MATERIALS AND METHOD

A. Depression and its Contributing Factors in the Social Media Context

Depression is a growing mental health concern among college students, with studies indicating that nearly one-third experience clinically significant symptoms (Auerbach *et al.*, 2018; Luo *et al.*, 2024). The contributing factors include biological markers such as inflammation and astrocyte dysfunction (Dolotov *et al.*, 2022; Wang *et al.*, 2021), psychosocial stressors such as childhood trauma and poor self-perceived health (Kim & Jang, 2021; Zheng *et al.*, 2022), and environmental pressures such as pandemic-related stress (Zhu *et al.*, 2020). In recent years, social media platforms have emerged as valuable data sources, with research showing that depressed individuals often display distinct behavioural patterns, including nocturnal activity, reduced posting frequency, and increased use of negative emotional language (De Choudhury *et al.*, 2013). Among college students in particular, social media serves as a channel for expressing emotional distress, often subtly, through indirect language and symbolic cues.

B. Machine Learning

Shatte *et al.* (2019) systematically reviewed machine learning applications in mental health and categorised them into detection, prognosis, and treatment support systems. Their work highlights how machine learning technologies can identify early symptoms and predict mental health trajectories through large-scale data analysis. Various algorithms have been employed, including support vector machines for classification (Cortes & Vapnik, 1995), random forests for feature importance analysis (Breiman, 2001), and logistic regression for interpretable predictions (Van den Goorbergh *et al.*, 2022). Descriptive techniques like hierarchical clustering and word clouds have proven effective for identifying thematic patterns in unstructured text data (Anandarajan *et al.*, 2019). Text mining and natural language processing (NLP) techniques have enhanced detection capabilities by extracting linguistic markers of depression, including sentiment patterns, pronoun usage, and topic clusters (Yadollahi *et al.*, 2017). Specific tools like Linguistic Inquiry and Word Count (LIWC) analysis have successfully identified depression in social media posts (Coppersmith *et al.*, 2014), while deep learning approaches have improved classification accuracy (William & Suhartono, 2021).

Despite these advances, significant challenges remain. Model generalisability is often limited across demographic groups (Musleh *et al.*, 2022), and ethical concerns regarding privacy, consent, and data sensitivity persist (William & Suhartono, 2021). These limitations highlight the need for more robust, context-sensitive detection approaches that address both technical and ethical considerations.

The current study addresses key gaps in existing research by applying text mining techniques to Facebook posts, offering a scalable approach to identifying latent depressive markers. While traditional methods have relied heavily on structured surveys and psychometric scales (Luo *et al.*, 2024; Narkbunnum & Wisaeng, 2022), this approach leverages naturalistic behavioural data for more timely and unobtrusive detection. Future research directions should focus on improving model robustness across populations while addressing critical ethical concerns in digital mental health assessment.

III. PROPOSED METHOD

This study adopts a quantitative, predictive, and non-experimental approach to develop a model for detecting depression among college students through social media text analysis. The methodology includes research framework, data collection, data selection and preprocessing, visualisation, model development, and evaluation. Word clouds and clustering are used to identify linguistic patterns associated with depression, providing insights into differences in language use between depressed and non-depressed individuals. These insights inform the development and validation of a predictive model using data mining techniques to identify signs of depression based on student-generated content.

A. Research Framework

This study follows a structured methodology to detect depression patterns in social media text among college students, as shown in Figure 1. It begins with defining the research problem and reviewing relevant literature. Data is sourced from secondary Facebook datasets, followed by text preprocessing to clean and structure the data for analysis.

Word clouds are used to visualise frequent terms, while clustering groups similar content to uncover depression-related patterns. Machine learning models are then developed and evaluated using metrics such as AUC, CA, F1-score, Prec, Recall, and MCC. The best-performing model is selected for potential deployment, supporting early depression detection and mental health interventions. Lastly, the selected model is deployed using external instances to evaluate its effectiveness in real-world scenarios.

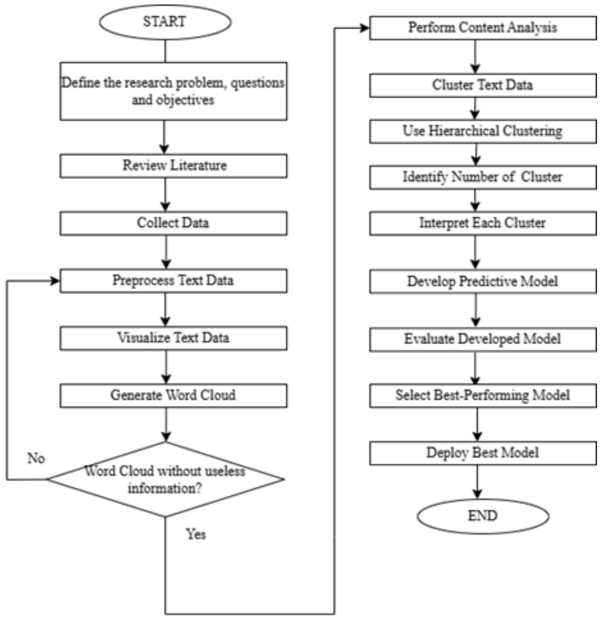

Figure 1. Flowchart of the text mining process

B. Data Collection

This study utilises a publicly available secondary dataset obtained from Kaggle titled Students Anxiety and Depression Dataset (Students' Anxiety and Depression Dataset, 2022). According to the dataset documentation, the dataset comprises approximately 6,982 manually annotated social media comments and posts, including Facebook textual data, with depression labels assigned through a majority-voting process by undergraduate student annotators. This makes the dataset suitable for supervised text classification and depression detection research.

The dataset contains 6,982 textual records. Among these, 6,247 records (89.47%) are labelled as non-depression (Label 0), 733 records (10.50%) are labelled as depression (Label 1), and two records (0.03%) contain missing labels. The two unlabelled records were excluded during the preprocessing stage to ensure data consistency before model development. Figure 2 illustrates the distribution of the labelled instances used in this study.

Since this study aims to develop a binary classification model for depression detection, both the depression (Label 1) and non-depression (Label 0) records were retained for analysis after removing the unlabelled instances.

This dataset was selected due to its relevance, accessibility, and academic suitability. It provides publicly available labelled textual data collected from Facebook-based social media interactions, enabling the application of text mining and machine learning techniques for depression detection. Furthermore, Facebook remains one of the most frequently used social networking platforms among university-aged populations, providing rich user-generated textual content for behavioural and linguistic analysis (Hainš *et al.*, 2020; Frison & Eggermont, 2016).

However, the dataset documentation does not provide detailed demographic information, such as participants' age, gender, ethnicity, institution, or geographical background. Consequently, the representativeness of the dataset cannot be fully verified, and the findings should therefore be interpreted with caution when generalising to the broader college student population.

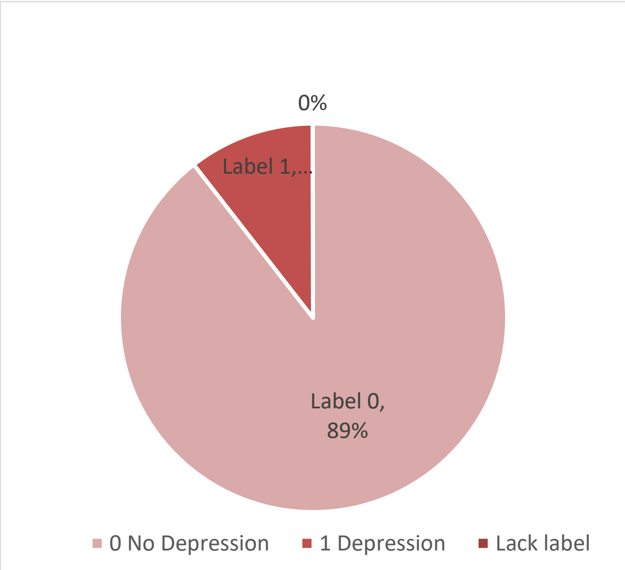

Figure 2. Depression status

C. *Data Selection and Preprocessing*

A multi-stage preprocessing pipeline was applied to ensure high-quality text. The cleaning process involved removing 12 records with missing values, 1,569 non-English entries identified using the langdetect tool, 103 all-uppercase posts that were likely spam, and 175 low-quality texts containing fewer than three meaningful words or with more than 30% punctuation.

This process yielded 5,123 high-quality English text records (73.37% of original data), with all steps implemented in Python using libraries such as 'pandas', 'langdetect', and 're' for text normalisation, URL and emoji removal, and structural filtering. The cleaned dataset was exported for reproducibility and later analysis.

To address class imbalance (677 depressed vs. 4,470 non-depressed records after cleaning), stratified under-sampling was applied using Orange Data Mining. A balanced dataset was created by randomly selecting 500 samples from each class using the "Data Sampler" widget, ensuring fair model training and evaluation.

D. *Data Analysis Strategies*

A systematic analytical framework was used to examine social media text and detect indicators of depression among college students. The analytical process consisted of three major stages: text preprocessing, exploratory text analysis, and predictive modelling. All procedures were conducted using Orange Data Mining Software.

1. *Text preprocessing*

Text preprocessing was performed to prepare raw social media posts for computational analysis. Noise such as punctuation, emojis, repeated characters, and URLs was removed to improve the overall quality of the text, consistent with established natural language processing guidelines (Jurafsky & Martin, 2020). The cleaned text was then tokenised and lemmatised to minimise lexical variation and ensure that different inflected forms of the same word were standardised (Bird *et al.*, 2009). Following these linguistic normalisation steps, the text was converted into numerical form using the Bag-of-Words (BoW) vectorisation model, which enabled the extraction of relevant features for subsequent machine learning procedures (Manning *et al.*, 2008).

2. *Exploratory text analysis*

Exploratory text analysis was conducted to identify linguistic trends, emotional patterns, and thematic structures within the dataset. Word clouds were generated for both depressed posts (Label 1) and non-depressed posts (Label 0), providing an initial visual summary of the most frequently used terms in each category. These visualisations facilitated early interpretation of the emotional and behavioural expressions present in the text.

Content analysis was then performed by examining prominent terms from the word clouds and categorising them according to emotional tone and psychological relevance.

Terms frequently used in depressed posts, such as "restless," were associated with indicators of anxiety, stress, or low mood. In contrast, terms commonly appearing in non-depressed posts reflect themes of motivation, daily routines, or positive emotional states. This thematic categorisation offered deeper insight into students' expression of mental health experiences on social media.

To further explore underlying textual structures, hierarchical clustering was applied to the depressed group. This technique grouped posts with similar linguistic and emotional characteristics into clusters, each representing a distinct psychological theme. The results helped identify subtypes of depressive expression and informed subsequent refinements to the predictive models.

3. *Predictive modelling*

Predictive modelling was conducted to classify social media posts based on the likelihood of depression. Five machine learning algorithms such as Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Naive Bayes were implemented in Orange Data Mining Software. Each model was trained on the vectorised text and assessed for its ability to distinguish between depressed and non-depressed posts. The modelling workflow used in this study is presented in Figure 3. Evaluation of the models provided insight into the most effective techniques for automated depression detection within social media contexts.

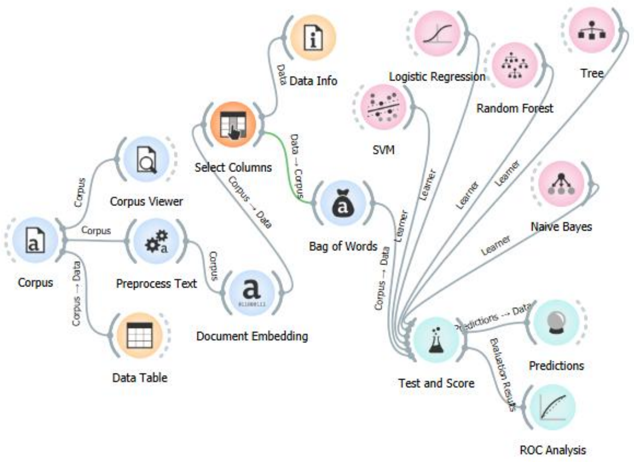

Figure 3. Predictive models' development

The modelling pipeline began with corpus import and preprocessing (tokenisation, stop word removal, lowercasing), followed by feature extraction using both Bag-of-Words (BoW) and Document Embedding to capture lexical and semantic information. Relevant features were selected before feeding the data into the five learning algorithms. Model performance was evaluated using cross validation, and predictions were compared to identify the most effective algorithm for depression detection.

4. *Evaluation metrics*

Model performance was evaluated using several complementary metrics to ensure a robust and reliable assessment. The Area Under the Receiver Operating Characteristic Curve (AUC) was employed to measure each model's ability to discriminate between depressed and non-depressed posts. Classification Accuracy (CA) measured the proportion of correctly classified instances among all observations, providing an overall indication of model performance. The F1-score, which represents the harmonic mean of precision and recall, was particularly important given the imbalance between the two classes. Precision quantified the proportion of correctly predicted positive cases out of all predicted positives, while recall measured the model's sensitivity and its ability to correctly identify depressed cases. The Matthews Correlation Coefficient (MCC) was included as an additional balanced measure because it incorporates true positives, true negatives, false positives, and false negatives, making it especially suitable for imbalanced datasets. Collectively, these metrics provided a comprehensive evaluation beyond simple accuracy, which is crucial in mental health detection where minimising false negatives is essential.

5. *Software and analytical environment*

A combination of software tools ensured an efficient, accurate, and reproducible analytical workflow. Python was used during the preliminary phase for data cleaning, including the handling of missing values and text normalization tasks. Microsoft Excel supported basic statistical analysis, such as calculating the proportions of depressed and non-depressed entries. Orange Data Mining Software served as the primary platform for text mining, visualisation, clustering, and predictive modelling. Its visual interface and support for multiple algorithms facilitated seamless model comparison and deployment of the best-performing classifier. The integration of these tools provided a coherent and methodologically sound framework for the analysis.

IV. RESULT AND DISCUSSION

This chapter presents the study's findings that address the research objective. The analysis was conducted on a dataset of Facebook comments that had been previously labelled as belonging to depressed and non-depressed groups. Rather than distinguishing users based on their depression status, this study focuses on identifying depression-related linguistic and behavioural patterns within the predefined groups. The findings are presented sequentially. First, word cloud and content analyses examine the most frequently occurring terms and their associated themes. Next, clustering analysis identifies potential subgroups within the depressed group. Finally, predictive models developed using Orange software are evaluated for their ability to classify comments according to the predefined labels. Model performance is assessed using relevant evaluation metrics, followed by a discussion of the findings.

A. Word Cloud Analysis

This section presents the visual analysis of frequently used words among two groups: students identified as depressed and non-depressed. Word cloud visualisation was employed to explore dominant emotional expressions and thematic differences in their social media comments.

B. Data Preparation

Before generating the word clouds, the raw textual data underwent several preprocessing steps using the Preprocess Text widget in Orange. These steps included lowercasing all text, tokenisation and removal of punctuation, stopping words filtering and removal of domain-specific meaningless words. For example, "iam," "wa," "whi," "ha," "you," "allah" and so on which were observed to frequently appear without contributing meaningful context.

In order to ensure a fair visual comparison, a Data Sampler widget was used to randomly select 500 records each from the depressed group (label = 1) and non-depressed group (label = 0), as the original dataset was imbalanced.

C. Word Cloud for Depressed Group

As shown in Figure 4, the depressed group exhibits a word cloud dominated by emotionally negative language, with high-frequency words including "anxious", "restless", "nervous", "cry", and "tired". Frequently occurring phrases such as "don't really", "feel like", "can't sleep", and "worried when" reinforce the overall emotional tone of the text. Sleep-related expressions, including "sleep", "night", and "tired", are also prominent, indicating that discussions related to sleep are common within this group. Furthermore, the presence of words such as "regret", "wrong", "doesn't", "tomorrow", and "future" reflects recurring references to self-evaluation and future-oriented concerns, collectively distinguishing the linguistic profile of the depressed group.

Figure 4. Word cloud for depressed group

D. Word Cloud for Non-Depressed Group

As illustrated in Figure 5, the word cloud for the non-depressed group is dominated by more neutral and positive language, with frequently occurring words such as "school", "friends", "holiday", "love", "morning", and "work". Additional terms, including "buy", "want", "make", and "home", reflect common daily activities, personal aspirations, and social interactions. Compared with the depressed group, the non-depressed group exhibits a more positive and socially oriented linguistic profile, suggesting a greater emphasis on everyday experiences and interpersonal engagement.

Figure 5. Word cloud for non-depressed group

E. Content Analysis

After generating the word clouds in Section *C* and *D*, a more in-depth content analysis was conducted to interpret the psychological and emotional implications of the most frequently used words in both the depressed and non-depressed groups. This section also includes an observation on the lexical variance across both groups.

1. Depressed group

Table 1 shows the most frequently used terms in the depressed group, along with their observed frequencies and interpreted meanings.

Table 1. Frequently Used Words in the Depressed Group

Word	**Frequency**	**Interpretation**
Restless	222	Indicates inability to relax, potential sleep issues, or inner agitation.
Worry	131	Suggests anxiety or overthinking, common in depressive thought patterns.
Feel	114	Reflects introspection, emotional awareness, or distress.
Nervous	88	Points to social anxiety or unease.
Like	80	Often appears in comparisons or expressions of dissatisfaction.
Anxious	79	Direct symptoms of anxiety related disorders, overlapping with depression.
Sleep	56	Possibly indicating insomnia or exhaustion.
Can't	38	Reflects perceived helplessness or barriers.
Afraid	28	Shows fear or vulnerability.
Cry	25	Emotional expressions are linked to sadness or frustration.
Myself	12	May signify isolation, loneliness, or self-focus
Depress	7	Direct references to depression, which reinforces the group label.

These results reveal a recurring use of emotionally charged language focused on negative feelings, isolation, and cognitive struggles. The repetition of certain words like "restless" and "worry" with high frequency suggests a tendency toward rumination, a hallmark of depressive thought processes. The word cloud analysis was based on a total of 1,069 unique words, highlighting the dense emotional vocabulary used by the depressed group and providing a strong foundation for content-based pattern recognition.

2. Non-Depressed Group

The non-depressed group displayed a different linguistic pattern, focusing on daily life and social engagement. Table 2 lists the top recurring words in this group.

Table 2. Frequently Used Words in the Non-Depressed Group

Word	**Frequency**	**Interpretation**
Work	13	Refers to productivity or professional life.
Buy	12	Indicates interests or consumption habits.
Love	12	Used in positive emotional expression. For example, hobbies and relationships.

Home	11	Denotes comfort, family, or belonging.
School	10	Reflects daily routine and student responsibilities.
Dream	10	May imply aspirations or hopeful thinking.
Friends	9	Indicates social support and connection.

The language used here is more externally focused, with a balanced and varied tone, reflecting ordinary aspects of life such as work, relationships, and aspirations. Their word cloud consisted of 2,022 unique words, indicating greater lexical diversity and more varied emotional expression.

F. Lexical Diversity and Emotional Repetition in Depressed vs. Non-Depressed Groups

Although both groups contain 500 comments, a clear difference emerges in the word frequency distribution and vocabulary usage. In the depressed group, the most frequent word (restless) appears 222 times, with a total of only 1,069 unique words. This suggests lower lexical diversity and a strong tendency to repeat certain emotional themes and expressions, such as distress, worry, and fatigue. Such repetition reflects ruminative thinking, where individuals fixate on negative thoughts and emotions.

In contrast, the non-depressed group shows a broader language range, with the most frequent word (want) appearing just 58 times, and a total of 2,022 unique words. This indicates greater lexical variety and more balanced, diversified expressions, typically associated with healthier, outward-focused thought patterns.

G. Clustering Analysis of Depressed Group Comments

In this stage, the pre-processed text data from the depressed group was further explored using hierarchical clustering to discover potential themes or groupings in students' expressions. Rather than using a traditional Bag of Words model, this study adopted Multilingual Sentence-Bidirectional Encoder Representations from Transformers (SBERT) inside the Document Embedding widget, a powerful embedding model that captures the semantic meaning of entire sentences across different languages. This method is more suitable for short, informal text data such as social media comments.

After transforming the text into sentence embeddings, we applied a Distance widget using Cosine similarity, followed by Hierarchical Clustering using the Ward linkage method. Ward linkage was chosen because it minimizes the total within-cluster variance, making it useful for identifying compact and meaningful clusters. The clustering results were grouped into five distinct clusters (C1 to C5). Figure 6 shows the hierarchical clustering dendrogram after setting the maximum depth to 20.

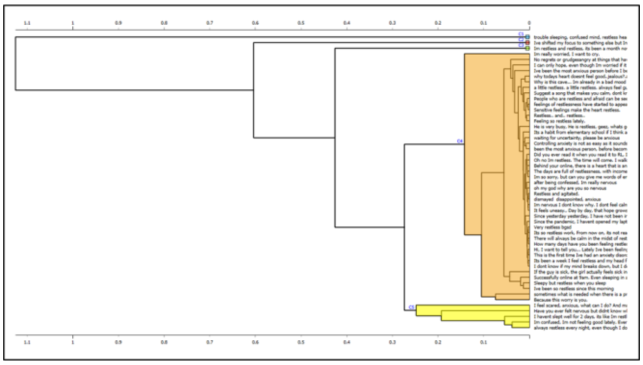

Figure 6. Hierarchical clustering dendrogram

To facilitate the interpretation of the clustering results, representative comments from each cluster are summarized in Table 3. These examples illustrate the common linguistic characteristics and emotional expressions associated with each identified cluster.

Table 3. Representative comments from each hierarchical cluster

Cluster	Theme	Representative comments
C1	Severe Emotional Distress	"Trouble sleeping, confused mind, restless heart. All out of tune."
C2	Mild Anxiety with Focus Shift	"I've shifted my focus to something else but I'm still worried."
C3	Situational Restlessness	"I'm restless and restless, it's been a month now. What do you mean?"
C4	Generalized Anxiety and Restlessness	"I can only hope, even though I'm worried if it starts like this."; "Can't sleep. Just restless by

		nature."; "Sleepy but restless when you sleep."; "Feeling so restless lately."; "Oh God, I'm restless."
C5	Confusion and Emotional Instability	"Have you ever felt nervous but didn't know why?"; "I haven't slept well for 2 days. It's like I'm restless."; "I'm really worried. I want to cry."; "Always restless every night, even though I don't know why."

To facilitate the interpretation of the clustering results, representative comments from each cluster are summarised in Table X. These examples illustrate the common linguistic characteristics and emotional expressions associated with each identified cluster.

Cluster 1 (C1), labelled Severe Emotional Distress, consists of only one comment (0.2%). The representative comment, *"Trouble sleeping, confused mind, restless heart. All out of tune,"* reflects intense emotional suffering accompanied by sleep disturbance and psychological confusion. Although this cluster contains only a single instance, it represents severe emotional distress and a possible mental health crisis. Cluster 2 (C2), also containing one comment (0.2%), was labelled Mild Anxiety with Focus Shift. The representative comment, "*I've shifted my focus to something else but I'm still worried,"* suggests an attempt to cope with anxiety by redirecting attention while persistent worry remains.

Cluster 3 (C3), consisting of one comment (0.2%), represents Situational Restlessness. The representative expression, *"I'm restless and restless, it's been a month now,"* indicates prolonged emotional discomfort associated with specific life circumstances rather than generalised anxiety.

Cluster 4 (C4) is the largest cluster, containing 492 comments (98.4%), and was labelled Generalized Anxiety and Restlessness. Representative comments include *"Can't sleep. Just restless by nature," "Feeling so restless lately," "Sleepy but restless when you sleep," and "I can only hope, even though I'm worried if it starts like this."* These comments consistently express persistent worry, sleep disturbances, emotional instability, uncertainty, and overthinking. The predominance of this cluster suggests that generalised anxiety and chronic restlessness are the most common linguistic patterns observed among the depressed group.

Cluster 5 (C5) consists of five comments (1.0%) and was labelled Confusion and Emotional Instability. Representative comments include "*Have you ever felt nervous but didn't know why?" and "I'm really worried. I want to cry."* These comments indicate emotional uncertainty, unexplained anxiety, and difficulty understanding one's emotional state.

Overall, the emotional distress levels of all participants have been clustered, and their most salient features captured. Comments reflecting generalised anxiety and restlessness (C4), more severe distress (C1), situational restlessness (C3), emotional bewilderment (C5), and mild anxiety with coping strategies (C2) were also noted. This suggests that participants have a range of psychological experiences, ranging from manageable worry to intense emotional crises.

H. Model Building and Evaluation

In this section, Table 4 shows the five classification models which is Logistic Regression, Decision Tree, SVM, Random Forest, and Naive Bayes were evaluated to identify the most effective algorithm for predicting depression among college students based on social media text. Since the dataset used in this research is highly imbalanced with a significantly larger number of non-depressed instances compared to depressed ones. Therefore, special attention was paid to metrics that can provide reliable insight in such scenarios.

The models were assessed using six key evaluation metrics which are AUC, CA, F1-score, Prec, Recall, and MCC. The performance of these models was assessed using the following six evaluation metrics as shown in Table 4.

Table 4. Predictive models' comparison

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.994	0.988	0.988	0.988	0.988	0.948
Tree	0.844	0.961	0.958	0.960	0.961	0.818
Naïve Bayes	0.971	0.935	0.938	0.946	0.935	0.755
Random Forest	0.955	0.933	0.925	0.934	0.933	0.675
SVM	0.973	0.922	0.908	0.925	0.922	0.608

The Logistic Regression model outperformed all other models across all evaluation metrics, indicating its strong ability to distinguish between depressed and non-depressed users based on their text data. It achieved the highest AUC (0.994), CA (0.988), F1-score (0.988), Prec (0.988), Recall (0.988), and MCC (0.948), which suggests it is highly reliable and balanced in detecting both classes correctly.

In contrast, although the Tree and Naive Bayes classifiers performed reasonably well, their overall performance was slightly lower than Logistic Regression. SVM and Random Forest had lower MCC scores, which indicates potential imbalance or misclassification

Overall, it appears that Logistic Regression is the most balanced, precise, and dependable model in terms of accuracy for text-based depression classification. This model could also work as a solid point of reference or benchmark model in other similar mental health detection systems.

1. *Pairwise Model Comparison*

Pairwise statistical comparison further confirmed that Logistic Regression was significantly better than the other models, with p-values as low as 0.000 and 0.001 compared to other algorithms. This statistically validates that the differences in performance are not due to chance.

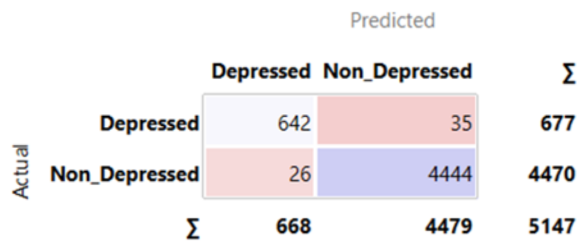

Figure 7. Confusion matrix of Logistic Regression

The confusion matrix of the Logistic Regression model in Figure 7 demonstrated the best performance among all tested models, provides a clear view of its classification capabilities. Out of 677 actual depressed cases, the model correctly identified 642 as depressed (true positives), while misclassifying 35 as non-depressed (false negatives). Among 4,470 actual non-depressed instances, 4,444 were accurately classified as non-depressed (true negatives), with only 26 misclassified as depressed (false positives). This results in a total accuracy of approximately 98.81%, indicating that the model is highly effective in overall classification tasks.

2. *Best Model Deployment*

After identifying Logistic Regression as the best model, it was further deployed using an external dataset comprising 115 social media comments. This new dataset consisted of 113 non-depressed and 2 depressed instances, which was intentionally chosen to simulate real-world class imbalance. The model was shown in Figure 8.

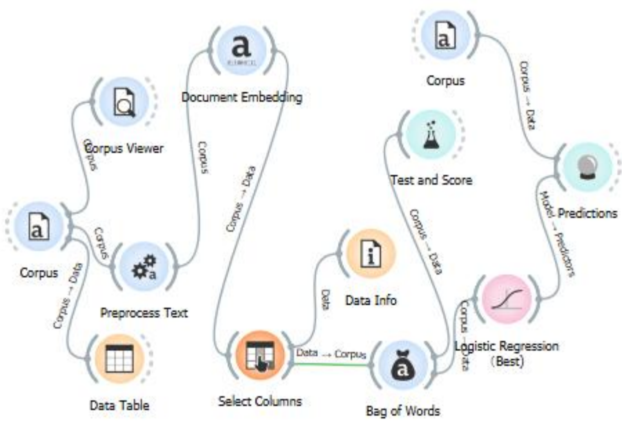

Figure 8. Best model deployment for predicting future instances

The Logistic Regression model demonstrated excellent predictive performance on the external dataset, achieving an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.994, Classification Accuracy (CA) of 0.988, Precision of 0.988, Recall of 0.988, F1-score of 0.988, and a Matthews Correlation Coefficient (MCC) of 0.948. The high AUC value indicates an excellent ability to distinguish between depressed and non-depressed instances, while the consistently high Precision, Recall, and F1-score demonstrate balanced predictive performance across both classes. Furthermore, the high MCC value confirms that the model remained robust despite the highly imbalanced class distribution. These findings suggest that the Logistic Regression model generalises well to previously unseen data and has strong potential for supporting practical depression detection using social media text.

V. CONCLUSION

This study demonstrates how Facebook-based social media text mining can be used to identify signs of depression among college students, who often experience significant emotional and psychological pressures. Through a systematic process that involved preparing the dataset, examining common words, analysing content, grouping similar expressions, and developing predictive models, the research revealed distinct linguistic patterns linked to depressive behaviour. Students who appeared to show signs of depression frequently used repeated and emotionally intense language that expressed worry, restlessness, fatigue, and mental strain. In comparison, students without depressive indicators used broader vocabulary and expressed themselves in a more varied and balanced manner.

The clustering analysis provided an even deeper understanding of these differences. It showed that students communicate their emotional experiences in many ways, ranging from mild uneasiness to more serious distress. Among all the predictive models tested, Logistic Regression achieved the strongest and most consistent performance, demonstrating excellent predictive capability on both the original dataset and the external validation dataset. This confirms its potential as a reliable tool for distinguishing between depressed and non-depressed text in real applications.

The findings also offer valuable contributions for real world use. Universities, counsellors, and mental health departments can consider adapting similar analytical tools to help identify students who may be struggling silently. With proper ethical safeguards, such systems could support early intervention efforts and make it easier for professionals to offer help before difficulties escalate.

Although the study produced promising results, several limitations should be acknowledged. The dataset was not balanced, as it contained far fewer depressed samples, which limited the range of patterns the models could learn. Many posts were short or similar in tone, which reduced the diversity of the clustering results. Furthermore, the dataset was derived primarily from Facebook-based social media posts and comments. Consequently, the identified linguistic patterns and predictive performance may not fully represent communication behaviours on other social media platforms, where user demographics, communication styles, and content formats differ. The study also focused solely on text and did not include other important factors such as posting time, emoji use, or sentiment values, which could provide a more complete picture of online behaviour.

Future research could address these limitations by using larger, more balanced datasets collected from multiple social media platforms to further evaluate the robustness and generalisability of the proposed framework. More advanced methods, including deep learning models such as Long Short-Term Memory (LSTM) or Bidirectional Encoder Representations from Transformers (BERT), may capture richer and more complex language patterns. Additional features such as emoji patterns, posting frequency, or sentiment scoring could further enhance predictive power. Real-time detection systems may also be developed to support continuous monitoring and quicker responses.

In conclusion, this study demonstrates that Facebook-based social media text contains meaningful linguistic signals related to the mental well-being of college students. The integration of descriptive text mining and predictive machine learning techniques provides both interpretable linguistic insights and accurate depression detection. Nevertheless, the findings should be interpreted within the scope of Facebook-based social media data and should not be generalised directly to other social media platforms without further validation. With continued improvement and careful attention to ethical considerations, text mining and predictive modelling can support more timely, accessible, and effective mental health interventions for young adults in an increasingly digital environment.

VI. CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

VII. REFERENCES

American Psychiatric Association 2013, Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Anandarajan, M, Hill, C, Nolan, T, Anandarajan, M, Hill, C & Nolan, T 2019, 'Text preprocessing', Practical text analytics: Maximizing the value of text data, pp. 45-59.

Auerbach, RP, Mortier, P, Bruffaerts, R, Alonso, J, Benjet, C, Cuijpers, P, Demyttenaere, K, Ebert, DD, Green, JG, Hasking, P, Murray, E, Nock, MK, Pinder-Amaker, S, Sampson, NA, Stein, DJ, Vilagut, G, Zaslavsky, AM, Kessler, RC & WHO WMH-ICS Collaborators 2018, 'WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders', Journal of Abnormal Psychology, vol. 127, no. 7, pp. 623–638. https://doi.org/10.1037/abn0000362

Banjanin, N, Banjanin, N, Dimitrijevic, I & Pantic, I 2015, 'Relationship between internet use and depression: Focus on physiological mood oscillations, social networking, and online addictive behavior', Computers in Human Behavior, vol. 43, pp. 308–312. https://doi.org/10.1016/j.chb.2014.11.013

Barry, CT, Sidoti, CL, Briggs, SM, Reiter, SR & Lindsey, RA 2017, 'Adolescent social media use and mental health from adolescent and parent perspectives', Journal of Adolescence, vol. 61, pp. 1–11. https://doi.org/10.1016/j.adolescence.2017.08.005

Bernama 2021, COVID-19 impacts mental health - Dr Noor Hisham. Astro Awani. https://international.astroawani.com/malaysia-news/covid19-impacts-mental-health-dr-noor-hisham-319050

Bird, S, Klein, E & Loper, E 2009, 'Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

Bono, SA, Othman, A, Jamaludin, SSS, Lim, MY & Mohamed Hussin, NA 2024, 'The impacts of social media on adolescent mental health: Malaysian adolescents' perspective', Malaysian Journal of Medicine & Health Sciences, vol. 20.

Breiman, L 2001, 'Random forests', Machine Learning, vol. 45, pp. 5-32.

Chen, LL, Magdy, W, Whalley, H & Wolters, M 2020, It's not just about sad songs: The effect of depression on posting lyrics and quotes. In Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12 (pp. 58-66). Springer International Publishing.

Coppersmith, Glen A, Dredze, Mark & Harman, Craig 2014, Quantifying Mental Health Signals in Twitter, pp. 51-60. http://doi.org/10.3115/v1/W14-3207

Cortes, C & Vapnik, V 1995, 'Support-vector networks', Machine Learning, vol. 20, pp. 273-297.

De Choudhury, M, Counts, S & Horvitz, E 2013, Social media as a measurement tool of depression in populations. In Proceedings of the 5th annual ACM web science conference (pp. 47-56).

Dolotov, OV, Inozemtseva, LS, Myasoedov, NF & Grivennikov, IA 2022, 'Stress-induced depression and Alzheimer's disease: focus on astrocytes', International Journal of Molecular Sciences, vol. 23, no. 9, p. 4999.

El-Den, S, Chen, TF, Gan, Y-L, Wong, E & O'Reilly, CL 2018, 'The psychometric properties of depression screening tools in primary healthcare settings: A systematic review', Journal of Affective Disorders, vol. 225, pp. 503–522. https://doi.org/10.1016/j.jad.2017.08.060

Evans-Lacko, S, Aguilar-Gaxiola, S, Al-Hamzawi, A *et al.* 2018, 'Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: Results from the WHO World Mental Health (WMH) surveys', Psychological Medicine, vol. 48, no. 9, pp. 1560–1571.

Friedrich, MJ 2017, 'Depression is the leading cause of disability around the world', JAMA, vol. 317, no. 15, p. 1517.

Frison, E & Eggermont, S 2016, 'Exploring the relationships between different types of Facebook use, perceived online social support, and adolescents' depressed mood', Social Science Computer Review, vol. 34, no. 2, pp. 153–171. https://doi.org/10.1177/0894439314567449

Grabmeier, J 2020, When college students post about depression on Facebook, Ohio State News, retrieved from <https://news.osu.edu/when-college-students-post-about-depression-on-facebook/>.

Gulliver, A, Griffiths, KM & Christensen, H 2010, 'Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review', BMC Psychiatry, vol. 10, no. 113. https://doi.org/10.1186/1471-244X-10-113

Hainš, VV, Kućar, M & Kovačić, R 2020, Student Social Media Usage and Its Relation to Free-recall Memory Tasks, in 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), pp. 731-736, IEEE.

Humble, N & Mozelius, P 2022, 'Content analysis or thematic analysis: Similarities, differences and applications in qualitative research, in European conference on research methodology for business and management studies, vol. 21, no. 1, pp. 76-81.

Institute for Health Metrics and Evaluation 2021, Global Health Data Exchange (GHDx), retrieved 1 January 2025, <https://vizhub.healthdata.org/gbd-results/>.

Isah, H, Trundle, P & Neagu, D 2014, Social media analysis for product safety using text mining and sentiment analysis, In 2014 14th UK workshop on computational intelligence (UKCI) pp. 1-7. IEEE. https://arxiv.org/abs/1510.05301

Jurafsky, D & Martin, JH 2020, Speech and language processing (3rd ed.). Pearson.

Karami, A, Dahl, AA, Turner-McGrievy, G, Kharrazi, H & Shaw Jr, G 2018, 'Characterizing diabetes, diet, exercise, and obesity comments on Twitter', International Journal of Information Management, vol. 38, no. 1, pp. 1-6. https://doi.org/10.1016/j.ijinfomgt.2017.08.002

Kim, Y & Jang, E 2021, Low self-rated health as a risk factor for depression in South Korea: A survey of young males and females, in Healthcare, vol. 9, no. 4, p. 452, MDPI.

Kmetty, Z & Bozsonyi, K 2022, 'Identifying depression-related behavior on Facebook—An experimental study', Social Sciences, vol. 11, no. 3, Article 135. https://doi.org/10.3390/socsci11030135

Koreaboo 2019, Police Confirms Sulli Has Passed Away by Suicide, Koreaboo, retrieved 14 October 2019, <https://www.koreaboo.com/news/sulli-confirmed-committed-suicide/>.

Kulatunga, U, Amaratunga, D & Haigh, R 2007, Structuring the unstructured data: The use of content analysis.

Lakhan, R, Agrawal, A & Sharma, M 2020, 'Prevalence of depression, anxiety, and stress during COVID-19 pandemic', Journal of Neurosciences in Rural Practice, vol. 11, no. 4, p. 519.

Lew, B, Kõlves, K, Lester, D, Chen, WS, Ibrahim, N, Khamal, NR, Mustapha, F, Chan, CMH, Ibrahim, N, Siau, CS & Chan, LF 2022, 'Looking into recent suicide rates and trends in Malaysia: A comparative analysis', Frontiers in Psychiatry, vol. 12, no. 770252. https://doi.org/10.3389/fpsyt.2021.770252

Luo, Ming-ming, *et al.* 2024, 'Prevalence of depressive tendencies among college students and the influence of attributional styles on depressive tendencies in the post-pandemic era', Frontiers in Public Health, vol. 12, no. 1326582.

Manning, CD, Raghavan, P & Schütze, H 2008, 'Introduction to information retrieval', Cambridge University Press.

Meta 2021, Facebook reports first quarter 2021 results, retrieved on 28 April 2021, <https://investor.atmeta.com/investor-news/press-release-details/2021/Facebook-Reports-First-Quarter-2021-Results/default.aspx>.

Mirowsky, J & Ross, CE 1992, 'Age and depression', Journal of Health and Social Behavior, vol. 33, no. 3, pp. 187–205.

Monsters and Critics, 'Sulli, South Korean pop star, found dead at her home at age 25', CBS News, retrieved on 14 October 2019, <https://www.cbsnews.com/news/sulli-died-former-fx-member-korean-pop-star-found-dead-her-home-cause-of-death-not-released-2019-10-14/>.

Murugesan, M 2024, HEALTH: Depression has doubled in Malaysia, New Straits Times, <https://www.nst.com.my/lifestyle/heal/2024/06/1065085/health-depression-has-doubled-malaysia>.

Musleh, DA, Alkhales, TA, Almakki, RA, Alnajim, SE, Almarshad, SK, Alhasaniah, RS & Almuqhim, AA 2022, 'Twitter Arabic Sentiment Analysis to Detect Depression Using Machine Learning', Computers, Materials & Continua, vol. 71, no. 2.

Narkbunnum, W & Wisaeng, K 2022, 'Prediction of depression for undergraduate students based on imbalanced data by using data mining techniques', Applied System Innovation, vol. 5, no. 6, p. 120.

Nasralah, T, El-Gayar, O & Wang, Y 2020, ‘Social media text mining framework for drug abuse: development and validation study with an opioid crisis case analysis’, Journal of medical Internet research, vol. 22, no. 8, p. e18350. https://doi.org/10.2196/18350

Normala, R, & Azlini, C 2018, ‘The factors of social media use on depression in Malaysia’, International Journal of Research and Innovation in Social Science, vol. 2, no. 12, pp. 352–356.

Pempek, TA, Yermolayeva, YA & Calvert, SL 2009, ‘College students' social networking experiences on Facebook’, Journal of Applied Developmental Psychology, vol. 30, no. 3, pp. 227–238. https://doi.org/10.1016/j.appdev.2008.12.010

Pyszczynski, T & Greenberg, J 1987, ‘Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression’, Psychological Bulletin, vol. 102, no. 1, p. 122. https://doi.org/10.1037/0033-2909.102.1.122

Rickwood, D, Deane, FP, Wilson, CJ & Ciarrochi, J 2005, ‘Young people’s help-seeking for mental health problems’, Australian e-Journal for the Advancement of Mental Health, vol. 4, no. 3, pp. 218–251. https://doi.org/10.5172/jamh.4.3.218

Shatte, Adrian BR, Hutchinson, D & Teague, Samantha J 2019, ‘Machine learning in mental health: a scoping review of methods and applications’, Psychological Medicine, vol. 49, pp. 1426 - 1448. http://doi.org/10.1017/S0033291719000151

Spitzer, RL & Williams, JBW 2001, ‘The PHQ-9’, Journal of General Internal Medicine, vol. 16, no. 9, pp. 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Students’ anxiety and depression dataset, Kaggle, retrieved on 12 November 2022, <https://www.kaggle.com/datasets/sahasourav17/students-anxiety-and-depression-dataset/discussion?sort=undefined>.

Tavchioski, I, Robnik-Šikonja, M & Pollak, S 2023, ‘Detection of depression on social networks using transformers and ensembles’, arXiv preprint arXiv:2305.05325.

Van den Goorbergh, R, van Smeden, M, Timmerman, D & Van Calster, B 2022, ‘The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression’, Journal of the American Medical Informatics Association, vol. 29, no. 9, pp. 1525-1534.

Wang, J, Zhou, D, Dai, Z & Li, X 2021, ‘Association between systemic immune-inflammation index and diabetic depression’, Clinical interventions in aging, pp. 97-105.

William, D & Suhartono, D 2021, ‘Text-based depression detection on social media posts: A systematic literature review’, Procedia Computer Science, vol. 179, pp. 582-589.

Woody, CA, Ferrari, AJ, Siskind, DJ, Whiteford, HA & Harris, MG 2017, ‘A systematic review and meta-regression of the prevalence and incidence of perinatal depression’, Journal of Affective Disorders, vol. 219, pp. 86–92.

World Health Organization 2017, ‘Depression and other common mental disorders: Global health estimates’, World Health Organization, <https://iris.who.int/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1>.

World Health Organization 2021, Suicide, retrieved from <https://www.who.int/news-room/fact-sheets/detail/suicide>.

Yadollahi, A, Shahraki, Ameneh Gholipour & Zaiane, Osmar R 2017, ‘Current State of Text Sentiment Analysis from Opinion to Emotion Mining’, ACM Computing Surveys (CSUR), vol. 50, pp. 1 - 33. http://doi.org/10.1145/3057270

Zheng, K, Chu, J, Zhang, X, Ding, Z, Song, Q, Liu, Z, Peng, W, Cao, W, Zou, T & Yi, J 2022, ‘Psychological resilience and daily stress mediate the effect of childhood trauma on depression’, Child Abuse & Neglect, vol. 125, p. 105485. https://doi.org/10.1016/j.chiabu.2022.105485

Zhou, S & Mohd, M 2025, ‘Mental Health Safety and Depression Detection in Social Media Text Data: A Classification Approach Based on a Deep Learning Model’, IEEE Access. https://doi.org/10.1109/ACCESS.2025.3559170

Zhu, J, Sun, L, Zhang, L, Wang, H, Fan, A, Yang, B, Li, W & Xiao, S 2020, ‘Prevalence and influencing factors of anxiety and depression symptoms in the first-line medical staff fighting against COVID-19 in Gansu’, Frontiers in Psychiatry, vol. 11, p. 386. https://doi.org/10.3389/fpsyt.2020.00386