Understanding Suicidal Tweets Using Sentiment Analysis and Topic Modelling

by

Reshma Manuchandran Nair

3109089

r.manuchandrannair@student.utwente.nl

Submitted in partial fulfillment of the requirements for the degree of Master of Science,

program Positive Clinical Psychology and Technology, University of Twente

2024

Supervisors:

Dr. Westerhof, Gerben, University of Twente

Dr. Gobyn, Saartje, University of Twente

Reshma Manuchandran Nair

MSc Positive Clinical Psychology and Technology, Faculty of Behavioral, Management and Social Sciences

Location:

University of Twente

P.O. Box 217

7500 AE Enschede

The Netherlands

Abstract

Introduction: This study investigates the complicated emotional experiences that individuals with suicidal ideation share on social media, particularly X. With an emphasis on suicidal thoughts, the study attempts to investigate the major themes and sentiment in tweets connected to depression.

Methods: Two major publicly available datasets of tweets on depression and suicide were utilized. With an emphasis on "depressed" tweets, the first dataset consists of 57,391 tweets classified into "indifferent," "happy," and "depressed" classifications. For this research, the classification “depressed” was selected which were 24,662. 2,460 tweets classified by content severity make up the second dataset, which focuses on messages that indicate the classification of significant disruption. 820 tweets were taken for this research from this classification. Therefore, a total of 25,482 tweets were used. To find the main themes associated with the tweets, R Studio was used for topic modelling. Using Orange's VADER and SentiArt tools, sentiment analysis was carried out, and Python was used to create correlation matrices and histograms.

Results: (1) Suicidal ideation and anxiety, (2) life reflection and negative experiences, and (3) depression and hopelessness were the three main themes that emerged from topic modelling. Sentiment analysis showed that pleasant sentiments were infrequent in tweets and that most messages expressed negative views, including fear, anger, sadness, and disgust. Correlation analysis revealed weak linear relationships between the themes and sentiments, indicating minor tendencies towards negative emotions, with Theme 2 generally showing neutral emotional expression. These findings highlight the complexity of emotional expression in depression-related tweets.

Conclusion: The results highlight how important social media is for giving up-to-date information about mental well-being and stress and the need for better algorithms to identify psychological discomfort and guide mental health interventions. The promise of sentiment analysis in the early detection and treatment of suicidal ideation is demonstrated by this work. It is encouraged to do additional research on longitudinal patterns and real-time monitoring to improve the efficacy of mental health support on social media platforms like X.

Introduction

In recent years, the increasing popularity of social media has completely changed the way people express themselves and communicate (Lahiry, Choudhury, Chatterjee, & Hazra, 2019). X has become one of those platforms where people can freely share their ideas, feelings, and experiences. This level of transparency has also drawn attention to certain concerns, particularly regarding mental health issues and suicidal ideation (De Choudhury, Counts, & Horvitz, 2013). The widespread occurrence of suicide-related tweets on X underlines the importance of understanding and dealing with this problem, both for the well-being of people and for the general online users (Coppersmith, Dredze, & Harman, 2014).

Suicidal ideation, defined as the thought of harming oneself or ending one's life, is a growing concern within online communities (Haque, Islam, Islam, & Ahsan, 2022). The significant presence of suicide-related tweets underscores the need to explore this issue deeply (De Choudhury, Counts, & Horvitz, 2013). Metzler (2022), demonstrated a sizable number of tweets about suicide, correctly labelled around 88% of the tweets, can be accurately identified and classified by deep learning models such as BERT and XLNet. This discovery validates the prevalence of suicidal tweets on X, enabling researchers to examine the influence of these tweets on suicide rates and behaviours related to seeking help. This capacity to swiftly filter and evaluate massive data sets can improve our comprehension of the ways in which social media content affects mental health and suicidal behaviour (Metzler, 2022).

Through our research, the main aim is to contribute to the understanding of suicidal ideation by examining the content and tone of tweets related to suicidal ideation. This decision is well-justified given the proven link between depression and suicidal ideation. Depression is a significant risk factor for suicide ideation and attempts, according to various psychological and psychiatric studies. (Hawton, Saunders, & O'Connor, 2012). Individuals suffering from depression frequently exhibit symptoms such as hopelessness, which is significantly connected with suicidal ideation (Nock M. K., Hwang, Sampson, & Kessler, 2010). Analysing tweets expressing depression can thus provide valuable insights into the patterns and triggers of suicidal thoughts. Furthermore, social media platforms such as X provide a real-time insight into users' thoughts and feelings, making it an invaluable source of information for studying the dynamics of mental health concerns.Analyzing these tweets can reveal insights into the frequency, mental states, and impact of suicidal thoughts on both the individuals who post them and those who interact with such content (Coppersmith, Dredze, & Harman, 2014). Therefore, one of the main objectives of this research is to understand the type and degree of distress conveyed in the suicidal tweets by different machine learning techniques (Zhang, Sun, Ren, & Shen, 2020). Further insightful analysis can be obtained by looking at the linguistic characteristics of these tweets, which can be help create targeted interventions and support networks (O'dea, et al., 2015); (Jashinsky, et al., 2014).

Despite advances in the mental health care field, understanding and preventing suicide still presents significant challenges. Studies have shown that suicidal ideations and the act of suicide are two separate constructs, with most people who go through suicidal ideations do not act on them (World Health Organization, 2020). Accurately predicting suicide risk is a significant challenge due to the complex factors involved, which can vary greatly from person to person. The stigma associated with discussing suicide often discourages individuals from seeking help, adding to the difficulty of providing appropriate interventions (Burnap & al., 2018). The regularity of suicidal tweets on twitter offers insight into the extent to which people express their distress and consider self-harm or suicide when using the platform (De Choudhury, Counts, & Horvitz, 2013).

At present, suicide is still a complex and difficult problem that must be effectively addressed by clinicians. It is important to understand that not everyone who has suicidal thoughts or ideations will commit suicide (Sher, 2004). Identifying and intervening for people at risk of suicide presents several challenges for clinicians. The challenge of accurately predicting suicide risk is one of the main obstacles. Suicidal thoughts and behaviours can be caused by a variety of complex factors that can vary significantly from one individual to another. Additionally, people are often discouraged from asking for help or speaking openly about their distress because of the stigma associated with discussing suicide. Moreover, time and resource limitations in the mental healthcare systems may hinder comprehensive suicide prevention efforts. Due to lack of time and high patient cases, it can be difficult for clinicians to conduct comprehensive risk assessments, provide appropriate support, and implement evidence-based interventions. Arendt, Scherr, Niederkrotenthaler, and Till, (2018) discovered that suicide tweets frequently contain significant levels of negative emotional content as well as specific language characteristics associated with suicidal ideation. Their findings suggested that exposure to such tweets could increase suicidal thoughts and acts among individuals who are vulnerable. This emphasizes the significance of monitoring and addressing suicidal content on social media sites with the goal to limit the harm that it causes.

There are multiple previous research which shows the effectives of studying suicidal tweets through text mining and NLP. For instance, Roy et al. (2020) uses the "Suicide Artificial Intelligence Prediction Heuristic (SAIPH)" which is a novel algorithm that uses publicly available data, specifically from X, to predict the likelihood of suicidal ideation. The researchers used X data to train multiple neural networks to do this. These data were compared to psychological factors linked to suicide, such as anxiety, despair, stress, loneliness, hopelessness, and feelings of burden. Going one step further, the researchers used the neural network outputs to train a random forest model to predict binary suicidal ideation (SI) status. An independent group of suicidal ideators and control events were used to test the model. The term "suicidal ideators" describes people who have disclosed suicidal thoughts or plans, which can be recognized by their posts and actions on social media. These expressions could be explicit statements of suicide like "I want to kill myself," "I don't want to be here anymore," or "I'm thinking of ending it all" or more subdued indicators of suicidal ideation like "I'm tired of everything," "nothing matters anymore," or "I feel trapped," which can be linked to a person's emotional state correlating with suicidal ideation that are picked up through language and interaction patterns analysis. The main predictors for this study were language characteristics including the frequency of terms associated with suicidal thoughts and feelings, the general tone of the tweets, specifically searching for suicidal thoughts, and recurrent themes pertaining to mental health problems like anxiety and depression. The SAIPH algorithm has a high accuracy rate, with the final model predicting suicidal ideation at around 85%. Some of the limitations that this study encountered would include the algorithm itself even though it is promising, its wide applicability is only extended to X. Not everyone uses X, and those who do might not be honest about how they're feeling when using the service. Furthermore, the volume and quality of the training data, which can vary widely, have a significant impact on the accuracy of the model. Roy, et al., (2020) used a huge dataset of tens of thousands of tweets to train and evaluate their models, ensuring that there was enough data to generate trustworthy predictions. In the future, the SAIPH algorithm might be modified to function as a clinical decision support tool, according to the researchers. This might use existing technology to support risk monitoring and suicide screening. Such an instrument has enormous potential because it can provide real-time risk assessment and perhaps save lives. Nonetheless, additional investigation is necessary to enhance the algorithm and evaluate its suitability for various contexts and demographics.

Similarly, Kumar, et al. (2020) focused on X data from the previous two years. They understood that early identification and detection are essential in avoiding suicide attempts, and that people are increasingly using online social media as communication channels to convey their suicidal ideations. To examine text descriptions and user language, the researchers used a variety of approaches, including sentiment analysis and supervised learning techniques. Their goal was to extract valuable information that would enable detection of suicidal ideation to function as an early warning system. They extracted multiple features and provided a set of features for training the model across the dataset to identify tweets displaying suicidal intent. They employed baseline and ensemble classifiers in this procedure. The enhanced ensemble random forest (RF) algorithm outperformed existing classification techniques for suicide ideation, with an accuracy rate of 99%, according to the study's promising results. The enormous potential of their work for early detection and prevention of suicide attempts is highlighted by this high accuracy rate. The existing system previous required manual identification and basic keyword searches to detect suicidal ideation in tweets. This outdated approach was less efficient and accurate than the researchers' machine learning strategy. The machine learning algorithms utilized in the study showed increased accuracy and precision, considerably boosting the identification process. One significant drawback is the possibility of false positives, which could result in situations where the model predicts suicidal ideation incorrectly. Furthermore, manual verification of the model is still necessary, which can be labour- and time-intensive.

Sakthi, Chen, & Sathiyanarayanan (2023) gives us a better understanding of different methods of text mining to understand suicidal ideation. This specific study focused on the possibilities of social media as a tool for monitoring and assessing students' mental health. The researchers analysed sentiment in social media data using an artificial neural network called a Deep Belief Network (DBN). With consideration for sentiment in the text and other variables, the DBN is taught to recognize patterns linked to suicide ideation. Posts were sorted into neutral and highly negative categories—the latter of which can suggest suicidal thoughts—was the primary goal. The DBN model effectively and significantly accurately distinguished between various sentiment levels, which was a promising outcome. Due to its success, the CyberHelp system may serve as a tool for early suicidal detection, assisting in the identification of students who may be at risk and triggering timely replies from appropriate law enforcement or mental health specialists. The study made clear how important it is to monitor mental health by using accurate sentiment analysis and treating sensitive data in an ethical manner. The promise of AI and machine learning technology for mental health monitoring and intervention underscores the importance of this work. Additionally, it highlights how important social media is as a data source for comprehending and forecasting mental health problems in students. According to the article, the model might be implemented on a variety of social media sites to obtain a thorough understanding of a person's mental health condition. Subsequent studies might also investigate tailored intervention approaches according to the level of suicidal thoughts identified. While the dragonfly optimization approach is quite useful for optimizing a Deep Belief Network (DBN)'s hyperparameters, it has a lot of drawbacks. This method seeks to optimize network topology and learning rate to enhance learning outcomes. It draws inspiration from the swarming behaviour of wild dragonflies. The DBN may underperform, though, if it is unable to precisely determine the ideal hyperparameters. This could be due to either underfitting—a failure to recognize the complexity of the data—or overfitting—a failure to learn an excessive amount of noise from the training set. As such, the algorithm's performance depends on how well it navigates the hyperparameter space.

The above studies underscore the promise of leveraging advanced machine learning techniques to understand and predict suicidal thoughts based on social media data. Previous research has proven to be able to effectively classify and evaluate tweets that show indicators of suicide thoughts or behaviour by using neural network models and algorithms such as BERT and XLNet. This holds great promise for efforts at prevention and early intervention, which may eventually result in lifesaving (Metzler, 2022). While prediction models have shown high accuracy, they often miss the nuances and variations in tweets that truly convey suicidal ideation. Therefore, this study aims to analyze a corpus of tweets linked to depression and mental heal tweets to determine suicidal ideation and to identify patterns and variations within this population. This nuanced approach can enhance detection algorithms and intervention strategies, leading to more focused and effective mental health support on social media platforms.

This research seeks to address the current gaps by investigating engagement patterns and word associations in mental health tweets to study suicidal ideation on X. By conducting topic modelling and sentiment analysis, this study will contribute to scientific knowledge by providing a comprehensive understanding of the emotional tone, prevalent themes, and engagement dynamics surrounding these tweets on X. The research questions for this study are as follows:

RQ1: What are the key themes of suicidal ideation expressed in depression-related tweets as identified through topic modelling?

RQ2: How do the sentiments associated with suicidal ideation vary in depression-related tweets, as identified through sentiment analysis?

RQ3: What are the correlations between the themes identified through topic modeling and the sentiments from sentiment analysis in depression-related tweets?

Methods

Design

This chapter outlines the methodological strategy that is used in this study to analyse tweets related to suicidal ideation by employing sentiment analysis and topic modelling techniques. The collection of tweets related to suicide presents several limitations due to the restrictions imposed by the Twitter API due to the end of free access after Elon Musk's takeover (Calma, 2023). Therefore, alternative strategies were employed to obtain the necessary data for this study. A comprehensive search was conducted across a range of online repositories, academic databases, and public data sources like GitHub and Harvard database to identify datasets relevant to the research objectives. As a result, manual selection process was undertaken to ensure that only tweets that unambiguously expressed suicidal ideation or depression, considering the substantial correlation between depression and suicide thoughts were included in the study.

Datasets

There are two main open datasets that is utilized for this study. The first dataset, obtained from the Harvard Dataverse repository (the link for the dataset: (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VK2D4R)) consists of 57,391 tweets that have been categorized by the authors into three distinct class: "depressed" tweets (24,662 instances), "indifferent" tweets (11,923 instances), and "happy" tweets (20,806 instances). For this study, tweets labelled as "depressed" were chosen. (Nassar, Helmy, & Ramadan, 2022). The tweets were collected between 1st January 2019 to 15th April 2022. The original search terms were in Arabic, and tweets were translated into English before analysis. The translation was performed using a combination of automated translated tools and manual verification to ensure accuracy. The main search words included "depressed", "death", "suicide", "anxious", "I want to die", "loneliness" etc., were used to find the tweets. The classification was done using the VADER lexicon which assigns an emotion score to each tweet based on its words and phrases, accounting for the intensity of the sentiment communicated. The categorization process achieved a high accuracy rate of 85%-87%, which was confirmed by hand annotation and review.

The second dataset, which focuses primarily on depression and its connection to suicide ideation, consists of tweets that have been classified according to the degree of content severity in the context of mental health. Using search phrases like #depression, #mentalhealth, #anxiety, #selfharm, and #killmyself, these tweets were extracted using the official Twitter API during a month, from January to February 2021 (Surana, Yusuf, & Singh, 2022). The dataset consists of 2460 tweets, with the labels '0' designating tweets that provide advice on mental health, '1' denoting sarcasm or irritation, and '2' designating tweets that indicate mild to severe disruption. The authors completed the labelling process in two basic steps. Initially, an automated sentiment analysis technique was employed to classify tweets based on predetermined language and sentiment parameters. This initial automatic classification entailed examining tweets for certain keywords, phrases, and sentiment markers that related to the three chosen groups. Following the automated method, the authors manually evaluated and confirmed the tagged tweets to verify their accuracy and reliability. The manual annotation includes cross-referencing the automated labels with the context and content of each tweet to rectify any misclassifications and improve the dataset's accuracy.

Negation terms received special attention since they have the potential to substantially alter the meaning of a Tweet. This was deemed critical for accurately determining the sentiment and intent of each tweet. Negative values in datasets are critical for effectively capturing sentiment and meaning in text, especially in sentiment analysis and natural language processing. Negations can reverse sentiment, add contextual meaning, and improve classification accuracy. For example, "I am not happy" transmits a negative attitude despite the positive word "happy" (Pang & Lee, 2008). Including negation terms assist models to resemble real-world communication, which improves their performance (Wilson, Wiebe, & Hoffmann, 2005). In the case of evaluating tweets about depression and suicidal ideation, correctly processing negations is critical for recognizing mental health signals (De Choudhury, Counts, & Horvitz, 2013). Following this classification, the tweets went through additional preprocessing, including lemmatization, to standardize the text by reducing words to their base forms (Hutto & Gilbert, 2014). For this study, the distinct class “depressed” tweets were considered for the text mining process (Nassar, Helmy, & Ramadan, 2022). This relates to the Dataset 1 which focuses on the distinct class “depressed” for this study. Therefore, the dataset from the Harvard Dataverse repository assures that the examined tweets appropriately represent depressive emotion, making it an appropriate and credible resource for this study (Nassar, Helmy, & Ramadan, 2022).

This balanced dataset, evenly dispersed across three categories, is critical for understanding the complex nature of mental health discussions on social media. It contributes to the study by showing how severe depressive symptoms manifest in online discourse, allowing researchers to uncover patterns and triggers linked with suicidal ideation in the setting of depression (Hawton, Saunders, & O'Connor, 2012); (Surana, Yusuf, & Singh, 2022). Both Dataset 1 and Dataset 2 went through pre-processing and only tweets with the label "2" from Dataset 2 were taken into account for this study since they offer the clearest window into statements of serious mental anguish and possible suicidal ideation. The authors of the dataset personally marked and validated the labels to verify their accuracy. Surana, Yusuf, and Singh explain particular data pre-processing methods used on Dataset 2. These pre-processing methods included removing numerals, URLs, usernames, and special characters, as well as stopwords and expanding traditional abbreviations. The stopwords were removed using the Natural Language Toolkit (NLTK) package, which contains a thorough list of frequent English stopwords (Bird, Klein, & Loper, 2009). This stage is critical for removing terms that provide no useful information to the study, hence increasing the efficiency and accuracy of the text mining process. which important for all three research questions of the study. Standard abbreviations were also enlarged to their full forms to maintain textual consistency and clarity. These preprocessing techniques are critical for preparing the dataset for future analysis and assuring the validity of the findings about mental health discussions on social media.

Programs Used

Orange

Orange is a free and open-source machine learning and data visualization tool. It's a strong tool with an easy-to-use interface that lets users handle intricate data analysis and visualization tasks. Orange's 'widgets' integrate many machine learning techniques that can be tailored to meet the unique requirements of any research undertaking. This makes it simple for researchers to carry out a variety of data analytic activities, including text mining, clustering, regression, and classification (Demsar, et al., 2013). Orange is especially helpful in the context of our research because of its natural language processing (NLP) capabilities. The tweets in our datasets could be analysed effectively and efficiently thanks to its text mining features. Furthermore, Orange's data visualization features enable us to present our findings clearly and visually. The word cloud that is generated, for example, provides an initial overview of the data by displaying the most used words in their unfiltered, raw form. This is essential for giving readers a thorough grasp of the themes and attitudes covered in the data.

R studio

Because the Orange software's results did not offer sufficient conclusions, as evidenced by the inconsistency and lack of clarity in the topic modelling outputs, another software, R, was used to produce more dependable results. The software frequently shut down due to the large size of the datasets, making it difficult to handle the data efficiently for topic modelling analysis. Additionally, the visualization of the heatmap for topic modelling was not clear, which posed challenges for accurately interpreting the results. Text mining and natural language processing activities can be facilitated by using the statistical program R in combination with many packages. This methodology's main goal is to identify patterns and latent themes in a dataset of tweets that are thought to be connected to suicide ideation and yield better outcomes. The application of topic modelling techniques aids in this inquiry by helping to uncover latent subjects in the text data (Blei, Ng, & Jordan, 2003).

Python

To visualize the results of the VADER sentiment analysis from Orange, python was used to generate the distribution levels of the results. The VADER (Valence Aware Dictionary and Sentiment Reasoner) algorithm was used first to assign positive, negative, neutral, and compound sentiment scores to each tweet. The compound score is an aggregate measure of sentiment that combines positive, negative, and neutral values into one statistic. The results of the sentiment distribution, as shown in Figure 5.

To visualize the results of SentiArt analysis from Orange, python was used again to generate the distribution levels of the sentiments using histogram. The illustrations were created using the Python packages 'matplotlib'. Each emotion score was plotted on the X-axis, frequency on the Y-axis, and separate colours were utilized for each emotion to improve clarity, which is also the same for the visualization of VADER results. Figure 6 represents the histogram which were generated for each emotion to display their distribution levels to understand the different sentiments to answer the second research question. Finally, to answer the third research question, Seaborn was used to build a heatmap of the correlation matrix, which visually represents the strength of correlations between distinct emotions and the three themes from topic modelling, with colour intensity signifying correlation magnitude.

Pre-processing

The datasets used in this study had undergone initial preprocessing by the original researchers. However, additional preprocessing steps were performed to ensure the data was suitable for this study's specific analyses. Both the datasets already included initial pre-processing steps such as the removal of numerals, URLs, usernames, special characters, and stopwords, as well as lemmatization to reduce words to their base forms (Hutto & Gilbert, 2014); (Surana, Yusuf, & Singh, 2022). Negation terms were given special attention because they can substantially alter the meaning of a tweet. For example, the phrase "I am not happy" transmits a negative attitude despite the positive word "happy." Including negations was critical for accurately determining the sentiment and intent of each tweet. Additional preprocessing performed in this study included further refinement of text normalization and ensuring consistency across both the dataset for accurate sentiment and topic analysis. After cleaning, the tweets were assembled into a corpus for tokenization and text transformation, conducted on R studio separately. The preprocessing steps for the R analysis were similar but tailored to fit the requirements of the tools and methods used in R Studio. This differs slightly from the preprocessing done for the data used in Orange, which was also cleaned and

Figure 1.1 WordCloud before Pre-processing

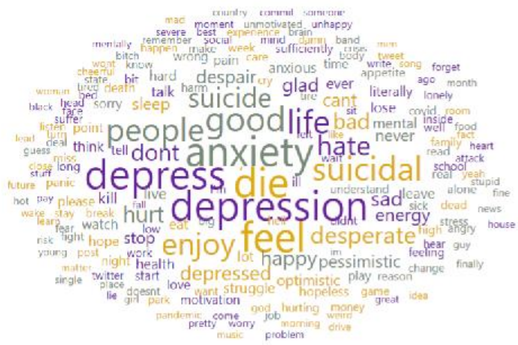

Figure 1.2 WordCloud after Preprocessing

transformed but with specific adjustments to suit Orange's processing capabilities. Tokenization involved breaking up the text into discrete terms or tokens, making stopword removal easier (Feinerer, Hornik, & Meyer, 2008). Figure 1.1 and 1.2 (page no. 13) shows the wordcloud before pre-processing and after pre-processing. It is essential to keep in mind that Figure 1.1 was generated prior to our data being pre-processed. Because it highlights the most important terms in the tweets, this visualization tool is quite helpful in providing some first understanding of the dataset's thematic substance (Fellows, 2014). Thus, this word cloud provides an initial look at the data we are working with by displaying the most commonly used words in their unfiltered, raw form. This is however different after pre-processing was done where we can see from Figure 1.2 that the major keywords for this dataset after pre-processing would be “depression”, “anxiety”, “die”, “suicidal" etc., in comparison to Figure 1.1 where the main keywords are “anxiety”, “hate”, “die”, “depress” etc.

Data analysis

Topic Modelling

To answer the first research question Latent Dirichlet Allocation (LDA) was done using the R Studio. For this, the first process is a Document-Term Matrix (DTM), a mathematical matrix that captures the frequency of phrases that appear in a collection of documents. After then, LDA is used on this DTM to identify themes, which are shown as word distributions. Different models with varying amounts of themes are developed to identify the best possible structure (Griffiths & Steyvers, 2004). The process also preprocesses the DTM to eliminate blank rows and use vocabulary trimming to improve the model. This process is critical for boosting the model's accuracy and efficiency throughout the analysis phase, as it guarantees that the data is clean and relevant, thereby enhancing the quality of the analytical results (Silge & Robinson, 2016). It is also important to know from LDA is the ideal number of topics to ensure that the topics are unique and do not intermix a lot. Therefore, using R, the ideal number of topics for the LDA model can be determined using the function “FindTopicsNumber”. This function determines which model has the best fit by calculating the fit of LDA models over a range of different topic numbers. The graphical outputs from the ldatuning software provide useful insights in examining the selection of the ideal number of topics for the use of Latent Dirichlet Allocation (LDA) inside the dataset under examination. The model tuning results using two different measures (CaoJuan2009 and Deveaud2014) are shown in Figure 2. (Cao, Xia, Li, Zhang, & Tang, 2009); (Deveaud, SanJuan, & Bellot, 2014) The number of themes that most accurately reflect the underlying structure of the data can be determined using these metrics as evaluation tools. When the CaoJuan2009 metric—represented by circular markers—is examined, it becomes clear that this measure has to be reduced. This means that the lower the value, the better and more unique the themes will be. There is a noticeable decline from topic 2 to topic 3, which is followed by an uptrend with slight peaks and valleys at topics 6, 8, and 11. After 11 topics, the metric displays a level trend, indicating that the ability of new topics to differentiate itself has reached a saturation point. On the other hand, maximizing is needed for the Deveaud2014 metric, which is represented by triangle markers. The measure shows a slow drop after a steep ascent from 2 to 3 subjects. The peak at 3 topics indicates a sharp boundary in the dataset, which decreases as the number of subjects rises. It is clear from the outcomes of the two measures that choosing a topic range of three to six is a wise decision. This range combines the two goals of model simplicity and topic distinctiveness by enclosing the peak of the Deveaud2014 metric and aligning with a stability region in the CaoJuan2009 metric. The selection of three topics presents a model of high clarity in topic distinction and is in line with the peak of the Deveaud2014 measure. Nevertheless, adding six topics to the model could allow for a more in-depth comprehension of the dataset while still preserving a logical and understandable model structure. For this study, we chose three themes because they exhibit high clarity in topic differentiation, aligning with the peak of the Deveaud2014 measure, and offer a basic yet effective model structure. This selection guarantees that the model is interpretable while capturing the key themes in the dataset. Although a greater selection of subjects (up to six) could provide a more in-depth understanding, the three themes selected achieve a compromise between comprehensiveness and clarity, making them appropriate for the given datasets (Schweinberger, 2023). (Griffiths & Steyvers, 2004)

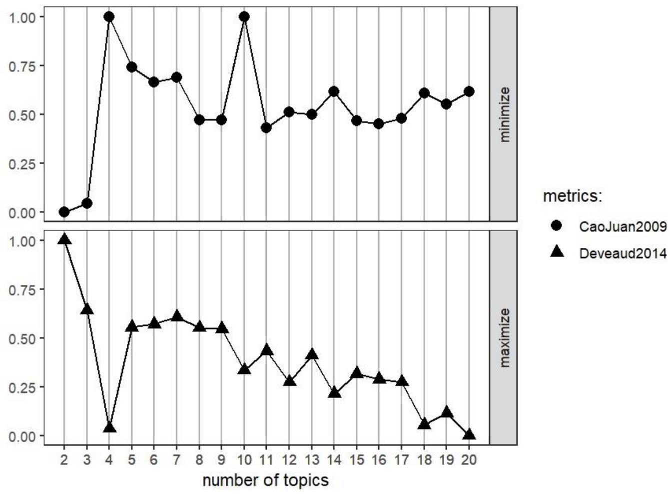

Figure 2: CaoJuan2009 and Deveaud2014 metrics for determining the ideal number of topics.

Sentiment Analysis – VADER and SentiArt

Sentiment analysis has grown in popularity because of its potential applications in a variety of sectors, including social media platforms like X. (Liu, 2022) More recent studies have used sentiment analysis to understand and interpret tweet sentiments. For example, in a 2020 study, Metzler and colleagues used sentiment analysis to identify potentially harmful and protective suicide-related information on X. They created a machine learning model that accurately identified tweets indicating suicide intent, demonstrating the value of sentiment analysis in mental health surveillance and intervention. (Metzler, 2022) Sentiment analysis can be approached in a variety of ways, including machine learning (ML), lexicon-based methods, and hybrid approaches. (Cambria, Schuller, & Xia, 2013) This study used the Orange data mining software to answer the second research question. This software recommends a user-friendly interface for complex analysis tasks, such as sentiment analysis. In the Orange application, two key sentiment analysis tools were employed: VADER (Valence Aware Dictionary and Sentiment Reasoner) and SentiArt. VADER is a vocabulary and rule-based sentiment analysis tool that is specifically suited to sentiments expressed on social media (Hutto & Gilbert, 2014). It uses a combination of qualitative and quantitative approaches to analyse text data and is adept at dealing with emoticons, slang, and other informal language commonly seen on social media sites such as X. The final result will comprise columns for tweet text and VADER sentiment ratings ('negative', 'neutral', 'positive', and 'compound'). The 'compound' score is a single unidimensional measure of the sentiment of a tweet, calculated by adding the valence ratings of each word in the tweet and normalizing the sum to be between -1 (most extreme negative) and +1 (most extreme positive)SentiArt, on the other hand, is a sophisticated sentiment analysis tool that analyses sentiments in text data using artificial intelligence (Yadav, 2020). It is specifically built to handle massive amounts of text data and deliver more nuanced insights on the sentiments expressed (like sadness, happiness, anger etc.) in the text. The process begins with feature extraction using word embeddings such as Word2Vec and BERT (Mikolov, Chen, & Corrado, 2013) followed by sentiment classification and emotion detection using deep learning models such as LSTM (Hochreiter & Schmidhuber, 1997). Sentiment scores for positive, negative, and neutral attitudes were combined to determine broad trends of emotions. Collectively, these methods enable us to undertake a thorough sentiment analysis of the tweets in the dataset. The sentiment scores obtained from these techniques provided useful insights into the emotional states and attitudes of X users who expressed suicidal ideation. The process is shown in Figure 3.

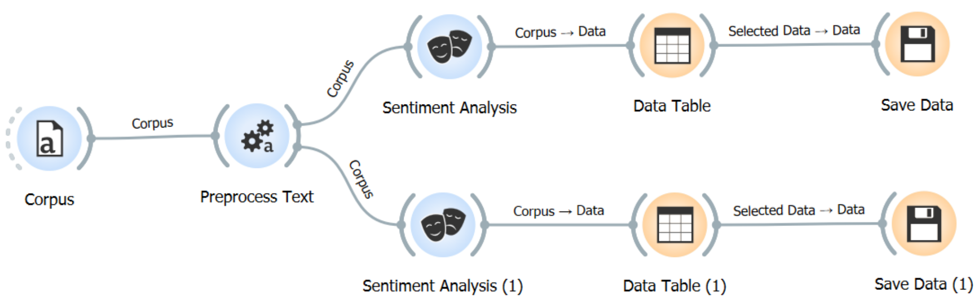

Figure 3: Process of Sentiment Analysis (VADER and SentiArt) in Orange

Results

RQ1: What are the key themes of suicidal ideation expressed in depression-related tweets as identified through topic modelling?

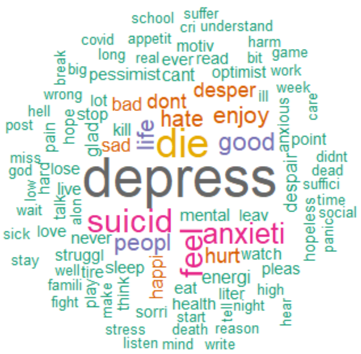

Figure 4.1: Worcloud of dataset 1 (Nassar, Helmy, & Ramadan, 2022)

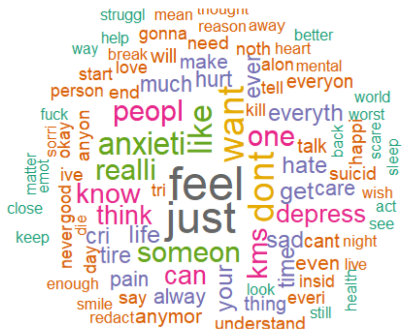

Figure 4.2: Wordcloud of dataset 2 (Surana, Yusuf, & Singh, 2022)

WordCloud comparison from Figure 4.1 and Figure 4.2

The two wordclouds investigated are from datasets 1 (Figure 4.1) and 2 (Figure 4.2) that are critical for understanding the expressions and patterns of suicidal ideation and depression on X. Both datasets place a strong emphasis on terms connected to mental health, such as "depress," "anxiety," "suicid," "die," and "feel," showing a concerted effort to capture the gravity of these discussions on social media. These phrases, which appear often in both wordclouds, highlight the key themes of depression and suicidal thoughts, demonstrating their significance in the study of mental health issues. The wordcloud from Figure 4.1 contains a broader range of terms, varied in size and frequency, indicating a diverse lexicon and a spectrum of sentiments. The use of both negative and positive words, as well as those denoting neutral feelings, demonstrates the dataset's broad commitment to collecting many facets of mental health. For example, terms like "good," "life," "enjoy," "happy," and "people" imply a more complex portrayal of the emotional landscape inside the dataset, which includes tweets labelled "depressed," "indifferent," and "happy." This diversity is most likely caused by the dataset's usage of negation and sophisticated sentiment scoring in the categorization of the dataset to better reflect the intricacies of real-world communication. In contrast, Figure 4.2's wordcloud shows a greater use of direct representations of personal feelings and aspirations. Words like "feel," "just," "want," "kms" (kill myself), and "realli" imply an emphasis on current emotional states and personal experiences. This dataset appears to capture a more conversational and introspective tone, as individuals explain their current mental state and interactions with others. The use of terms like "someone," "people," and "like" implies a preference for discussing relationships and social situations alongside mental health difficulties. This more detailed depiction of personal feelings and relationships sheds light on how people express distress and seek help or understanding on social media platforms.

Themes from Topic Modelling

The topic modelling analysis revealed three unique themes in the X data as given in Table 1. Each theme covers a different aspect of the emotional and psychological states stated by users in tweets, with a particular emphasis on suicide ideation.

Theme 1 - Suicidal Ideation and Anxiety: This theme is characterized by acute anxiety and suicide thoughts. The example tweets in this theme demonstrate a high level of mental suffering, with people frequently describing their worry, depression, and feelings of desperation. This theme implies that people are experiencing great emotional distress, as evidenced by key phrases such as "suicid," "anxieti," "depress," and "desper" etc. This shows that those showing suicidal ideation and anxiety may require further mental health aid and interventions. It also shows that social media is being utilized as a forum for people to freely express their mental health difficulties, emphasizing the significance of acknowledging and responding to these expressions to provide timely and appropriate help.

Theme 2 – Life reflection and negative experiences:

This topic represents the users' overall emotional anguish and unpleasant experiences throughout their life. This category's tweets express dissatisfaction with repeated negative experiences, as well as a sense of being trapped in undesirable environments. Words like "die," "hate," "bad," and "despair" are regularly used to emphasize the severity of emotional discomfort and negativity. However, the usage of words like "good," "life," "enjoy," and "happi" implies a complicated emotional environment in which users encounter fleeting instances of optimism amidst persistent negative experiences. This theme emphasizes the need to understanding emotional state's fluctuation when evaluating mental health care and therapies.

Theme 3 – Depression and Hopelessness

This theme focuses on symptoms like depression and hopelessness. The example tweets in this theme show a great sense of tiredness and frustration, especially while attempting to express their problems to others. The frequent use of phrases like "depress," "feel," "dont," "sad," and "hopeless" imply deep-seated feelings of grief and despair. Sleep difficulties and a lack of energy are two of the most frequently stated symptoms of depression. The frequent use of phrases such as "hurt," "tire," and "struggl" emphasizes the continual struggle people have with their mental health.

Table 1: Themes from Topic Modelling

Topic	**Example Tweets**	**Most Used Words**
Topic 1: Suicidal Ideation and Anxiety	1. "Feel anxious but feel mistake." 2. "I think I'll buy a gun, I'll kill myself anxiety." 3. "You're so afraid to tell people how you feel because you fear rejection so you bury it deep inside yourself where it only destroys you more." 4. “I can’t see any reason to keep going on I’ll kill myself anxiety.” 5. “Im losing my motivation for everything and the stress is piling up. I just want to be dead.” 6. “Just like the veterans who so desperately need help with their healthcare & mentalhealth the PTSD.”	Suicide, anxiety, people, desper, mental, health, ill, lose, panic, stress
Topic 2: Life reflection and Negative Experiences	1. "So sick of going through the same shit over and over again." 2. "Completely understand upset sorry." 3. "Cute forever but lay depressed." 4. “I hate my life I just want to end it.” 5. “Even you are at home, you’ll face unsafety, life is sad and hurtful, I wanna end me.” 6. “First meal of the day i enjoyed it very much anxiety levels high so anorexia rears its head everyday is a challenge.”	Die, good, life, enjoy, hate, bad, happi, glad, despair, kill, depress, hope, optimist, hurt, sorry
Topic 3: Depressi	1. "Twitter cautiously pessimistic." 2. "Silence doesn't always mean yes, sometimes it means I'm	Depress, feel, sad,

on and Hopelessness	tired of explaining to people who don't even care to understand." 3. "Bad can't sleep depression." 4. "Every night I dread having to sleep very night I pressure myself into the same routine to cope with the constant anxiety." 5. "Yeah I always struggle at this time of year, knowing all my family will be together and there's me who hates them all." 6. "I feed off of bad energy. The sadder I feel the better."	hurt, energi, sleep, never think talk, live, struggle, anxious, tired, eat

The representative examples provided in Table 1 elucidate the manner in which these themes manifest in real-world tweets, thereby offering concrete illustrations of the emotional states associated with suicidal ideation. These examples serve to underscore the complexity and multifaceted nature of the emotions expressed by individuals on social media platforms such as X. The tweet exemplifying the Suicidal Ideation and Anxiety theme, " I can't see any reason to keep going on I'll kill myself anxiety," reveals a profound sense of despair intertwined with overwhelming anxiety. This tweet not only communicates a clear desire to end one's life but also highlights the unbearable nature of the anxiety experienced. The explicit mention of anxiety as a contributory factor to suicidal thoughts underscores the need for interventions that address both anxiety and suicidal ideation simultaneously. Another tweet associated with the Life Reflection and Negative Experiences theme, " So sick of going through the same shit over and over again " reflects a deep-seated sense of hopelessness and futility derived from a history of negative experiences. This retrospective view, where past failures and traumas are perceived as continuous and unending, contributes significantly to the individual's current emotional state. One example for the Depression and Hopelessness theme, "Every night I dread having to sleep very night I pressure myself into the same routine to cope with the constant anxiety", vividly conveys the depth of emotional pain and the pervasive sense of hopelessness experienced by the individual. The act od dreading to sleep suggests an overwhelming and inescapable nature of constant anxiety, indicating a state of mental suffocation. This profound hopelessness and constant anxiety is a critical risk factor for suicidal ideation. In summary, the complexity of these emotional expressions highlights the necessity for tailored mental health interventions that address the specific psychological needs of individuals. By understanding the nuanced ways in which suicidal ideation is articulated, mental health professionals can develop more effective strategies to support those at risk, ultimately contributing to better mental health outcomes.

RQ2: How do the sentiments associated with suicidal ideation vary in depression-related tweets, as identified through VADER and SentiArt sentiment analysis?

Sentiment Analysis – VADER

The positive sentiment distribution (shown in Figure 5), depicted on the X-axis with scores ranging from 0 to 0.843, and the Y-axis indicating tweet frequency, reveal that the vast majority of tweets have low positive sentiment scores. A considerable percentage of tweets received a 0 rating, which comes from both the datasets indicating a low level of positive sentiment in the context of suicide ideation. For example, one tweet read, " millions of people around the world struggle silently with anxiety fear isolation depression lacking acce," with a positive score of 0.0, a negative score of 0.576, a neutral score of 0.424, and a composite score of -0.9118. Another tweet said, "hello twitter hello universe sucks to be alone again tonight with no one to talk to not sure why i even joined" with a positive score of 0.0, a negative score of 0.338, a neutral score of 0.662, and a compound score of -0.7692. A third tweet read: "had a bad night of anxiety and looked into the link between urination and anxiety every" received a positive score of 0.0, a negative score of 0.365, a neutral score of 0.635, and a compound score of -0.7096. The mean positive score is 0.0796 (SD = 0.13), showing that positive sentiment is mostly absent from the tweets.

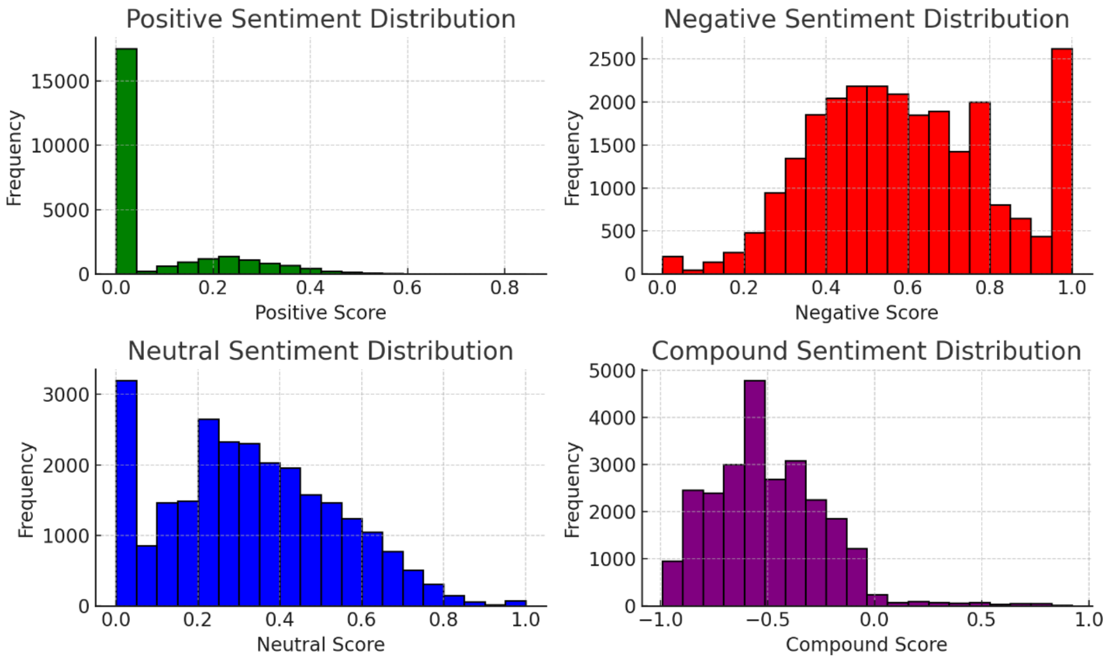

Figure 5: Histogram of VADER Sentiment Analysis

In contrast, the negative sentiment distribution, which ranges from 0 to 1 on the X-axis, shows a prevalence of high negative sentiment scores. The Y-axis displays the frequency of tweets with matching negative scores, with a mean negative score of 0.5936 (SD = 0.22). This implies a strong negative attitude throughout the tweets, emphasizing the unpleasant emotional reaction linked with the issue of suicidal ideation. Neutral sentiment ratings have a moderate distribution, with values ranging from 0 to 1 on the X-axis and frequency on the Y-axis. The mean neutral score is 0.3267, indicating that a considerable proportion of tweets are neutral in tone, neither excessively positive nor negative. This suggests a balanced presence of neutral sentiment in the tweets, implying objectivity or a lack of significant emotional bias in the writing. The compound sentiment distribution, with scores ranging from -0.9912 to 0.9206 on the X-axis and frequency on the Y-axis, demonstrates that compound sentiment scores are primarily negative. The mean compound score is -0.5015, reinforcing the analysis' overall negative sentiment trend. In summary, the sentiment analysis results from the VADER results that tweets about suicidal ideation are predominantly negative, with limited positive sentiment and a moderate quantity of neutral emotion. These findings demonstrate the ubiquitous negative emotional environment around discussions about suicidal thoughts on social media platforms.

Sentiment Analysis – SentiArt

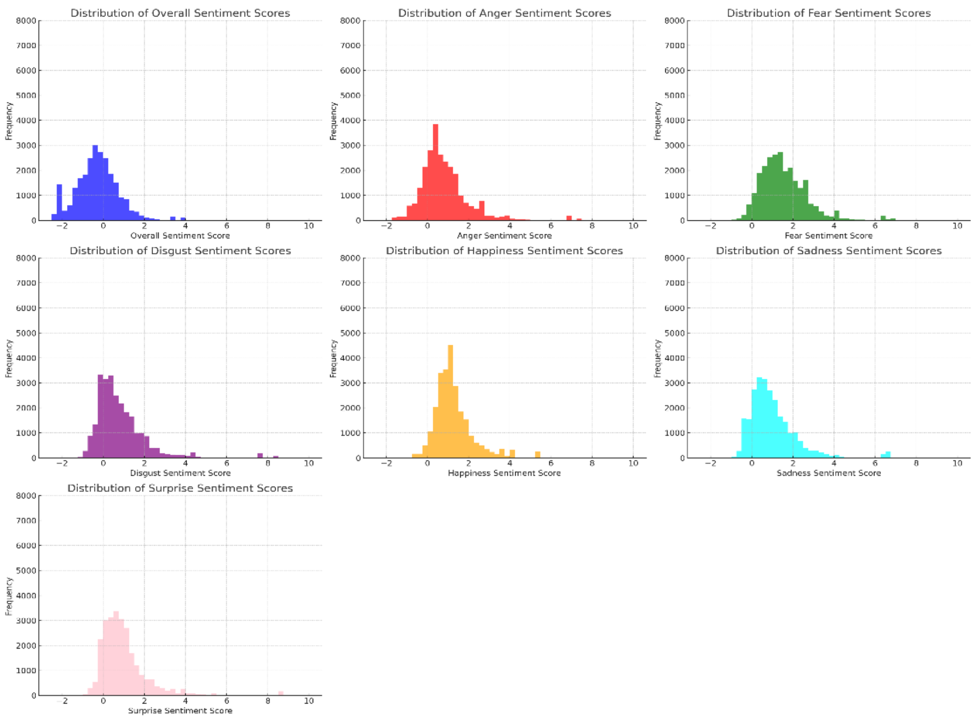

Figure 6: Histogram of SentiArt Sentiment Analysis

Results from the Histogram (shown in Figure 6) compares distribution levels across different emotions. The X-axis for all histogram depicts scores ranging from -2.5 to 10, while the Y-axis represents tweet frequency ranging from 0 to 8000. The first histogram denotes the overall sentiment of the datasets. The histogram shows that sentiment scores are widely spread, with the highest frequency of values falling between -2.5 and 10 and mean of -0.35 (SD = 1.11) indicating a significant presence of neutral to slightly negative attitudes in the tweets. This range indicates that many tweets have a neutral or balanced emotional tone, rather than extreme positive or negative feeling. The second histogram denotes the anger sentiment of the datasets. The histogram shows that the anger sentiment scores vary greatly, ranging from around -1 to 4 and a mean of 1.381 (SD = 0.261). The maximum frequency of anger scores occurs between 1 and 2, indicating a strong presence of anger in the tweets. This implies that anger is a frequently stated emotion among individuals talking about suicidal ideation, with a significant percentage of tweets expressing moderate to high degrees of anger. The third histogram denotes the overall fear sentiment scores of the datasets. The most frequency of fear scores was found between 0.5 and 2.5, with a mean of 1.58 (SD = 1.17) indicating that many tweets indicate moderate degrees of fear. This distribution illustrates fear as a common emotion in the context of suicidal ideation, with a large number of tweets expressing varied levels of fear and anxiety. The fourth histogram denotes the overall disgust sentiment scores of the datasets that were utilized. The histogram shows that disgust scores are concentrated between 0.5 and 2.5, with the maximum frequency around 1.5, and a mean of 1.227 (SD = 0.522) showing that disgust is commonly expressed in tweets concerning suicidal ideation. The occurrence of high disgust scores indicates that this feeling is highly linked to discussions related to this issue. The fifth histogram denotes the overall happiness of the datasets. The highest frequency of happiness ratings ranges between 0.5 and 1, with a mean of 0.757 (SD = 0.146) indicating that positive sentiment is less common and often low in intensity. This suggests that displays of happiness are uncommon in tweets concerning suicidal ideation, and when they do appear, they are typically mild in nature. The sixth histogram denotes the overall sadness sentiment scores of the datasets. The histogram indicates a wide range of sadness scores, with the largest frequency happening between 1.5 and 2.5, with a mean of 1.927 (SD = 0.283) indicating that sorrow is an important emotion in discussions about suicidal ideation. This distribution shows that sadness is the most common emotional response in these tweets. The final histogram denotes the overall surprise sentiment scores of the datasets. The surprise sentiment scores are broadly spread, with the maximum frequency falling between 1 and 2, with a mean of 1.379 (SD = 0.271) showing a variety of expressions of surprise in the tweets. This shows that surprise, which is frequently associated with unexpected events or revelations, plays an important role in discussions about suicide ideation.

In conclusion, the most common sentiments noted in tweets are sadness and fear, emphasizing the severe psychological distress associated with suicidal ideation. Anger, disgust, and surprise occur at mid-level frequency, although happiness is significantly less often and often weak in intensity. This distribution highlights the overwhelmingly unpleasant emotional landscape around conversations of suicidal thoughts on social media.

RQ3: What are the correlations between the themes identified through topic modeling and the sentiments from SentiArt analysis in depression-related tweets?

To understand the relationship between the three themes found by topic modelling and the sentiment scores from the SentiArt study, a correlation matrix was created in Python. This matrix computes the linear correlations between the independent topic scores and the sentiment scores. The values vary from -1 (perfect negative correlation) to +1 (perfect positive correlation), with values near zero indicating no linear relationship. The topic modelling analysis revealed three unique themes: Theme 1 (suicidal ideation and anxiety), Theme 2 (life reflection and negative experiences), and Theme 3 (depression and hopelessness). The scores for these topics, generated from the topic modelling process, indicate the extent to which every theme appears in each tweet. Similarly, SentiArt's sentiment scores quantify numerous emotional expressions in tweets, such as overall sentiment, anger, fear, disgust, happiness, sadness, and surprise. The Pearson correlation coefficient, which is appropriate for continuous data, was employed in this investigation. To demonstrate the correlation matrix, a heatmap (Figure 7) was used, which depicted correlation values with colour intensity. The heatmap's colour scale ranged from blue (negative correlation) to red (positive correlation), with annotations indicating the exact correlation values. This graphical representation helps to comprehend how different themes and attitudes interrelate in the dataset.

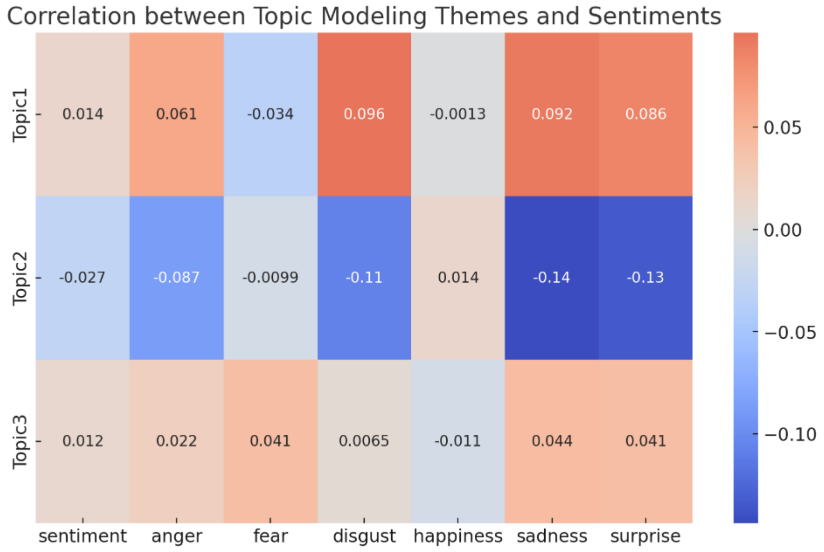

Figure 7: Correlation matrix between the topics from topic modelling and sentiments from sentiart analysis

The analysis revealed that Theme 1: Suicidal Ideation and Anxiety had extremely weak positive correlations with overall sentiment (0.014), anger (0.061), disgust (0.096), sadness (0.092), and surprise (0.086), as well as a weak negative correlation with happiness (-0.001) and fear (-0.034). The results of this analysis indicate that tweets related with Theme 1 reflect small tendencies toward negative emotions such as anger, disgust, and sadness. However, the total sentiment correlation is nearly zero, showing a weak link with positive sentiments.

Theme 2: Life Reflection and Negative Experiences showed weak negative associations with overall sentiment (-0.027), anger (-0.087), fear (-0.009), disgust (-0.11), sadness (-0.14), and surprise (-0.133), as well as a weak positive correlation with happiness (0.014). These findings suggest that Theme 2 is highlighted by lower levels of strong negative emotions like sadness and surprise, as well as significantly higher display of positive sentiments like happiness. While the general emotional tone is less strong than the other themes, it tends toward a subdued positive sentiment rather than being entirely neutral. The very small positive association with happiness indicates that there are few instances of positive sentiment among the mostly negative sentiments.

For Theme 3: Depression and Hopelessness revealed very slight positive correlations with overall sentiment (0.012), anger (0.022), fear (0.041), sadness (0.044), and surprise (0.041), as well as very weak negative correlations with happiness (-0.011) and disgust

(0.006). These associations imply that tweets about this theme have a relatively small inclination to communicate negative emotions like fear and sadness. The theme's emphasis on negative emotional states is reinforced by its lack of correlation with overall sentiments and happiness.

Ultimately, the correlation analysis of the topics found by topic modeling and the sentiments from SentiArt analysis reveals weak linear relationships. Theme 1 (Suicidal ideation and Anxiety), and Theme 3 (Depression and Hopelessness) are associated with minor tendencies towards negative sentiments, whereas Theme 2 (Life reflections and Negative experiences) have a typically neutral emotional sentiment. The low correlations suggest that while themes provide some insight into tweet content, they do not accurately predict sentiment. This emphasizes the need for more nuanced analytical techniques to fully comprehend the emotional dynamics in depression-related discourse influencing suicidal ideation on social media.

Discussion

The current study aimed to investigate the intricate and multifaceted emotional experiences related with depression tweets which contributes to suicidal ideation as they are expressed on X. The primary focus of this study was the analysis of themes and sentiments in labelled tweets. The first research question answers three main themes. Theme 1: Suicidal Ideation and Anxiety contains tweets about severe anxiety and suicidal ideation. Theme 2: Life Reflection and Negative Experiences collects tweets on emotional anguish and unhappiness with life, which include both negative and positive phrases. Theme 3: sorrow and Hopelessness focuses on intense feelings of sorrow and hopelessness, with tweets emphasizing sadness, irritation, and symptoms such as insomnia. The second research question was answered using The VADER and SentiArt tools which were used to analyse sentiment and found that the majority of it was negative. VADER research revealed a low mean positive sentiment score, a high mean negative sentiment score, and moderate neutral sentiment. SentiArt analysis revealed that sadness and fear had the highest mean levels, which corresponded closely to the primary themes identified. This was followed by anger, disgust, and surprise, all of which were observed to varying degrees. Positive emotions were found, but only on occasion, reflecting the tweets' overall negative emotional landscape.These data illustrate the prevalence of negative emotions in tweets. The third research question was answered through correlation analysis of the three themes and SentiArt sentiment ratings. The analysis of the correlation between themes and sentiments in depression-related tweets revealed generally weak linear relationships. Specifically, Theme 1 (Suicidal Ideation and Anxiety) exhibited weak positive correlations with overall sentiment, anger, disgust, sadness, and surprise, while showing a general trend of weak positive correlation with happiness and a weak negative correlation with fear. Theme 2 (Life Reflection and Negative Experiences) had weak negative correlations with most sentiments, indicating a neutral to slightly negative emotional expression. Theme 3 (Depression and Hopelessness) showed slight positive correlations with negative sentiments like anger, fear, sadness, and surprise, but a weak negative correlation with happiness. These findings suggest that while themes offer some insights into tweet content, they do not accurately predict the sentiment, highlighting the complexity of emotional expression on social media.

The study's findings highlight the crucial importance of social media, particularly X, as an environment where people can openly express their ideas and feelings. This provides essential insights on their emotional state, resulting in a rich dataset suitable for further analysis and research. The first theme, suicidal ideation, and anxiety highlighted the extreme psychological distress that people are experiencing. This suffering frequently showed as anxiety and emotions of desperation. Importantly, this theme is compatible with existing psychological literature, which has consistently shown that anxiety disorders are highly associated with an increased risk of suicide behaviour (Barlow, 2004). Recent studies have also shown the link between anxiety disorders and a higher risk of suicide behaviour. For example, a study discovered that anxiety remained an independent risk factor for suicide behaviours even after controlling for other variables (Ruth Lynfield, Lori Freeman, & Scott Becker, 2021). Furthermore, research conducted during the COVID-19 pandemic revealed a considerable increase in anxiety and linked suicide ideation among public health personnel, underscoring the need for focused mental health interventions. (Bryan, et al., 2014)

The second theme, "life reflection and negative experiences," captured users' overall emotional distress as well as the difficult incidents they had encountered throughout their lives. This theme emphasizes the significance of recognizing emotional variations while assessing mental health and therapeutic approaches. However, research has revealed that those with suicidal ideation prefer brooding over reflection, but non-suicidal people with a history of depression exhibit the opposite pattern. (Crane, 2007) This could mean that many have expressed difficult incidents as a method of brooding rather than life reflections. This finding is corroborated by the literature on emotional dysregulation, which is a substantial contributor to suicide ideation and actions (Pisani, Murrie, & Silverman, 2016). The presence of both positive and negative emotional expressions in this topic reflects the concept of emotional ambivalence, which has been linked to an elevated risk of suicide (Hawton, i Comabella, & Haw, 2013).

The third theme, depression, and hopelessness, emphasizes feelings of exhaustion, dissatisfaction, and a lack of energy, which are common symptoms of depression. A significant amount of research, including a study conducted by (2012), has indicated that depression and suicide are both likely to be linked to hopelessness, which could be the result of psychological stresses caused by social structure and life events (Zhang & Li, 2013). This theme is also consistent with a large body of research demonstrating the strong link between depression, hopelessness, and suicide (Nock M. K., et al., 2009). This theme therefore emphasizes the crucial importance of understanding, support, and therapies designed to reduce the severe impact of depression on daily life. In relation to one another, these themes create a full picture of the mental health challenges associated with suicide ideation as reported by X users.

The sentiment analysis of the tweets, carried out using the VADER and SentiArt tools, provided a better grasp of the emotional undertones inherent in the tweets. For the results from VADER analysis, the main sentiment in the tweets was negative. This could also be because of the selection of only depressed/disruption classification of tweets. The overwhelming presence of negative sentiment was contrasted with the relatively low presence of positive sentiment. The study conducted by Melton, (2022) also confirms this finding as the average sentiment expressed on X was more negative (54.8%) than positive sentiment related to the COVID-19 Vaccine. Furthermore, the lack of positive attitude in these exchanges highlights the severity and gravity of the situations the users experience. The little occurrence of positive sentiment may imply a lack of optimism or positivity in the users' lives, exacerbating their distress. The data from (Hollander, 2016) study also shows that negative sentiments are more widespread in social media posts than positive sentiments. This prevalence of negative emotion is due to the nature of social media sites, where users are more prone to express concerns, setbacks, and unpleasant experiences. Positive sentiments, on the other hand, are much less common. When compared to X as a whole, it is clear that while negative attitudes are frequent, the current dataset's concentration on mental health issues exacerbates this trend. According to Wang et al., (2016), although X users normally express a mix of happy and negative experiences, tweets about mental health difficulties are disproportionately negative. This distinction emphasizes the different emotional landscape of mental health-related discussions on social media, where expressions of sadness and negative emotions are far more prominent than in ordinary X usage. It is crucial to remember that the sentiment analysis is naturally biased by the tweet selection process, especially when using VADER to find "depressed" tweets. Given that VADER was utilized to identify tweets categorized as "depressed," it is unsurprising that these tweets had a high degree of negative sentiment and a low level of positive sentiment. Additionally, a moderate level of neutral attitude was noticed. The moderate level of neutral sentiment found may represent tweets in which users share their experiences or views without overtly expressing positive or negative emotions. These could be true statements or observations about their experiences. This can be related to the results of a study done by Abdukhamidov, (2022) where during COVID-19, neutral sentiment reactions dominate social media, with health-related subjects being the most discussed, and posts from the top countries impacting worldwide reactions.

On the other hand, SentiArt sentiment analysis of tweets on suicidal ideation reveals a rich emotional setting. According to the analysis, fear and sadness had highest mean levels, indicating that these were the most often reported emotions. The moderate presence of anger, disgust, and surprise is indicated by the mid-level means of these sentiments, whilst the noticeably low level of happiness highlights the predominantly negative emotional landscape. This can be related to the study done by (Rogers, 2017) shows that negative emotions, such as anger and humiliation, add to veterans' suicide risk by creating feelings of burden and a lack of belonging. This could confirm that a lot of negative sentiments expressed by the users could add on to their suicidal ideation. Positive sentiment, indicated by instances of happiness, was uncommon and frequently low. This can also be confirmed by the study done by Moreno-Ortiz, (2022), demonstrate that positive lexical words play a minor influence in describing pleasant events. The study discovered that positive words are not commonly employed to indicate happiness. Although neutral sentiment was identified by the research, it is crucial to remember that SentiArt does not provide a specific value for neutral sentiment. Rather than belonging to a specific neutral category, the neutral tone that was detected might instead reflect the balance between the positive and negative feelings. This emphasizes the challenge to explain emotions clearly since even with accurate observations or allegations, there may be an underlying emotional bias present. The study on sentiment analysis of epidemiological surveillance data discovered that a considerable number (34%) of the comments examined had neutral sentiment tones. This implies that neutral sentiment can convey true observations or assertions without a strong positive or negative bias (Stefanis, 2023). This analysis therefore gives useful information for building methods to identify and support at-risk persons for future discussions/study.

The low correlations across all themes and sentiments indicate that the prevalence of specific themes in tweets regarding depression is not strongly linked to the expressed sentiments. The investigation highlights the complexities of emotional expression on social media, where a variety of factors influence how people express their experiences and emotions. The low correlations show that, while themes provide some insight into tweet content, they do not accurately predict sentiment. This emphasizes the need for more nuanced analytical techniques to fully comprehend the emotional dynamics in depression-related discourse that influence suicidal ideation on social media. The low correlations between themes and attitudes in the analysis have numerous significant implications. One possible explanation for these low correlations is that the tweets studied were primarily negative. Given that the majority of tweets reflect negative sentiments, it may be difficult to establish strong relationships between themes and sentiments, as the emotional tone is generally consistent. This uniformity in negative sentiment can hide more complex relationships that may exist in a more diverse data collection. Despite this, they show that thematic content in tweets does not strongly predict emotional tone, implying that users may convey their experiences and emotions in ways that are difficult to define by theme. This intricacy in emotional expression underlines the challenges of utilizing simple topic analysis to understand mental well-being on social media. Calvo et al. (2017), relates to the finding that emotional expression in text can be highly subtle and context-dependent, making simple thematic models difficult to apply. In addition, the weak correlations highlight the need for more advanced and nuanced analytical tools to capture the complex ways in which people express distress and suicidal ideation. Current methodologies may miss subtle emotional cues that are not immediately related to the identified themes. This gap highlights the need for more advanced algorithms and models that can better grasp the various ways people express their mental health difficulties. Guntuku et al. (2017) argue that combining several analytical methodologies improves the accuracy of detecting mental health disorders in social media datasets. Finally, the findings also suggest that mental health interventions based primarily on topic analysis alone may be insufficient. To more effectively identify and support those who are at risk, multifaceted analytical approaches that take into account both thematic content and the broader emotional context must be implemented. This can lead to more accurate distress signal recognition and increase the timeliness and relevance of mental health interventions delivered via social media platforms. De Choudhury et al. (2013) emphasize the need of combining various data sources and analytical tools in order to better detect and analyze mental health expressions online.

Despite the useful insights gained from this study, few limitations must be addressed that influenced the study's findings. The main issue was the inability to obtain data directly from X. According to Calma, (2023), X's rigorous constraints on data extraction can significantly limit the scope and size of data available for analysis. This constraint may have resulted in the elimination of a wide range of opinions and attitudes about suicidal thoughts, thereby reducing the data' generalizability. Another drawback is the potential for overlap between the two datasets used in this study. While we used many datasets to capture a wide range of depression-related tweets, it is possible that the topics that were discovered just partially represent the specific characteristics of these datasets rather than a generalizable set of themes. To address this, we carried out a detailed analysis to guarantee that the selected themes were consistent and indicative of both datasets. However, additional studies with larger datasets is needed to validate the robustness and generalizability of these themes.

Another significant drawback was the usage of Orange software for text mining. While Orange is a robust text analytics tool, its ability to handle massive amounts of tweets is limited (Demsar, et al., 2013). However, this does not apply to Sentiment Analysis done on Orange. This constraint may have resulted in using additional resources like R studio to conduct an in-depth analysis which is more time-consuming. Furthermore, along with previous studies in the fields of text mining and sentiment analysis, this study encountered difficulties in effectively reading and categorizing the sentiment of social media posts (Cambria E. D., 2017). These obstacles included recognizing sophisticated language and slang, as well as dealing with misspellings and informal grammar. Despite rigorous efforts to resolve these concerns through careful thematic and sentiment analysis, there is always the risk of misinterpretation or oversimplification of sentiments. Preprocessing of the datasets does help with tackling this issue.

This study emphasizes the enormous potential and certain limitations of machine learning and sentiment analysis in understanding and managing suicidal ideation, particularly on online platforms such as X. Despite the obstacles and constraints of data extraction and data integrity, examining the language, sentiment, and engagement patterns in suicide tweets provides essential knowledge. It emphasizes the importance of building innovative, adaptable data extraction and analysis tools to improve the reliability of the findings. The study by Tsugawa et al. (2015) provides a foundational understanding of how X activity can be used to recognize depression. In their research, they analyzed various features of X activity, including tweet content, frequency, and interaction patterns, to identify users showing signs of depression. Their findings demonstrated that depressive states could be inferred from social media activity with a reasonable degree of accuracy (69%). While Tsugawa et al. addressed depression generically, our study focuses on the specific and acute issue of suicidal ideation within the depression dataset from X. Using comparable machine learning approaches, the research we conduct finds and analyses tweets expressing suicidal ideation, studying the relationship between numerous themes and attitudes. The results of this study show that, while correlations between themes and sentiments are often low, there are identifiable patterns that can help with the early detection of suicidal ideation. In essence, our study expands on Tsugawa et al.'s work by verifying the use of sentiment analysis in mental health interventions while also demonstrating its potential for treating more severe and acute mental health emergencies, such as suicide ideation. This connection emphasizes the importance of our research into improving mental health support mechanisms using modern analytical methodologies.

This study's findings have important implications for mental health research and practice, notably in detecting and preventing suicidal ideation on social media platforms. The identification of themes such as suicidal ideation and anxiety, life reflection and negative experiences, and depression and hopelessness provide a deeper understanding of the emotional states exhibited by users on social media. (Lin, 2014) While each topic focuses on a different component of emotional discomfort, they all agree that people who communicate their emotions on social media platforms need to be understood, supported, and treated effectively. Understanding the distinct features and interrelationships of these themes might help generate more targeted and effective measures for understanding suicidal ideation and promote mental wellbeing in online communities. Each of the themes revealed in this study offers various perspectives on comprehending suicidal ideation expressed on X. Each theme sheds light on different aspects of emotional discomfort, all of which point to the need for awareness, support, and effective interventions. Betton et al, (2015) corroborate this up by emphasizing the significance of social media platforms as both a medium for mental health discourse and a tool for offering support and intervention for suicidal ideation. The overlap and unique elements of these topics emphasize the interplay of numerous factors in mental health difficulties. According to Reavley and Pilkington (2014), personal, societal, and environmental factors all influence a person's impression of mental health. A thorough grasp of these complex influences is necessary to comprehend the emotional suffering that is conveyed through tweets. Furthermore, Chancellor and De Choudhury, (2020) found that the expression of mental distress on social media platforms is typically complicated and not always obvious. Their findings highlight the necessity of understanding the diverse and subtle language patterns employed in tweets conveying psychological distress.This can help lead future research toward developing complex algorithms for identifying emotional distress online. (De Choudhury M. G., 2013)

The substantial link between anxiety and suicide behaviour, as demonstrated by studies such as Nock., et al., (2009), highlights the need for more research into how different emotional states interact and contribute to suicidal ideation. The use of machine learning methods like VADER and SentiArt for sentiment analysis illustrates their capacity to identify nuanced patterns in big datasets, implying that these technologies can be used for various other contexts. (Hutto & Gilbert, 2014) While this study provides a snapshot of emotional expressions on X, longitudinal studies are necessary to examine changes over time, providing insights into the evolution of suicide ideation and intervention effectiveness. (Franklin, 2017) According to Niederkrotenthaler, et al. (2010), there are studies that emphasize the role of social media in early detection and intervention, as well as the importance of training therapists to use social media data in their practice, increasing evaluation and intervention procedures. Creating automated methods to flag high-risk content and linking users with mental health professionals is also important, and something to consider for future research. (Metzler, 2022) This study also demonstrates the potential of machine learning and sentiment analysis to address suicidal ideation on social media, paving the way for better detection and support for at-risk individuals. (Chancellor & De Choudhury, 2020) The findings of this study could have a substantial impact on early diagnosis and prevention of suicide, ultimately leading to more effective mental health support systems.

References

Calma, J. (2023). *Twitter just closed the book on Academic Research.* Retrieved from The Verge: https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research

Abdukhamidov, E. J.-S. (2022). Sentiment analysis of users' reactions on social media during the pandemic. . *Electronics*, 11(10), 1648.

Arendt, F., Scherr, S., Niederkrotenthaler, T., & Till, B. (2018). The role of language in suicide reporting: Investigating the influence of problematic suicide referents. *Social Science & Medicine*, 208, 165-171.

Barlow, D. H. (2004). Anxiety and its disorders: The nature and treatment of anxiety and panic. . *Guilford press.*

Betton, V., Borschmann, R., Docherty, M., Coleman, S., & Brown, M. &. (2015). The role of social media in reducing stigma and discrimination. *The British Journal of Psychiatry*, 206(6), 443-444.

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

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Bryan, C. J., Rudd, M. D., Wertenberger, E., Etienne, N., Ray-Sannerud, B. N., & Morrow, C. E.-M. (2014). Improving the detection and prediction of suicidal behavior among military personnel by measuring suicidal beliefs: An evaluation of the Suicide Cognitions Scale. *Journal of affective disorders*, 159, 15-22.

Burnap, P., & al., e. (2018). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making. *Policy & Internet,*, 10(2), 216-235.

Calvo, R. A. (2017). Natural language processing in mental health applications using non-clinical texts. . *Natural Language Engineering*, , 23(5), 649-685.

Cambria, E. D. (2017). Affective computing and sentiment analysis. . *A practical guide to sentiment analysis*, 1-10. .

Cambria, E., Schuller, B., & Xia, Y. &. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28(2), 15-21.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. . *Neurocomputing*, 72(7-9), 1775-1781.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. . *NPJ digital medicine*, 3(1), 43.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. . *In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* , (pp. 51-60).

Crane, C. B. (2007). Reflection, brooding, and suicidality: A preliminary study of different types of rumination in individuals with a history of major depression. . *British Journal of Clinical Psychology*, 46(4), 497-504.

De Choudhury, M. G. (2013). Predicting depression via social media. . *In Proceedings of the international AAAI conference on web and social media*, (Vol. 7, No. 1, pp. 128-137).

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).

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. *In Proceedings of the 2016 CHI conference on human factors in computing systems*, (pp. 2098-2110).

Demsar, J., Curk, T., Erjavec, A., Gorup, C., T, H., Milutinovic, M., . . . Zupan, B. (2013). Orange: Data Mining Toolbox in Python.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1), 61-84.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54. doi:https://doi.org/10.18637/jss.v025.i05

Fellows, I. (2014). Wordcloud: Word clouds. *R package version 2.5.*

Franklin, J. C. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. . *Psychological bulletin*, 143(2), 187.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1), 5228-5235.

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. . *Current Opinion in Behavioral Sciences*, 18, 43-49.

Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. *Technologies*, 10(3), 57.

Hawton, K., Bergen, H., Cooper, J., Turnbull, P., Waters, K., Ness, J., & Kapur, N. (2015). Suicide following self-harm: findings from the multicentre study of self-harm in England, 2000–2012. *Journal of affective disorders*, 175, 147-151.

Hawton, K., i Comabella, C. C., & Haw, C. &. (2013). Risk factors for suicide in individuals with depression: a systematic review. *Journal of affective disorders*, 147(1-3), 17-28.

Hawton, K., Saunders, K. E., & O'Connor, R. C. (2012). Self-harm and suicide in adolescents. *The lancet*, 379(9834), 2373-2382.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation.* , 9(8), 1735-1780.

Hollander, J. G.-K. (2016). A (Short) History of Social Media Sentiment Analysis. 15-25.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *In Proceedings of the international AAAI conference on web and social media* , (Vol. 8, No. 1, pp. 216-225).

Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & & Argyle, T. (2014). Tracking suicide risk factors through Twitter in the US. *Crisis*.

Kumar, E., Rao, K., Nayak, S., & Chandra, R. (2020). Suicidal ideation prediction in twitter data using machine learning techniques. *Journal of Interdisciplinary Mathematics*, 23, 117 - 125.

Lahiry, S., Choudhury, S., Chatterjee, S., & Hazra, A. (2019). Impact of social media on academic performance and interpersonal relation: A cross-sectional study among students at a tertiary medical center in East India. . *Journal of education and health promotion*, 8.

Lin, H. J. (2014). User-level psychological stress detection from social media using deep neural network. . *In Proceedings of the 22nd ACM international conference on Multimedia*, (pp. 507-516).

Liu, B. (2022). *Sentiment analysis and opinion mining.* Springer Nature.

Melton, C. W.-N. (2022). Fine-tuned Sentiment Analysis of COVID-19 Vaccine–Related Social Media Data: Comparative Study. . *Journal of Medical Internet Research.* , 24.

Metzler, H. B. (2022). Detecting potentially harmful and protective suicide-related content on twitter: machine learning approach. *Journal of medical internet research.*, 24(8), e34705.

Mikolov, T., Chen, K., & Corrado, G. D. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

Moreno-Ortiz, A. P.-H.-G. (2022). The language of happiness in self-reported descriptions of happy moments: Words, concepts, and entities. . *Humanities and social sciences communications*, 9(1), 1-13.

Nassar, R., Helmy, A., & Ramadan, N. (2022). *Multi Labeled Depression Corpus of 57,000 English tweets (Version V1) [Dataset]*. Retrieved from Harvard Dataverse: https://doi.org/10.7910/DVN/VK2D4R

Niederkrotenthaler, T. V. (2010). Role of media reports in completed and prevented suicide: Werther v. Papageno effects. *The British Journal of Psychiatry*, 197(3), 234-243.

Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2010). Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. *Molecular psychiatry*, 15(8), 868-876.

Nock, M. K., Hwang, I., Sampson, N., Kessler, R. C., Angermeyer, M., & Beautrais, A. .. (2009). Cross-national analysis of the associations among mental disorders and suicidal behavior: findings from the WHO World Mental Health Surveys. . *PLoS medicine*, 6(8), e1000123.

O'dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & & Christensen, H. (2015). Detecting suicidality on Twitter. *Internet Interventions*, 2(2), 183-188.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2), 1-135.

Pisani, A. R., Murrie, D. C., & Silverman, M. M. (2016). Reformulating suicide risk formulation: from prediction to prevention. *Academic Psychiatry,*, 40, 623-629.

Reavley, N. J., & Pilkington, P. D. (2014). Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. *PeerJ*, 2, e647.

Rogers, M. K.-R. (2017). Negative emotions in veterans relate to suicide risk through feelings of perceived burdensomeness and thwarted belongingness. *Journal of affective disorders.* , 208, 15-21.

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., & Kaminsky, Z. A. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. . *NPJ digital medicine,*, 3(1), 1-12.

Ruth Lynfield, M. D., Lori Freeman, M. B., & Scott Becker, M. S. (2021). Symptoms of Depression, Anxiety, Post-Traumatic Stress Disorder, and Suicidal Ideation Among State, Tribal, Local, and Territorial Public Health Workers During the COVID-19 Pandemic—United States.

Sakthi, U., Chen, T., & Sathiyanarayanan, M. (2023). CyberHelp: Sentiment Analysis on Social Media Data Using Deep Belief Network to Predict Suicidal Ideation of Students. *International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 206-211.

Schweinberger, M. (2023). *Topic Modeling with R. Brisbane: The University of Queensland.* . Retrieved from https://slcladal.github.io/topicmodels.html (Version 2023.04.05).

Sher, L. (2004). Preventing suicide. *Qjm*, 97(10), 677-680.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, 1(3), 37.

Stefanis, C. G. (2023). Sentiment analysis of epidemiological surveillance reports on COVID-19 in Greece using machine learning models. . *Frontiers in Public Health*, 11, 1191730.

Surana, P., Yusuf, M., & Singh, S. (2022). *ADAM-SDMH: A DAtaset from Manipal for Severity Detection in Tweets related to Mental Health [Data set].* . Retrieved from figshare: https://doi.org/10.6084/m9.figshare.19029656.v2

Tsugawa, S. K. (2015). Recognizing depression from twitter activity. . *In Proceedings of the 33rd annual ACM conference on human factors in computing systems* , (pp. 3187-3196).

Wang, W. H. (2016). Twitter analysis: Studying US weekly trends in work stress and emotion. . *Applied Psychology*, 65(2), 355-378.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *In Proceedings of human language technology conference and conference on empirical methods in natural language processing*, (pp. 347-354).

World Health Organization. (2020). Retrieved from Suicide: https://www.who.int/news-room/fact-sheets/detail/suicide

Yadav, A. &. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review.*, 53(6), 4335-4385.

Zhang, J., & Li, Z. (2013). The association between depression and suicide when hopelessness is controlled for. *Comprehensive psychiatry*, 54(7), 790-796.

Zhang, Y., Sun, M., Ren, Y., & Shen, J. (2020). Sentiment analysis of Sina Weibo users under the impact of super typhoon Lekima using natural language processing tools: A multi-tags case study. . *Procedia Computer Science*, 174, 478-490.