Research Proposal: Classifying the Causes of Depression from Social Media Tweets

Thesis Proposal

Department of Applied Computing Technology

Unitec - Te Pukenga

Ayushi Thakur (1580598)

Supervisor: Dr. Mohsen Pashen

October 16, 2024

Abstract

In this proposal, we propose a machine learning-based approach to identify and classify Depression caused by tweets on social media. This paper endeavors to sort four classes as Depression, Depression with Domestic Violence, Depression with Stress, and Depression with Gender Inequality by employing two powerful machine learning models of XGBoost and SVM and preprocessing the data (cleaning the text, tokenizing it, and extracting Features using a few TF-IDF to make it more effective for classification. To improve generalization, the model is evaluated using cross-validation and hyperparameter tuning. Such an approach offers a potent opportunity to illuminate the root origins of Depression and would be helpful for future-directed approaches. However, there are also difficulties to tackle, such as the need for more sufficient data, potential model bias, and the requirement of a vast number of labeled data for proper classification. This research will contribute essential knowledge to mental health by improving cause-specific depression detection by overcoming these limitations.

KEY WORDS: social media, Causes of Depression,XGboost,SVM,Machine Learning

Chapter 1

Introduction

Over the past few years,mental health problems, like depression, have been getting much focus, particularly in terms of how common they are and their effects on people and communities. Depression is a serious issue worldwide and has been recognized as one of the primary reasons for disability globally by the WHO. Traditionally, mental health assessments relied on face-to-face discussions with professionals, interrogation interviews, and self-report questionnaires about oneself and their society. However, the advent of digital communication and social media platforms like Twitter, in particular, also opened doors for locating signs of depression within user-written data. The development of research on depression in social media has thus given birth to an increasing number of studies that focus on the intersection of mental health and technology, specifically designed to examine and identify depressive symptoms represented among various social media platforms. Social media offers a plethora of data created by users who openly share their emotions and feelings on different platforms such as Twitter, with its large user base being a resource for this information. The platform is often used for sharing personal stories, daily challenges, and discussions about mental health issues such as depression. By examining this information, researchers can identify trends correlating to degrees of depression, classify depression, and Detect depression early through feature extraction using NLP and hybrid ML algorithms[43]. In this paper, we apply NLP and ML techniques to classify the depression caused by tweets automatically. It also provides hope for the diagnosis of best triggers–home violence, academic stress, social insecurities, etc.–as well as expectations of how these causes may alter over time.

The application of NLP and machine learning ML in the examination of social media data has represented a fast-growing area of study in recent years[57]. These methods enable the extraction of insights from data to identify symptoms of depression and categorize their root triggers. While previous studies have primarily concentrated on detecting depression and analyzing undertones in social media content, there has been some limited exploration into categorizing the causes of depression based on posts shared online. These research endeavors to bridge this gap by developing a model for categorization.

However, this research does come with its share of challenges. Social media information can sometimes be messy and disorganized, which poses a challenge in identifying attributes from it all. This is further complicated when trying to grasp the meaning behind unclear messages posted online, especially in places like Twitter. Moreover, there are concerns to take into account, like safeguarding user privacy and obtaining consent before delving into mental health details. Despite facing these obstacles, the significance of this research cannot be understated as it presents an opportunity to gain insights into health patterns and enhance assistance for individuals battling depression by delving into the root causes of the condition and tracking changes throughout time. However, this study faces obstacles, as social media information is frequently chaotic and disorganized, which makes it challenging to identify attributes accurately extracted features are complex to ascertain the context behind brief and uncertain messages, especially prominent in platforms like Twitter, which poses a challenge to ethical aspects also come into play like safeguarding user privacy and consent when examining personal psychological well-being data notwithstanding these hurdles the possible repercussions of this research are significant providing fresh insights, into mental health patterns and enhancing assistance mechanisms for individuals experiencing depression.

1.1 Inspiration and Motivation

The motivation behind this study stems from the growing recognition of health concerns and the substantial effects that depression can have on individuals and communities alike in New Zealand and beyond, like in numerous other nations where mental well-being is now a key focus due to the escalating instances of depression and suicide rates as highlighted by the Ministry of Health as a primary contributor to disabilities impacting individuals across various demographics. Social networking sites such as Twitter are progressively being utilized as platforms for individuals to openly express their challenges with wellness. These platforms offer researchers a chance to delve into the patterns and causes of depression by providing real-time insights into how people experience and express their struggles –an inaccessible resource for study.

In addition to this development in Machine Learning (ML) and Natural Language Processing (NLP) technology, resources have emerged that can assist in comprehending and evaluating quantities of unorganized information from platforms such as Twitter. The drive to implement these state-of-the-art innovations in an area like wellness stems from the hope of creating an impact. The goal of this study is to pinpoint the root reasons behind depression as seen on platforms in order to provide insights for healthcare professionals and policymakers to grasp the factors and patterns influencing episodes of depression.

I have personally witnessed the effects of health challenges on loved ones and members of our community here in Auckland. Sadly, many people feel alone and overwhelmed in their battles with depression. This has inspired me to delve into research that can enhance understanding and potentially improve lives by detecting depression triggers and offering intervention options.

Moreover, the ability to remain unidentified and the unrestricted environment of media allow individuals to convey their emotions and personal encounters in manners they may not feel comfortable doing in therapeutic environments. This presents an opportunity to identify signs of depression and grasp the underlying reasons within scenarios. The task of sorting through unorganized information to discover trends resonates with my enthusiasm for technology and advocating for mental health awareness. The objective is to create a scalable, easily comprehensible system that can precisely detect the elements causing depression. This could assist health professionals in intervening and with greater effectiveness.

This study is driven by a goal to connect technology and mental health. Two fields are usually perceived as entities in society today. This study uses NLP and ML capabilities to show how technology can influence the comprehension and tackling of health challenges such as depression.

1.2 Objectives

The primary objective of this research is:

- Develop machine learning algorithms, like XGBoost and SVM for recognizing and categorizing factors that contribute towards depression based on the content of users' tweets including issues like violence and academic pressure along, with social disparities.

- Prepare the text data by preprocessing it to ensure it is clean and well organized before analysing it for the model's input accuracy and reliability.

- Utilize Natural Language Processing (TF-IDF, Word Embedding) to identify aspects of the text and improve the models capability to categorize factors contributing to depression.
- Utilize machine learning algorithms to categorize factors contributing to depression alongside conducting sentiment and emotion assessments for a grasp of depressive conditions.
- The goals of these objectives are to offer a method, for examining and comprehending the elements that play a role, in depression as shared on social platforms.

1.3 Questions and Challenges

During The research, the following questions will be explored.

- How well can machine learning tools such, as XGBoost and SVM categorize tweets about depression into groups based on causes, like domestic Violence or stress related problems including gender disparities?
- Which method is more effective, for predicting the underlying Causes for depression using Twitter information; TF-IDF or word embeddings (Word2Vec/GloVe)?
- How much can categorizing real time tweets into depression related groups help us understand the mental health trends, on social platforms?
- What ethical considerations arise in the use of Twitter data for classifying depression-related causes, and how can they be addressed?

1.4 Challenges

While working on these questions, we will need to crack these challenges. Twitter data is unstructured and primarily noisy, smattered with abbreviations, emojis, hashtags, and URLs. Misspellings, code-switching (mixing two or more languages), short-form text, and emojis can make preprocessing too complex. There can be specific reasons for depression, which may be less represented in the dataset. Therefore, classes are imbalanced, and when we train a model, it will be biased toward majority dominant categories. Advanced models such as ML (XGBoost,SVM) may provide high accuracy but are more challenging to interpret since advanced models can be black-box models that do not show why specific predictions are made. One of the most challenging situations comes when tweets have sarcasm, irony, or emotional complexity in expressions.

These are complex signals for models to detect, and hence, mis-classification occurs. Even with TF-IDF or word embeddings, the feature sizes for these models become relatively more significant. If the feature spaces are large, this can tremendously increase computation time and make it more prone to overfitting. Temporal analysis to track changes in causes of depression over time requires aligning these tweets with real-world events and pinpointing trends across long stretches. Analyzing UGC that relates to sensitive topics like depression is an ethical gray area, as this type of research uses publicly accessible but very personal data. Although Twitter may readily crowdsource the references, it could not be transferable to other platforms with distinct user demographics and communication styles. However, training deep learning models in complex architectures like Transformers on large datasets is computationally expensive and time-consuming. The problem with using either of these datasets is that, as we already discussed, labeling tweets for supervised learning is very labor intensive, especially when you delve into searching for tweets that highlight subtle causes of depression.

1.5 Proposal Structure

This proposal report has three sections: an introduction, objectives, and research questions were briefly discussed in Section 1. In section 2, relevant research on Social media Tweets and computational approaches to Depression detection, Domestic Violence detection, and stress detection are comprehensively discussed. Section 2 ends with a research gap. The proposed research methodology, including the datasets, analysis, and evaluation of existing algorithms, the design of the proposed architecture, and objective evaluation criteria, are all formed in section 3. Finally, a timeline for the thesis is presented.

Chapter 2

Literature Review

As social media grows in popularity,numerous research have looked at and forecast real-world activities using social platform, including user sentiment analysis [7], opinion mining during political campaigns[85, 58], natural disasters [72, 19], epidemic surveillance [66, 1], event detection [62, 14], e-healthcare services [79, 77], and so forth. There are 200 billion tweets shared annually and 500 million tweets shared daily on Twitter, a significant part of many people's lives [1]. However, other research focused on security and privacy concerns[90, 80, 48, 89] because the growing sophistication of social media data seriously threatens users' privacy. Unlike When opinions are immediately shared online with a status update, Twitter turns into a reliable source of data for statistical surveys. Which enables academics to identify real-world occurrences in information retrieval and decision-making on an effective platform. Sentiment analysis has been examined on several platforms, including blogs and forums, and is currently being examined in social media [62].By taking out the comment on the characteristics and aspects of a specific aspect, author [51] have performed sentiment analysis and classified the comments as Positive, Negative, or neutral. A product, event, person, or subject can all be considered aspects. According to opinion mining on political occasions, Twitter is an effective way to convey opinions about events. Studies by author [85] and Tumasjan [72] found a strong correlation among voters' political choices and emotive analysis of tweets, which closely matched the election outcomes. Public tweets have been quite influential not just in politics but also in the economics sector. Tweet sentiments directly correlate with stock trends and can be used to predict them, according to an analysis by Bollen et al. [14][14]. Bruns and colleagues (2015), Burns and colleagues (2016), and Gaffney and colleagues (2017) noted that A powerful tool for collecting public views and enacting social change is Twitter. According to Sakaki et al. [72]research, Twitter can detect earthquakes and alert people to impending danger. To determine the probability of an earthquake, they employed the time and geolocation information of every Twitter user in Japan as a mobile sensor. Kalman and particle filters were used to estimate earthquake sites by treating posting time and volume as an exponential distribution. Their investigation proved that an earthquake can be felt before it is officially announced. Culotta et al. [23, 20] used Twitter analysis to find influenza epidemic outbreaks, which is faster and less expensive than established techniques. Compared to queries, users data like their location, age , and gender can be used to provide more detailed information on demographic insights. Quincey et al.[27] used predetermined keywords and the term co-occurrence approach to identify swine flu from Twitter, whereas they used multiple regression models to diagnose influenza. These methods are analyzed using specific keywords to search tweets to identify any unusual changes in the quick flow of message traffic relevant to the specified keywords. This technique helps to gather more targeted data from the Twitter feed. Twitter has proven to be a valuable tool for doing research on health-related topics and analyzing a wide range of disorders, including depression [2], mood swings [33], the Ebola epidemic [60], alcoholism [24], tobacco use [21], dental discomfort [34], and cholera [19]. To track the public's health care on diseases, symptoms, and treatments, authors suggested a method called the Ailment Topic Aspect Model [65, 64, 63]. Conventional methods Topic Detection and Tracking has discussed how to find real-world events using traditional media sources, such as news articles, research papers, and so forth [73] of event detection. Because of its cacophony and small size, event identification on Twitter is challenging. The interactive topic browsing system developed by Bernstein et al. [9] uses tag clouds for visualization and groups Twitter users into subjects. Sankaranarayanan et al. [74] created TwitterStand, which allows users to peruse news according to their preferred location. The feature sets are extracted using frequency-based online clustering, and topics are identified. Another comparable method is TwitInfo [53], which computes peaks based on tweet frequency and offers users sentiment analysis support. Bursty topic detection, which identifies more prevalent trends in the time series, is an additional method for Twitter trend analysis. A sliding window model that takes into account the term frequency's time factor was created by Lee et al. [46]. It uses the arrival rates within a given time range to compute the term weights. The term's burstiness is employed in the specific event detection to identify its popularity at the specified time interval. Topic modeling is one of the most often used study areas to determine the semantic relationship from text content. The problem of topic detection is inferred as a probabilistic distribution by Blei et al. [11] using their probabilistic topic model. It depicts a topic as a distribution of words and the document as a distribution over themes. To identify events in Twitter streams, online LDA [5], dynamic topic models [10], and labeled LDA [67] employed probabilistic topic models as a baseline model. However, these approaches might be more appropriate given the unique features of Twitter streams, which include their brief duration, loud, informal content, spelling errors, and rapid arrival rate. To extract the topics, LDA models are used. They train various classifiers to predict the model's performance. [37].

Most topics use term-based models, making it hard for users to understand better and anticipate. Synonyms and polysemy plague this. One word might have several meanings when it is polysemic. Apple, for example, is a name that describes fruit and technology. The exact meaning of two distinct words is what makes a word synonym. For example, a financial institution and the bank river are called banks. Because patterns allow the meaning of a word to be appropriately communicated through the co-occurrence of neighboring phrases, they are, therefore, the optimal approach for subject modeling.

The typical method employed in traditional databases is called pattern mining. Apriori is a highly recognized and significant algorithm for mining association rules. When an item set's association rules surpass the minimum support and confidence levels, they are all found through association rule mining. Key components of data mining applications are Interest and Patterns. The interest factor establishes whether or not a pattern is valuable, while association rule mining and classification extract the patterns. The method of make data into useful information and then that information into value that users can utilize is another way to conceptualize data mining . It could be advantageous for the community, organization, and individuals to mine the actionable trends. Therefore, through analyzing intriguing patterns about a specific topic, actionable knowledge can be gleaned from the significant social media Twitter data stream .HUPC , EDCoW , Twitter Monitor , and SFPM [36] are a few of the strategies that have been studied recently. Bubbly keyword analysis forms the foundation of the first two methods. Finding more bursty terms involves using keywords with a higher absolute frequency than typical. With the EDCoW approach, terms are modeled using frequencies and wavelet analysis. Trends in Twitter data streams were identified by HUPC and SFPM using pattern mining techniques.

2.0.1 Depression From Social media

Author [25] determined a significant yet underreported mental health condition that affects a lot of women. Their focus was on employing Twitter posts to construct a predictive model regarding the upcoming impact of childbirth on the temperament and conduct of recently delivered women. Tweets from 376 moms were assessed using metrics like feelings, social media interactions, language style, and social media use to identify changes that occur after giving birth. [59] sought to determine whether the degree of concern for a tweets regarding suicide thoughts could be exclusively based on the information included in the tweet and using machine learning classifiers as well as human coders.

According to the authors of [92], people regularly use social media to interact with other people in their daily lives and emotions. Author work on Twitter messages to create a well-labeled depression data collection, and they mined six feature groups made up of clinical and online social behavior characteristics associated with depression. Feature groups were used to develop the multimodal depression dictionary learning model for detecting depressed individuals on Twitter. In [45], the writers carefully investigated the connection between mental health illness and social media posts. They connected depression and anxiety with erratic thought patterns, restlessness, and sleeplessness. They presented a real-time tweet prediction technique based on posting trends and linguistic clues for Dire forecasts of depression.

The word, emotion, execution, rate, comparison, and similarity make up the feature vector, and an ensemble voting classifier is used for majority vote. Using various ML methods, the authors of [52] categorized users' emotions into positive and negative categories. They then used deep learning techniques to compare the outcomes. The scientists concluded that there is a substantial difference in performance between the CNN-based deep learning model and the machine learning models. The Naive Bayes classifier Work in [69] to categorize Twitter tweets into train and test datasets on depression.

As social media use increases, reseachers are talking about depression—and specially , suicide —on social media, according to an observation made by the authors [54]. The researchers examined multiple Twitter accounts and utilized many accounts and Twitter-related tools for finding profiles of suicide. An array of individuals was utilized to identify those who had previously taken their own lives and the method's efficacy. Additionally, current research in [47] highlights the places where the writers thought of using categorization exercises to identify depressive symptoms by responding to Reddit users' text messages. In addition to the TF-IDF, other baseline metrics were taken into consideration. More intricate characteristics include morphology, stylometrics, bigrams, and embedding.

The authors of [76] create a supervised model for prediction automatically identifying early signs of depression by utilizing the user's earlier Reddit writings, a standard bag of terms, surface features, and additional linguistic-related data. In [75], researchers developed a feature attention network comprising several feature networks, such as those for ruminative thought, depressive symptoms, feelings, and writing style. They applied deep learning to identify unhappy users. In [31], the writers looked into several techniques for early depression identification that use dual and singleton machine learning methodologies. A Random forest methods with two threshold functions is used in the singleton technique. However, two independent Random Forest classifiers are used in the dual approach, one for identifying users who are sad and another for identifying users who are not. For early symptom detection, the authors of [68] employ two distinct approaches: the first looks at the user's writing habits and time, while the second looks for hints in tweets and shared texts. The authors of [68] discussed the use of ML and deep learning LSTM-CNN techniques for early suicide identification in social media posts like Reddit . Using a mix of learning algorithms, they examined user posts on the internet one after the other over an extended period and produced an early detection.

2.1 Domestic Violence from Social media

One of the main reasons why women get Depressed is Domestic Violence (DV), and the issue is widespread around the world [35]. In addition to physical assault, the victims frequently experience verbal, emotional, and sexual abuse. Formal and informal assistance can significantly contribute to victims of abuse's increased safety and improved physical and emotional health consequences [71, 49]. In order to offer a variety of services, including counseling, a crisis hotline, and advocacy, as well as emergency housing for victims of domestic violence [8]. The Victims discovered that revealing their circumstances and getting emotional support is beneficial and enhances their psychology [81]. Nevertheless, because victims must actively seek out these assistance, they frequently do not use them appropriately. Due to obstacles arising from their socioeconomic and religious backgrounds, many victims have opted not to get in touch with DVCS groups and report their circumstances [44]. The concept of violence as a private matter has been called into question by the existence of social media [50]. By exposing stories to the public and spreading knowledge, social media has been utilized to reduce violence [20, 77]. Social media's advantages in information sharing have been applied to several situations, including crisis response, planning, and recovery [30] in the event of natural catastrophes such as earthquakes [38], tsunamis [2], and flooding [15]. Still, social media has not fully realized its potential to locate and instantly support victims of domestic violence who are in dire need.

Previous research has been done on identifying online Domestic Violence DV by using deep learning techniques. Previous research has used traditional machine learning models such as SVM and decision trees to identify distress signals, but these methods have scalability limitations when combined with feature extraction (Coppersmith et al., 2014). Similarly, deep learning models with GloVe word embeddings indicated that Recurrent Units (GRUs) have performed well in classifying DV-related posts, with up to 94% accuracy [78]. This method helps in better real-time detection and mitigation supported by context features learned from unstructured texts, which is a step ahead of traditional approaches. Nevertheless, constraints such as small data samples and platform dependence call for research opportunities, including real-time usage or more extensive social media coverage [78].

2.2 Stress and gender Inequality from social media

Numerous text-based methods for stress detection have been developed; among them, some have shown that the most successful method is BLSTM with mechanism [40]. The second study is centered on the BLSTM model and continuous lexical embedding [93]. Another study introduces the KC-Net, a contrastive and mental state knowledge-aware network that can be used for early stress detection [33]. A simple yet reliable technique for identifying depressing texts on social media was put out [34]. The authors employed a CNN-based model and a bidirectional LSTM based on attention. Achieving 0.75 accuracy, some research uses the SVM-based technique. Using distributional representations with lexicon-based characteristics can improve classification accuracy [40].

The other study offers a TensiStrength technique that uses a lexical approach and a set of rules to determine the text's stress or relaxation level . Some of the publications focus on textual data analysis from social media. To classify sentiment, the authors of Nijhawan et al. [56] use BERT, machine learning techniques, and massive Twitter datasets. They also employ an unsupervised machine learning technique called LDA to forecast relevant topics and identify document patterns. These algorithms streamline determining users' emotions on social media networks.

The authors of the other study examine how emotion recognition might be used to identify psychological stress, which can be explained [86].They study how to learn multitasking and modify language models with empathy. The readdit dataset was used to train their models, and the results are similar to those of BERT models. A different study used the readdit dataset [55].To prepare tweeter text data for machine learning models, they used a variety of NLP techniques, BERT tokenizers, and the Bag of Words technique. Results from each method are examined, with a 76% accuracy rate. Classifying the Dreaddit data was the work [39]. With the published LSTM, BLSTM, and VM algorithms, BLSTM's accuracy was the best. Still, they did not

mention the level of precision attained. Author [61] work on Dreddit and self label reddit post and use BOW, and logistic regression and achieve 77% accuracy and the author get success to predecit the acedmic stress like PHD student.
Social media is becoming an extremely effective instrument for promoting gender equality through campaigning, awareness-raising, and action. Social media platforms like as Twitter, Facebook, and Instagram have been essential in drawing attention to issues, rallying support, and affecting institutional and policy changes[16, 88, 28] This is a detailed examination of how social media marketing and Social media efforts have been crucial in increasing awareness of the movements that have led to gender equality. Concerns regarding gender equality, include unequal compensation, violence against women, and a lack of representation in leadership roles parts. For instance, the #MeToo movement, which got its start on Twitter, started a worldwide dialogue about sexual harassment and violence, resulting in heightened consciousness and demands for reform for change. The value of investing in women and girls has been made increasingly evident by social media. Governments and companies have been encouraged to engage in programs and projects that help women's economic empowerment through campaigns such as #InvestInWomen. Social networking has moreover given female entrepreneurs a venue to obtain capital and introducing them to possible backers and investors, as well as investment prospects. For instance, consider the Using social media, the #BringBackOurGirls campaign spread awareness of Nigerian schoolgirl abductions. via the terrorist organization Boko Haram. The initiative attracted attention from around the world and exerted pressure on the Nigerian Action by the authorities to save the girls is needed [29, 70].
Below is the list of Depression detection using different ML Technique and Deep Learning method from Last five years.

Table 2.1: Related work on depression from 2019 to 2024

Year	**Methodology Used**	**Social media platform**	**Results/Remarks**
2019 [17]	SVM , SS-BED. NB and LSTM-SSWE with LSTM-GloVe.	Twitter, ISEAR, SemEval2007, WASSA 17SS	This method makes better use of sentiment and semantic-based elements to predict user moods from their speech.

Year	Methodology Used	Social media platform	Results/Remarks
2019 [82]	NLP-TF-IDF,LIWC Dictionary,LR,SVM,MLP and Random Forests	Reddit	Author got the accuracy of 91% from MLP and 80% accuracy from SVM considering LDA as feature extraction process.
2020 [42]	CNN and XGBoost	Reddit	Author have work on different mental health issues like anxiety bipolar and BPT and got the 96.69% accuracy from CNN and 89.01% from CNN they have used e TF-IDF vectorizer for feature extraction.
2022 [87]	LSTM, RNN with LDA and XAI using AI	Facebook, Norwegian dataset	Extracted data is actually based on young people only but got a good accuracy 98%.
2022[6]	Hybrid model (RNN, LSTM)	Twitter	The suggested approach outperforms frequency-based deep learning models in terms of accuracy, achieving 99.0%, while reducing the false positive rate. .

Year	Methodology Used	Social media platform	Results/Remarks
2023[12]	NLP, Xgboost, TF-IDF	Reddit	author mention posts of depression and suicide of reddit by using TF-IDF for feature extraction and XGBoost help to increase the accuracy of the modal .
2024[84]	CNN, LSTM, BiLSTM	CLEF2017 dataset	Models Give 90% Acc, but author have still doubt of and suggested to perform on a large dataset
2024[83]	Fasttext, CNN, LSTM (FCL)	Reddit, Twitter	FCL model is used to achieved an accuracy of 88% for twitter data, LSTM-RNN is used to achieve the good results and predection.

2.3 Research Gap

There are still few studies that attempt to predict specific types of depressive causes and even less into mental health problems using social media data. Many current studies concentrate on depression as one problem that does not contain differentiating factors such as domestic violence, stress, or being part of gender inequality. Moreover, existing methods mainly support only binary classification in depression detection and hardly any multi-class classification that can identify the reason for depression. There is also an issue with the lack of labeled datasets that distinguish between different factors that trigger depression in individuals. The current datasets have limitations in capturing the subtleties of language used in media platforms, like slang terms and emojis, as well as abbreviations, which play a crucial role in comprehending these triggers. Moreover, there is a missed opportunity to leverage sentiment analysis and emotion assessment alongside machine learning models to address this issue.

While mental health and social media data have been receiving attention lately, there are research gaps when it comes to predicting the causes of depression. Existing studies often lump depression into one category without differentiating between factors such as violence or gender inequality that could contribute to it. Moreover, current models focus on classification for detecting depression and lack the sophistication needed for multi-class classification that could pinpoint specific causes.

This study seeks to address these deficiencies by creating a classification model utilizing advanced NLP and ML methods to pinpoint the factors that contribute to depression based on the data from social media platforms and enhance the accuracy of mental health forecasts by providing more comprehensive insights.

Chapter 3

Methodology

In this proposal, we introduced a system that helps us to utilize multi-class classification of Tweets to predict specific causes of depression. The classification is mainly based on four classes: Depression(If the Cause is not from one of them), Depression with Domestic Violence, Depression with Stress, and Depression with Gender Inequality. To put it into practice, we must gather information from multiple sources or Twitter datasets. We use XGBoost and SVM (Support Vector Machine) as the main ML models to build a system that can accurately predict what causes depression based on available textual content from Twitter. This approach uses NLP techniques like word embedding(word2Glove), BOW, and TF-IDF to perform feature extraction and machine learning techniques for better prediction accuracy. The following are detailed step-by-step instructions to help better understand the techniques applied.
The first and foremost step is identifying four classes in which tweets can be divided and representing different reasons for depression. Depression is often triggered by various life circumstances such as domestic violence, stress, gender discrimination, study stress, failing in life, loss of loved ones, and many more.Unlike binary classification, which predicts whether depression will occur or not, we require a more nuanced method to comprehend the underlying reasons of this issue.
The research thus bridges a significant gap in the existing literature, which typically classifies depression based on binary categories.

3.1 Data Collection

The initial phase includes assembling a data collection encompassing tweets discussing Depression and its possible triggers. The dataset had to be created from scratch because there is no publicly accessible dataset for the Causes of Depression in tweets. We can use datasets, such as the Sentiment140 dataset and the Twitter dataset from

Keagle, which were already classified as Depression and non-depression Tweets, and then manually remove the Non-depression Tweet, and classify the dataset into four different Causes using some lexicon based on the previous study and also o verify the lexicons with the psychiatrist, filter relevant hashtags and keywords.

- General depression: #depression, #mentalhealth, #sadness
- Domestic violence: #domesticviolence, #abuse, #survivor
- Stress: #stress,#burnout, #anxiety
- Gender inequality: #genderinequality, #womensrights, #metoo

The open-source platform Apify was selected for the scraping task. Each of the above lexicons was entered into the search field one at a time, and the tweets were subsequently recorded in CSV format.

3.2 Dataset Annotation

The process of adding labels or metadata to a dataset in order to facilitate its usage and analysis in machine learning and other data-driven methods is known as dataset annotation. In this research study, the scraped tweets underwent manual annotation. The tweets were thoughtfully categorized, considering the context. Tweets indicating did not have a diagnosis of depression until and unless it included the term abusive, stress-related, or any other kind of depression, the situation and the context matched as well with general depression. A person with depression or any other type of depression has been called depressed. Consequently, it was made sure that the prepared data had excellent quality. Consequently, it can be applied to another study.

3.3 Data Preprocessing

Causes of Depression identification by tweets require the data preprocessing stage. The Twitter data must be cleaned and organized to prepare for model training. This is known as preprocessing. Data preparation also involves removing irrelevant information and noise from the tweets. The hashtags and URLs Were eliminated, as well as the tweets that used foul language or profanity. Thus, high-quality data and characteristics are extracted in this step by utilizing a few preprocessing approaches. Twitter stream preprocessing eliminates noise that impedes performance and has adverse impacts. Because there is more noise in the tweets at this stage, it is the most crucial in micro blogs. Stop word lists include frequently used English terms such as pronouns,

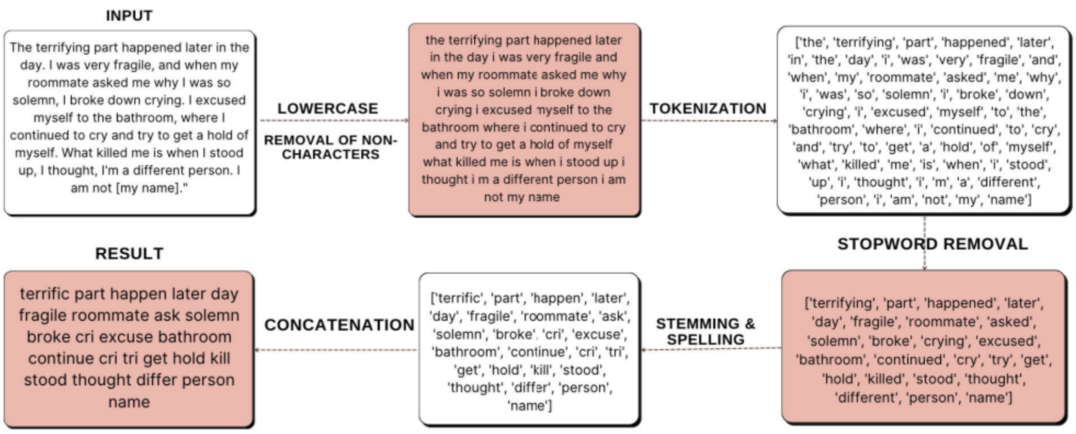

Figure 3.1: Preprocessing Steps,[61]

articles, and prepositions. The, a, an, the, in, at, and so forth are examples. Some steps need to be followed to clean the Tweets.

- Removing Special Characters and Punctuation: Remove hashtags, mentions (@user, #), URLs, and Applicable special characters.
- Lowercasing: Any text case is converted to lowercase for better uniform formatting.
- Stop Words Removal: Remove the most common insignificant stopwords when differentiating between the like and the is classes.
- Tokenization: This is when we split the text into individual words (or tokens)
- Stemming/Lemmatization - the heuristic process of reducing different word forms to a core form may not always work, such as "running" is not mapped to "run."
- To tackle balanced data: If an imbalance exists among the four classes in the dataset, over sample the minority classes or under sample the majority class.

3.4 Dataset labeling and Transformation

We accept that most of the dataset includes tweets with labels indicating positive or negative sentiment. Using a negative sentiment analysis of the tweets, we deduced the existence of depression in order to modify this dataset for depression identification. Research indicates that disparaging remarks or language usage may be a sign of a depressive propensity [26, 22]. The algorithm may be able to detect early warning

signals or indicators of depression by keeping an eye on language trends in social media, such as those seen in the Sentiment140 dataset [91]. In particular, we used a unique algorithm that examined each tweet's text.
Calculating Average Polarity: A sentence or word's polarity score, which ranges from -1 to +1, indicates how positive or negative it is sentiment. Among tweets categorized as negative emotion, we calculated the average polarity score for a collection of terms that are substantially connected with depression in order to differentiate between depressed and non-depressive tweets. These terms were chosen based on the theory put forth by Yazdavar et al. (2017)[91] that the system may be able to detect early warning signals or indicators of depression by keeping an eye on language patterns in social media, such as those recorded in the Sentiment140 dataset. This proved their applicability in identifying depression. Once the Negative content is detected, we can manually assign one of the four classes, or semi-automated labeling according to keyword matching can be performed. A tweet could be marked as 'Depression with Domestic Violence' if it contained words such as "violence," "abuse," or "assault." For the rest of the classes, based on types of keywords in our context

Tweets	**Class**
"Feeling really down today. #depression"	Depression
"He keeps hurting me, I can't take it"	Depression with Domestic Violence
"So much work stress, I'm breaking down"	Depression with Stress
"Still getting paid less as a woman..."	Depression with Gender Inequality

Table 3.1: Tweets Labeling based on Causes of Depression

3.5 Feature Extraction

Feature extraction converts unprocessed data into beneficial characteristics that may be applied to deep learning and natural language processing applications[13]. Stated differently, it comprises identifying and removing the appropriate information from the incoming data and presenting it correctly in a machine-learning prototype. Likewise, this procedure may involve multiple approaches, contingent upon the type of information and the specific natural Linguistic exercises required. In natural language processing tasks, feature extraction could use methods like bag-of-words or TF-IDF to represent the text input in a machine-readable form. Alternatively, any deep learning-based model, like BERT, could be used for feature extraction. BERT was employed in this study to extract features. BERT is a powerful feature extractor that can be used in NLP applications in addition to other methods like BoW, TF-IDF, and Glove. BERT is a linguistic model already taught, and rich text representations that change depending on context can be learned due to this skill.

Data input for various machine learning models must be in particular forms [41]. In order to convert the textual tweet data into numerical representations appropriate for particular models, we utilized the TF-IDF vectorization technique in this study. Every tweet is represented by TF-IDF as a numerical vector that takes into account the overall dataset as well as the importance of each individual word in the tweet. [4]. It determines each word's term frequency (TF) in a tweet and scales it according to the word's inverse document frequency (IDF) for the entire dataset. Models like SVM and XGBoost frequently use TF-IDF vectorization.

3.5.1 TF-IDF:

TF-IDF (Term frequency-inverse document frequency) is a statistical measure of how frequently a word appears in a text, representing a word's scarcity throughout the corpus, and term frequency (TF) is combined to evaluate a word's significance inside a document or set of documents. TF-IDF assigns larger weights to terms common in a text but uncommon in the corpus to capture their relative significance.

3.5.2 Word Embedding:

Word embedding is a characteristic of the learning approach that maps words to vectors using their contextual hierarchy; similar words will have similar feature vectors; for instance, the meaning of the words ”dog” and ”puppy” leads to tighter placements between them.

3.6 Model Selection:

One of the crucial steps is Finalising the Machine learning models that fit best to identify the causes of depression. In this research project, XGBoost and SVM (Support Vector Machine) were selected for their successful classification of big datasets.

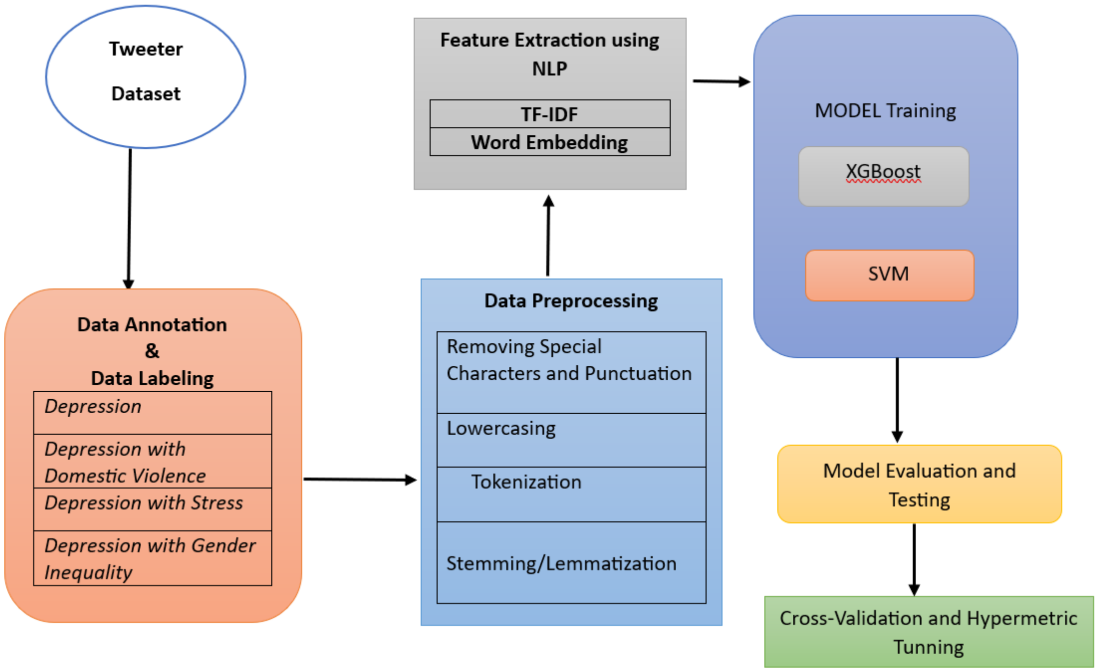

Figure 3.2: Proposed Architecture

3.6.1 XGBoost

The classifier for tree boosting is called XGBoost. This ensemble classifier used the decision tree concept to separate the targets by segmenting the data into smaller portions. For various NLP applications and machine learning models, this classifier aids in achieving good accuracy. Also, it is scalable in several scenarios, and the classifier is significantly quicker than other well-known classifiers [32]. XGBoost was chosen for its strength in bringing classification yields and applying them to structured and unstructured data. It is a group of weak learners making up to form a robust model. With the power of handling missing data, minimal overfitting, and gradient boosting features, XGBoost accelerates this multi-class classification problem.XGBoost algorithm has been widely used in text classification and has been proven to have state-of-the-art performance across various NLP applications. It also supports multi-class classification directly, allowing for effective categorization into our four targeted classes while incorporating regularization to minimize overfitting. In the field of machine learning, XGBoost has become very popular and widely used. Its broad acceptance can be ascribed to its competence in handling large datasets and its capacity to achieve cutting-edge performance throughout a range of machine learning applications, including as categorization and regression.[3]

3.6.2 SVM

The SVM is chosen for another secondary model as it is particularly effective with high-dimensional data, such as the multi-million element feature vectors created by TF-IDF. According to SVM, it tries to draw a hyper-plane that will distinguish one class from another. This classifier is reliable, especially when data/observations are linearly separable. SVM has been widely utilized in multi-class tasks, performing well on tasks like text classification and sentiment analysis. In order to maximize the separation distance from the hyperplane, it first uses a hyperplane to divide the various classes. The inaccuracy that the classifier produces decreases with increasing distance. The performances of the two models were trained and compared to choose the best classifiers for this task. The SVM is chosen for another secondary model as it is particularly effective with high-dimensional data, such as the multi-million element feature vectors created by TF-IDF. According to SVM, it tries to draw a hyper-plane that will distinguish one class from another. This classifier is reliable, especially when data/observations are linearly separable. SVM has been widely utilized in multi-class tasks, performing well on tasks like text classification and sentiment analysis [18]. In order to maximize the separation distance from the hyperplane, it first uses a hyperplane to divide the various classes. The inaccuracy that the classifier produces decreases with increasing distance. The performances of the two models were trained and compared to choose the best classifiers for this task.

3.7 Model Training and Evaluation

After preprocessing and extracting features, the dataset is divided for training and testing. The training data is used to train both the XGBoost and SVM models, which are then evaluated on the test data with standard evaluation metrics: accuracy, precision, recall, and F1-score. A confusion matrix will be used to understand how well the model can differentiate between all four classes[57].

- Precision (P): P stands for the anticipated positive post, and it is calculated as the ratio of true positives (TP) to the total of false positives (FP) and TP.

$$P = \frac{TP}{TP + FP}$$

- Recall (R): R is the sum of the false negative (FN) and positive prediction percentages, the ratio of TP, and the total number of adequately predicted positive postings.

$$R = \frac{TP}{TP + FN}$$

- Weighted average F1-score (F): A weighted average of P and R is used to calculate F. Compared to the A, R, and P, the evaluation metric F is more significant.

$$F_1 = 2 \times \frac{P \cdot R}{P + R}$$

- Accuracy (A): Using TP, FP, TN (true negative), and FN, A calculates the percentage of correctly classified postings.

$$A = \frac{TP + TN}{TP + FP + TN + FN}$$

In addition to the aforementioned conventional criteria, performance is also evaluated using other metrics like specificity, false positive rate, and false negative rate. The primary evaluation of the accuracy metric is how well In comparison to earlier research, the expected study performs. Notes that follow indicate the parameters that the metrics utilized

1. True positive
2. False Positive
3. True Negative
4. False Negative

3.8 Cross-validation and Hyperparameter Tuning

All experiments were carried out utilizing the N-fold Cross Validation (CV) approach, as Fig. 2 shows. CV is usually employed in regression and classification models to lessen bias between training and testing sets and the full dataset [57], as well as to prevent overfitting [58]. Data is divided into n disjoint folds, or partitions, in CV. The remaining fold is utilized for testing, and n-1 folds, or subsamples, are used for training. For the purposes of this study, we used n = 10, meaning that 10-fold CV was used in every trial. The findings of each process were averaged to determine the overall measures.

To enhance the more efficiency of the models used in validation and hyper-parameter tuning processes are utilized effectively by employing techniques such, as GridSearchCV or RandomizedSearchCV to systematically investigate various hyperparameter setups and determine the optimal configurations, for our models.

GridSearchCV thoroughly explores a given range of hyperparameter values whereas RandomizedSearchCV randomly selects parameter combinations, from a distribution pattern. In both methods cross validation is used by dividing the dataset into subsets for training and validation to assess the models performance, under setups and avoid overfitting by ensuring the selected parameters can generalize effectively to new data. In the end Fine tuning the hyperparameters guarantees that the models are optimized effectively improving prediction accuracy and steering clear of overfitting or concerns.

Chapter 4

Limitaions and Conclusion

4.1 Limitation

The suggested approach for categorizing the causes behind depression based on tweets shows potential. It comes with limitations to consider. Primarily, The challenges of Difficulty Data availability and annotation quality in social media. Public datasets frequently lack the tags required to pinpoint classifications of depression causes like those linked to domestic conflicts or pressure. This could require tagging processes. Social media information could lead to bias in the model because the people discussing depression might not accurately reflect the population, which could affect how well the model can apply to groups or platforms. Moreover, short and unclear tweets can also be a problem since they often need more context, making it easier for the model to understand correctly, especially when the language is casual or includes slang tones or references related to cultures. Additionally, The model is based on only four classifications: Depression, Depression with domestic violence, Depression from stress, and Depression from gender inequality. However, depression can be caused by a wide range of other factors not covered in this study, potentially limiting the breadth that this particular model can capture and thereby overlooking critically essential facets of mental health. The final hurdle is that training machine learning models, especially for large datasets and hyperparameter tuning, can be a resource-intensive process requiring dedicated computational power, and significant time may limit the system's scalability for real-time applications.

4.2 Conclusion

This proposal seeks to improve our knowledge of depression and its root causes by categorizing tweets into four groups; Depression Depression, along with Domestic violence;

Depression with stress;. Depression associated with gender inequality issues on Twitter platform. Through the utilization of sophisticated machine learning algorithms like XGBoost and SVM along with NLP methods such as TF IDF and word embeddings this study offers a method for pinpointing factors that may trigger depression by analysing textual content, from Twitter posts.

By utilizing a real time forecasting system, in health and natural language processing (NLP) research studies have found benefits; aiding academic progress and offering valuable tools for early intervention strategies in practice as well It enables mental health practitioners' policymakers and support groups to grasp the unique obstacles individuals encounter by instantly categorizing tweets This fosters more precise and prompt assistance for those, in need

To sum up this research contributes to the progress of identifying depression by going beyond categorizations and providing a nuanced view of the factors leading to depression as portrayed in media posts.T his approach sets the stage, for studies, on monitoring health and gives important understanding into the social aspects affecting mental health

Chapter 5

Timeline

The Following section is showing the Progress graph where in y axis we have Tasks and in x-axis we have Weeks in counting assuming that i am submitting my proposal in week 1st, and according to this graph Proposal submission is already done and after that the first and foremost step is data collection where I need to collect the data from different datasets and once the data is there, we need to do labelling of data and classifying the data in different classes which will take at least 6 weeks. Once the dataset is Prepared, we need to do the cleansing and ensuring the quality of data cleansing I need at least 3 weeks for cleansing which involved several steps and then the process of extracting features using Natural Language Processing techniques to pinpoint characteristics that will aid in creating models we are plan to complete feature extraction by the end of February 2025, and furthermore, we can start creating our model and moving to Calculating Results stage comes the task of assessing how well these models perform and studying the results they yield once everything including results , hypothesis and precision in our hand we can start writing our thesis from the mid of march 2025 and till June 2025 we can achieve our milestone to submit the thesis

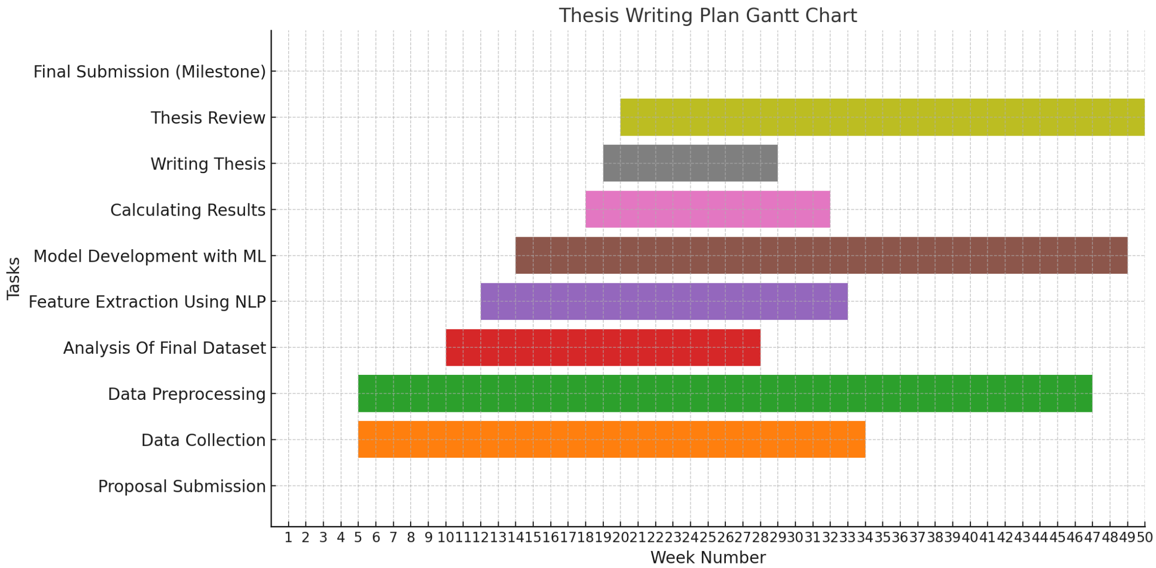

Figure 5.1: Project Timeline from October 2024 to June 2025.

Bibliography

[1] *Twitter statistics*, 2010.

[2] A. Acar and Y. Muraki, *Twitter for crisis communication: Lessons learned from japan' s tsunami disaster*, International Journal of Web Based Communities, 7 (2011), pp. 392–402.

[3] A. Agrawal and H. Mishra, *Enhancing depression detection in clinical interviews: Integration of fused bert-bilstm model and xgboost*, in Proceedings of the 2024 10th International Conference on Computing and Artificial Intelligence, 2024, pp. 174–179.

[4] M. I. Alfarizi, L. Syafaah, and M. Lestandy, *Emotional text classification using tf-idf (term frequency-inverse document frequency) and lstm (long short-term memory)*, JUITA: Jurnal Informatika, 10 (2022), pp. 225–232.

[5] L. AlSumait, D. Barbar, and C. Domeniconi, *On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking*, in Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), IEEE, 2008, pp. 3–12.

[6] A. Amanat, M. Rizwan, A. R. Javed, M. Abdelhaq, R. Alsaqour, S. Pandya, and M. Uddin, *Deep learning for depression detection from textual data*, Electronics, 11 (2022), p. 676.

[7] L. Barbosa and J. Feng, *Robust sentiment detection on twitter from biased and noisy data*, in Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010, pp. 36–44.

[8] L. Bennett, S. Riger, P. Schewe, A. Howard, and S. Wasco, *Effectiveness of hotline, advocacy, counseling, and shelter services for victims of domestic violence: A statewide evaluation*, Journal of Interpersonal Violence, 19 (2004), pp. 815–829.

[9] M. S. Bernstein, L. H. B. Suh, J. Chen, S. Kairam, and E. H. Chi, *Eddi: Interactive topic-based browsing of social status streams*, in Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, ACM, 2010, pp. 303–312.

[10] D. M. Blei and J. D. Lafferty, *Dynamic topic models*, in Proceedings of the 23rd International Conference on Machine Learning, ACM, 2006, pp. 113–120.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent dirichlet allocation*, Journal of Machine Learning Research, 3 (2003), pp. 993–1022.

[12] B. G. Bokolo and Q. Liu, *Deep learning-based depression detection from social media: Comparative evaluation of ml and transformer techniques*, Electronics, 12 (2023), p. 4396.

[13] ——, *Deep learning-based depression detection from social media: Comparative evaluation of ml and transformer techniques*, Electronics, 12 (2023).

[14] J. Bollen, H. Mao, and A. Pepe, *Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena*, in Proceedings of the International Conference on Web and Social Media (ICWSM), vol. 11, 2011, pp. 450–453.

[15] C. Caragea, A. Silvescu, and A. H. Tapia, *Identifying informative messages in disaster events using convolutional neural networks*, in Proceedings of the International Conference on Information Systems for Crisis Response and Management, 2016, pp. 1–6.

[16] B. J. Casad, J. E. Franks, C. E. Garasky, M. M. Kittleman, A. C. Roesler, D. Y. Hall, and Z. W. Petzel, *Gender inequality in academia: Problems and solutions for women faculty in stem*, Journal of neuroscience research, 99 (2021), pp. 13–23.

[17] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, *Understanding emotions in text using deep learning and big data*, Computers in Human Behavior, 93 (2019), pp. 309–317.

[18] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, *A textual-based featuring approach for depression detection using machine learning classifiers and social media texts*, Computers in Biology and Medicine, 135 (2021), p. 104499.

[19] R. Chunara, J. R. Andrews, and J. S. Brownstein, *Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera out-*

break, The American journal of tropical medicine and hygiene, 86 (2012), pp. 39–45.

[20] R. Clark, '*hope in a hashtag' : The discursive activism of #whyistayed*, Feminist Media Studies, 16 (2016), pp. 788–804.

[21] N. K. Cobb, A. L. Graham, M. J. Byron, and D. B. Abrams, *Online social networks and smoking cessation: a scientific research agenda*, Journal of Medical Internet Research, 13 (2011), p. e119.

[22] G. Coppersmith, M. Dredze, and C. Harman, *Quantifying mental health signals in twitter*, in Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality, 2014, pp. 51–60.

[23] A. Culotta, *Detecting influenza outbreaks by analyzing twitter messages*, arXiv preprint arXiv:1007.4748, (2010).

[24] ——, *Lightweight methods to estimate influenza rates and alcohol sales volume from twitter messages*, Language Resources and Evaluation, 47 (2013), pp. 217–238.

[25] M. De Choudhury, S. Counts, and E. Horvitz, *Predicting postpartum changes in emotion and behavior via social media*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, 2013, ACM.

[26] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, *Predicting depression via social media*, in Proceedings of the international AAAI conference on web and social media, vol. 7, 2013, pp. 128–137.

[27] E. de Quincey and P. Kostkova, *Early warning and outbreak detection using social networking websites: The potential of twitter*, in Electronic Healthcare, Springer, 2009, pp. 21–24.

[28] C. Ezeafulukwe, O. R. Owolabi, O. F. Asuzu, S. C. Onyekwelu, C. U. Ike, and B. G. Bello, *Exploring career pathways for people with special needs in stem and beyond*, International Journal of Applied Research in Social Sciences, 6 (2024), pp. 140–150.

[29] T. Falaiye, O. A. Elufioye, K. F. Awonuga, C. V. Ibeh, F. O. Olatoye, and N. Z. Mhlongo, *Financial inclusion through technology: a review of trends in emerging markets*, International Journal of Management & Entrepreneurship Research, 6 (2024), pp. 368–379.

[30] H. Gao, G. Barbier, and R. Goolsby, *Harnessing the crowdsourcing power of social media for disaster relief*, IEEE Intelligent Systems, 26 (2011), pp. 10–14.

[31] G. Geetha, G. Saranya, K. Chakrapani, J. G. Ponsam, M. Safa, and S. Karpagaselvi, *Early detection of depression from social media data using machine learning algorithms*, in 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), 2020, pp. 1–6.

[32] S. Ghosal and A. Jain, *Depression and suicide risk detection on social media using fasttext embedding and xgboost classifier*, Procedia Computer Science, 218 (2023), pp. 1631–1639.

[33] S. A. Golder and M. W. Macy, *Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures*, Science, 333 (2011), pp. 1878–1881.

[34] N. Heaivilin, B. Gerbert, J. Page, and J. Gibbs, *Public health surveillance of dental pain via twitter*, Journal of Dental Research, 90 (2011), pp. 1047–1051.

[35] D. Houry, K. M. Feldhaus, B. Peery, J. Abbott, N. Crews, K. Viswanathan, and N. J. Kaslow, *Does screening in the emergency department hurt or help victims of intimate partner violence?*, Annals of Emergency Medicine, 51 (2008), pp. 433–442.

[36] J. Huang, M. Peng, and H. Wang, *Topic detection from large scale of microblog stream with high utility pattern clustering*, in Proceedings of the 8th Workshop on Ph. D. Workshop in Information and Knowledge Management, 2015, pp. 3–10.

[37] J. Huang, M. Peng, H. Wang, J. Cao, W. Gao, and X. Zhang, *A probabilistic method for emerging topic tracking in microblog stream*, World Wide Web, 20 (2017), pp. 325–350.

[38] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, *AIDR: Artificial intelligence for disaster response*, in Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 159–162.

[39] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, *Machine learning driven mental stress detection on reddit posts using natural language processing*, Human-Centric Intelligent Systems, 3 (2023), pp. 80–91.

[40] S. Jadhav, A. Machale, P. Mharnur, P. Munot, and S. Math, *Text based stress detection techniques analysis using social media*, in 2019 5th Inter-

national Conference On Computing, Communication, Control And Automation (ICCUBEA), IEEE, 2019, pp. 1–5.

[41] T. Jo, *Machine learning foundations*, Machine Learning Foundations. Springer Nature Switzerland AG. https://doi. org/10.1007/978-3-030-65900-4, (2021).

[42] J. Kim, J. Lee, E. Park, and J. Han, *A deep learning model for detecting mental illness from user content on social media*, Scientific reports, 10 (2020), p. 11846.

[43] S. S. Korti and S. G. Kanakaraddi, *Depression detection from twitter posts using nlp and machine learning techniques*, in 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2022, pp. 1–6.

[44] A. Kulwicki, B. Aswad, T. Carmona, and S. Ballout, *Barriers in the utilization of domestic violence services among arab immigrant women: Perceptions of professionals, service providers community leaders*, Journal of Family Violence, 25 (2010), pp. 727–735.

[45] A. Kumar, A. Sharma, and A. Arora, *Anxious depression prediction in real-time social data*, in Proceedings of the International Conference on Advanced Engineering, Science, Management and Technology (ICAESMT19), 2019.

[46] C.-H. Lee, C.-H. Wu, and T.-F. Chien, *BursT: A dynamic term weighting scheme for mining microblogging messages*, in Advances in Neural Networks - ISNN 2011, Springer, 2011, pp. 548–557.

[47] V. Leiva and A. Freire, *Towards suicide prevention: Early detection of depression on social media*, in Lecture Notes in Computer Science, International Conference on Internet Science, 2017.

[48] J. Li, H. Wang, H. Jin, and J. Yong, *Current developments of k-anonymous data releasing*, Electronic Journal of Health Informatics, 3 (2008), p. 6.

[49] B. Liang, L. Goodman, P. Tummala-Narra, and S. Weintraub, *A theoretical framework for understanding help-seeking processes among survivors of intimate partner violence*, American Journal of Community Psychology, 36 (2005), pp. 71–84.

[50] C. Liou, *Using social media for the prevention of violence against women.* `http://www.partners4prevention.org/sites/default/files/resources/socialmedia_final.pdf`. Accessed: 2024-09-14.

[51] B. Liu, *Opinion mining*, in Encyclopedia of Database Systems, Springer, 2008. Invited contribution.

[52] S. K. Lora, N. Sakib, S. A. Antora, and N. Jahan, *A comparative study to detect emotions from tweets analyzing machine learning and deep learning techniques*, International Journal of Applied Information Systems, 12 (2020), pp. 6–12.

[53] A. Marcus, M. S. Bernsteina, O. BadarD., R. Karger, S. Madden, and R. C. Miller, *Twitinfo: Aggregating and visualizing microblogs for event exploration*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2011, pp. 227–236.

[54] A. Mbarek, S. Jamoussi, A. Charfi, and A. Ben Hamadou, *Suicidal profiles detection in twitter*, in Proceedings of the 15th International Conference on Web Information Systems and Technologies (WEBIST 2019), 2019, pp. 289–296.

[55] S. Muñoz and C. A. Iglesias, *A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations*, Information Processing & Management, 59 (2022), p. 103011.

[56] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, *Stress detection using natural language processing and machine learning over social interactions*, Journal of Big Data, 9 (2022), p. 33.

[57] M. O. Nusrat, W. Shahzad, and S. A. Jamal, *Multi class depression detection through tweets using artificial intelligence*, arXiv preprint arXiv:2404.13104, (2024).

[58] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, *From tweets to polls: Linking text sentiment to public opinion time series*, in Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), vol. 11, 2010, pp. 1–2.

[59] B. O'Dea and et al., *Detecting suicidality on twitter*, Internet Interventions, 2 (2015), pp. 183–188.

[60] M. Odlum and S. Yoon, *What can we learn about the ebola outbreak from tweets?*, American Journal of Infection Control, 43 (2015), pp. 563–571.

[61] N. Oryngozha, P. Shamoi, and A. Igali, *Detection and analysis of stress-related posts in reddit' s acamedic communities*, IEEE Access, (2024).

[62] B. Pang and L. Lee, *Opinion mining and sentiment analysis*, Foundations and trends in information retrieval, 2 (2008), pp. 1–135.

[63] M. J. Paul and M. Dredze, *You are what you tweet: Analyzing twitter for public health*, in Proceedings of the International Conference on Weblogs and Social Media (ICWSM), vol. 20, 2011, pp. 265–272.

[64] ——, *A model for mining public health topics from twitter*, Health, 11 (2012), pp. 16–6.

[65] ——, *Discovering health topics in social media using topic models*, PLoS One, 9 (2014), p. e103408.

[66] S. Petrovi, M. Osborne, and V. Lavrenko, *Streaming first story detection with application to twitter*, in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 181–189.

[67] D. Ramage, S. T. Dumais, and D. J. Liebling, *Characterizing microblogs with topic models*, ICWSM, 10 (2010), pp. 1–1.

[68] D. Ramírez-Cifuentes, A. Freire, R. Baeza-Yates, J. Puntí, P. Medina-Bravo, D. A. Velazquez, J. M. Gonfaus, and J. Gonzàlez, *Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis*, Journal of Medical Internet Research, (2020).

[69] D. T. S. M. Rao, D. V. S. Kompalli, and U. Kompalli, *A model to detect social network mental disorders using ai techniques*, Journal of Critical Reviews, 7 (2020).

[70] R. F. Rodgers, S. J. Paxton, and E. H. Wertheim, *# take idealized bodies out of the picture: A scoping review of social media content aiming to protect and promote positive body image*, Body Image, 38 (2021), pp. 10–36.

[71] L. E. Rose and J. Campbell, *The role of social support and family relationships in women' s responses to battering*, Health Care for Women International, 21 (2000), pp. 27–39.

[72] T. Sakaki, M. Okazaki, and Y. Matsuo, *Earthquake shakes twitter users: real-time event detection by social sensors*, in Proceedings of the 19th international conference on World wide web, 2010, pp. 851–860.

[73] G. Salton, *Automatic text processing: The transformation, analysis, and retrieval of information by computer*, Addison-Wesley, Reading, MA, 1989.

[74] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, *Twitterstand: News in tweets*, in Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2009, pp. 42–51.

[75] H. Song, J. You, J.-W. Chung, and J. C. Park, *Feature attention network: Interpretable depression detection from social media*, in 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32), Hong Kong, December 1–3 2018.

[76] M. Stankevich, V. Isakov, D. Devyatkin, and I. Smirnov, *Feature engineering for depression detection in social media*, in Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018), 2018, pp. 426–431.

[77] S. Subramani, H. Wang, M. R. Islam, A. Ulhaq, and M. O'Connor, *Child abuse and domestic abuse: Content and feature analysis from social media disclosures*, in Proceedings of Australasian Database Conference, Cham, Switzerland, 2018, Springer, pp. 174–185.

[78] S. Subramani, H. Wang, H. Q. Vu, and G. Li, *Domestic violence crisis identification from facebook posts based on deep learning*, IEEE access, 6 (2018), pp. 54075–54085.

[79] L. Sun, H. Wang, J. Soar, and C. Rong, *Purpose based access control for privacy protection in e-healthcare services*, Journal of Software, 7 (2012), pp. 2443–2449.

[80] X. Sun, H. Wang, J. Li, and Y. Zhang, *Satisfying privacy requirements before data anonymization*, The Computer Journal, 55 (2012), pp. 422–437.

[81] K. M. Sylaska and K. M. Edwards, *Disclosure of intimate partner violence to informal social support network members: A review of the literature*, Trauma, Violence, Abuse, 15 (2014), pp. 3–21.

[82] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, *Detection of depression-related posts in reddit social media forum*, Ieee Access, 7 (2019), pp. 44883–44893.

[83] V. Tejaswini, K. Sathya Babu, and B. Sahoo, *Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model*, ACM Transactions on Asian and Low-Resource Language Information Processing, 23 (2024), pp. 1–20.

[84] J. P. Thekkekara, S. Yongchareon, and V. Liesaputra, *An attention-based cnn-bilstm model for depression detection on social media text*, Expert Systems with Applications, 249 (2024), p. 123834.

[85] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, *Predicting elections with twitter: What 140 characters reveal about political sentiment*, in Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), vol. 10, 2010, pp. 178–185.

[86] E. Turcan, S. Muresan, and K. McKeown, *Emotion-infused models for explainable psychological stress detection*, in Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies, 2021, pp. 2895–2909.

[87] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, *Deep learning for prediction of depressive symptoms in a large textual dataset*, Neural Computing and Applications, 34 (2022), pp. 721–744.

[88] N. Varghese and N. Kumar, *Feminism in advertising: irony or revolution? a critical review of femvertising*, Feminist Media Studies, 22 (2022), pp. 441–459.

[89] H. Wang, J. Cao, and Y. Zhang, *A flexible payment scheme and its role-based access control*, IEEE Transactions on Knowledge and Data Engineering, 17 (2005), pp. 425–436.

[90] H. Wang, Z. Zhang, and T. Taleb, *Special issue on security and privacy of IoT*, World Wide Web, (2017), pp. 1–6.

[91] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan, J. Pathak, and A. Sheth, *Semi-supervised approach to monitoring clinical depressive symptoms in social media*, in Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, 2017, pp. 1191–1198.

[92] L. Zhang, T. Fu, R. Yin, Q. Zhang, and B. Shen, *Prevalence of depression and anxiety in systemic lupus erythematosus: a systematic review and meta-analysis*, BMC psychiatry, 17 (2017), pp. 1–14.

[93] Y. Zheng, Y. Li, Z. Wen, B. Liu, and J. Tao, *Text-based sentential stress prediction using continuous lexical embedding for mandarin speech synthesis*, in 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), IEEE, 2016, pp. 1–5.