Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Depression

A Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Technology
in
Computer Science & Engineering

by
Rohit Singh(20214021)
Rohan Singh (20214152)
Pratik Kumar Das(20214247)
Rishi Dilawari(20214171)

to the
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY
ALLAHABAD PRAYAGRAJ
May, 2024

UNDERTAKING

We declare that the work presented in this report titled "*Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Depression* ", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, for the award of the ***Bachelor of Technology*** degree in ***Computer Science & Engineering***, is our original work. We have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, We accept that our degree may be unconditionally withdrawn.

May, 2024
Allahabad

Rohit Singh(20214021)
Rohan Singh (20214152)
Pratik Kumar Das(20214247)
Rishi Dilawari(20214171)

CERTIFICATE

Certified that the work contained in the report titled "*Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Depression*", by Rohit Singh, Rohan Singh, Pratik Kumar Das and Rishi Dilawari as been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Dr D.K.Yadav)
Computer Science and Engineering Dept.
M.N.N.I.T, Allahabad

May, 2024

Preface

This report aims to explore the potential application of Sentiment Detection Through an investigation into various concepts pertinent to web application development,.

Conducted as part of our Computer Science and Engineering curriculum at Motilal Nehru National Institute of Technology, Allahabad, this research project allowed us to gain practical experience in Sentiment Detection

Under the mentorship of Dr D.K.Yadav, we received consistent support and guidance throughout the rigorous research process. His assistance was invaluable in grasping crucial concepts, navigating the stages of coding, testing, analyzing, and comprehending the project. We express profound gratitude for his unwavering support and encouragement.

Acknowledgements

We are deeply thankful to Professor D.K. Yadav for his invaluable guidance, mentorship, and support throughout the completion of the Mental Disorders in Social Media Through Emotional Patterns - The Case of Depression. His expertise, encouragement, and unwavering dedication have been pivotal in shaping the success of this endeavor.

Acknowledgment is also extended to Motilal Nehru National Institute Of Technology for providing the essential resources and a conducive research environment. Moreover, I appreciate the collaborative effort demonstrated by our team. Working together seamlessly as a group of four, we tackled challenges, exchanged ideas, and collectively achieved our shared objective of completing the Mental Disorders in Social Media Through Emotional Patterns - The Case of Depression. Our joint endeavor, mutual encouragement, and steadfast support have been pivotal to our success.

The fruition of this project owes much to the combined efforts and contributions of all those involved. I am genuinely privileged to have had the opportunity to work under Professor D.K. Yadav's guidance and to have gained from his wealth of knowledge and experience.

Contents

Chapter 1

Introduction

The discussion about mental health has changed a lot in recent years, reflecting the prevalence and impact of various mental disorders around the world. From depression to anxiety, bipolar disorder to schizophrenia, these diseases not only affect the individual, but also the whole society and society, putting pressure on healthcare systems and public health. Despite research and treatment, solving mental health problems remains a difficult task. This is often compounded by stigma, less access to care, and many of these cases. Conventional treatments generally focus on pharmacological and psychological interventions; While these interventions are helpful for many people, they may not capture the complexity of the human experience or address thought patterns that cause stress. There is interest in exploring the role of mental health-important thought patterns in understanding and solving mental health problems. As the foundation of human cognition and behaviour, emotions play a critical role in shaping our thoughts, actions, and relationships. By examining thought patterns (repeating sequences of thoughts and feelings), we gain insight into the processes and intervention methods of psychology. How can understanding emotional states inform treatment strategies and improve clinical outcomes?

By comprehensively reviewing scientific research and theoretical frameworks, we aim to elucidate the mechanisms underlying various psychological disorders and clarify the conceptual models and how they influence symptoms and treatment.

1.1 Mental Disorder

Mental illness causes many disturbances in the thinking and behavior of those affected [1]. These effects can range from mild to severe and may lead to the inability to meet daily activities and basic needs [2]. Mental illnesses such as depression and anorexia affect millions of people worldwide. It may be related to a single stressful situation for the person, or it may be related to many different stressful situations. It is known that mental disorders will be more common in countries where violence or natural disasters are intense. For example, a 2018 study on mental health in Mexico showed that 17 percent of the population suffers from at least one mental illness, and one in four people suffers from the mental illness of dementia at least once in their life. Here in our project we will be focusing on depression so lets understand more about mental disorder(depression).

Depression and symptoms of depression: Depression is a mental illness characterised by sadness, hopelessness, and a lack of interest or pleasure in activities. Great result.

Common symptoms of depression include persistent depression, irritability, fatigue, changes in appetite or weight, sleepiness or excessive sleeping, feelings of worthlessness or guilt, and difficulty thinking or making decisions. Its duration varies and can range from mild to severe; It may take weeks, months or years. [8]

1.2 Use of emotional pattern

We realised that in today's world, social life can be experienced both in the real world and in the virtual world created by social media such as Facebook, Twitter, Reddit or similar platforms. This reality brings some challenges, but also offers great opportunities that, if used correctly, can lead to an understanding of what and how we communicate [3]. In this case, the aim of this study is to check whether there are signs of depression or negative emotions in the local people by analysing social media data by analysing thought patterns. By leveraging emotional patterns in conjunction with ML techniques, like supervised learning we can create a model that can detect the sign of depression and via this way we can help people before it goes into the worst condition.

1.3 Earlier studies

Several earlier studies have explored the use of emotional patterns for depression detection. For instance, research has utilised sentiment analysis of social media posts to identify linguistic markers associated with depressive symptoms [4], such as negative affect, self-disclosure, and social withdrawal. Other studies have investigated behavioural cues, such as changes in posting frequency or engagement patterns, as indicators of depression. Additionally, machine learning algorithms have been applied to detect subtle variations in emotional expression and language use that may signal underlying mental health issues. Overall, these studies highlight the potential of leveraging emotional patterns in digital data for early detection and intervention in depression.

Former studies focused on the detection of depression and anorexia have mainly considered linguistic and sentiment analysis. Note that the use of emotions, i.e., polarity, precedes the use of emotions in the same task [5] . This idea shows the possibility of using not language features or general emotions (such as positive, negative), but a different attitude such as “anger”, “surprised” or “happy” [6]. In this direction, we introduced a novel representation that was built using information extracted from emotions lexicons combined with word embeddings as a way to represent the information contained in users documents. Then, using a clustering algorithm,we created sub-groups of emotions, that we conveniently named as sub-emotions. These discovered sub-emotions provided a more flexible and fine-grained representation of users and a better performance for the detection of depression. In a few words, the idea behind this representation was to capture the presence of sub-emotions in users’ posts. The intuition of our approach is that users suffering from depression would show a distribution of emotions different from healthy users.

Lets see some result that shows how mental disorder is affecting lifes :
According to the World Health Organization (WHO), 970 million people worldwide were living with a mental disorder in 2019, which is 1 in every 8 people. In 2020, the

number of people with anxiety and depressive disorders increased by 26 percentage and 28 percentage, respectively, due to the COVID-19 pandemic [7].
World Health Organization (WHO) According to the Institute for Health Metrics Evaluation, 970 million people worldwide, or 31 percentage of the global population, live with mental disorders. The most common mental disorders are:
Anxiety disorders: 31percentage
Depressive disorders: 28.9percentage
Developmental disorders: 11.1percentage
Attention deficit/hyperactivity disorder: 8.8percentage
Bipolar disorder: 4.1percentage

Chapter 2

Related work

In this section we present an overview of previous works about the detection of depression using social media data. we describe their strengths and opportunities, and contrast the strategies used to our proposal.

2.1 Depression detection

Depression is a mental health disorder characterized by persistent loss of interest in activities, which can cause significant difficulties in everyday life. Studies focusing on the automatic detection of this disorder have used crowdsourcing as their main strategy to collect data from users who expressly have reported being diagnosed with clinical depression [8]. Among these studies, the most popular approach considers words and word n-grams as features and employs traditional classification algorithms [9]. The main idea is to capture the most frequent words used by individuals suffering from depression and compare them against the most frequent words used by healthy users. This approach suffers because there is usually a high overlap in the vocabulary of users with and without depression. Another group of works used a LIWC-based representation, aiming to represent user's posts by a set of psycho- logically meaningful categories like social relationships, thinking styles, or individual differences. These works have allowed a better characterization of the mental disorder conditions, nevertheless, they have only obtained moderately better results than using only the words.

These works show that in social media texts exist useful information to determine if a person suffers from depression, but the results are sometimes hard to interpret. This is an important limitation since these types of tools are naturally aimed to support health professionals and not to take the final decisions. In, the authors conduct studies to tackle this problem. They characterize users affected by mental disorders and provide methods for visualizing the data in order to provide useful insights to psychologists. Lastly, some works have also considered representations based on sentiment analysis techniques [10]. These works have shown interesting results, indicating that negative comments are more abundant in people with depression than in users who do not suffer from this disorder.

In a recent study [11], it was found that not only considering sentiments but also emotions to identify depression on Twitter users. This work was motivated by a psychological theory that relates the manifestation of feelings and emotions with depression, and its objective was to improve the interpretability of the results. In a previous work, we proposed to use a finer concept, called sub-emotions, reporting promising results to detect depression. Here, is where this study continues exploring this path, by proposing a new sub-emotion based representation, this time considering emotional changes through time, and also extending the potential use of this representation to detect anorexia.

2.2 Technology for healthcare

The integration of artificial intelligence (AI) and machine learning (ML) technologies in healthcare has revolutionized various aspects of medical practice, research, and management. Predictive analytics models can analyze patient data, including electronic health records (EHRs), genetic information, and environmental factors, to identify individuals at risk of developing certain diseases or complications, enabling proactive interventions and personalized treatment plans. For example, deep learning models have shown promise in accurately detecting and classifying tumors in radiological images.

Drug Discovery and Development: AI-driven drug discovery platforms leverage ML algorithms to analyze vast datasets of molecular structures, biological pathways, and clinical trial outcomes, accelerating the process of identifying potential drug candidates.

Natural language processing (NLP) technologies enable the automation of administrative tasks, such as medical coding, documentation, and billing, improving efficiency and reducing administrative burden on healthcare providers.

AI and ML technologies have emerged as powerful tools for improving detection, treatment, and management of conditions such as depression. Here's how they are being utilized: **Early Detection and Diagnosis:** AI algorithms analyze diverse data sources, including social media posts, electronic health records (EHRs), and smartphone sensor data, to identify patterns indicative of depressive symptoms. Natural language processing (NLP) techniques can detect linguistic cues and sentiment in text data, while behavioral analytics can assess changes in communication patterns and social interactions, aiding in early detection and diagnosis

Personalized Treatment Approaches: ML models leverage patient data, including genetic markers, neuroimaging scans, and treatment history, to tailor personalized treatment plans for individuals with depression. By analyzing large datasets, these algorithms can predict treatment response, recommend appropriate interventions, and optimize medication dosages, improving outcomes and reducing adverse effects. AI-driven analytics platforms analyze large-scale datasets, including electronic health records, clinical trials, and genomic data, to uncover novel insights into the underlying mechanisms of depression. By identifying biomarkers, genetic risk factors, and treatment response predictors, these tools advance our understanding of the disorder and inform the development of targeted interventions. Overall, AI and ML technologies hold promise for transforming the landscape of depression care by enabling early detection, personalized treatment, and innovative interventions.

2.3 Transformers based learning model

Transformers have emerged as a groundbreaking architecture in the field of natural language processing (NLP) and have significantly advanced the state-of-the-art in various NLP tasks.

Transformer Architecture: The transformer architecture, introduced in the paper "Attention is All You Need" [12]relies on a self-attention mechanism to process input sequences. It consists of an encoder-decoder architecture with multiple layers of attention and feedforward neural networks. The encoder processes input sequences by attending to all positions in the input sequence simultaneously, capturing long-range dependencies efficiently. The decoder generates output sequences autoregressively, attending to both the input sequence (via encoder-decoder attention) and previously generated tokens in the output sequence.

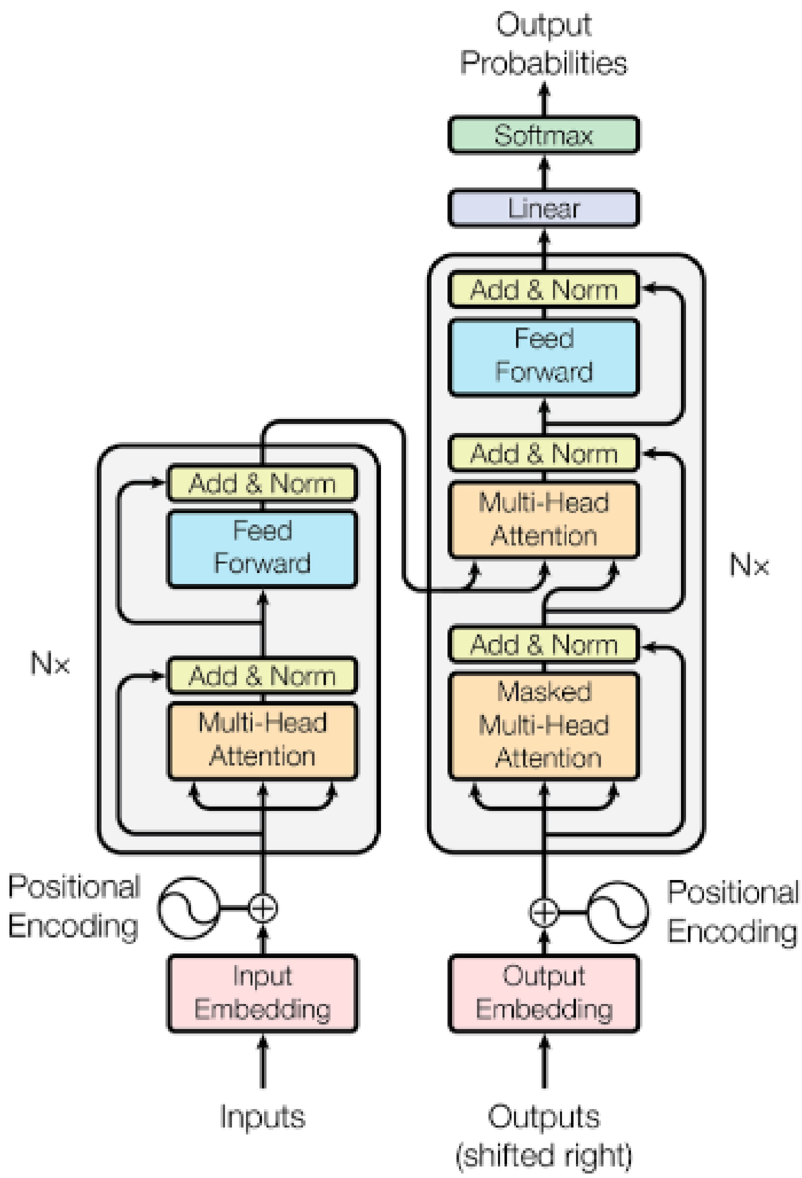


Figure 1.The Transformer Model Architecture:

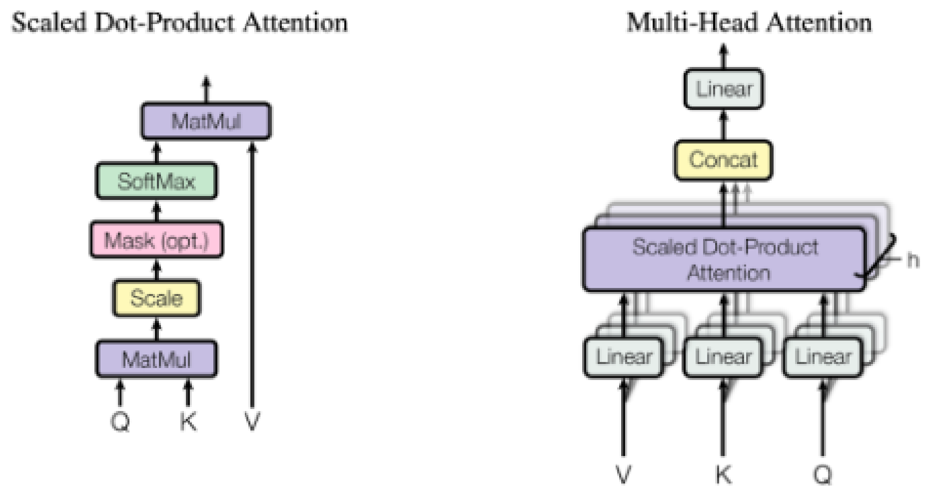
- **Input Embeddings**: The input sequence is first represented as a sequence of embeddings, where each token in the input is mapped to a high-dimensional vector representation.

  These embeddings may include pre-trained word embeddings (e.g., Word2Vec, GloVe) or learned embeddings specific to the task at hand.

- **Positional Encoding**: Transformers incorporate positional encoding to provide information about the position of tokens in the input sequence.

  Positional encodings are added to the input embeddings to capture the sequential order of tokens, allowing the model to differentiate between tokens based on their position in the sequence

- **Self-Attention Mechanism**: The core component of the transformer architecture is the self-attention mechanism, which computes attention scores between all pairs of positions in the input sequence. Self-attention allows the model to weigh the importance of each token in the input sequence when computing representations, enabling it to capture dependencies between distant tokens effectively.



- **Feedforward Neural Networks**: Following the self-attention layers, the transformer architecture typically includes feedforward neural networks (FFNs) at each position in the sequence.

These FFNs consist of two linear transformations with a non-linear activation function (e.g., ReLU) applied in between, enabling the model to learn complex, non-linear transformations of the input representations.

- **Layer Normalization and Residual Connections**: Layer normalization is applied after each sub-layer (self-attention and FFN) to stabilize training and improve generalization.

  Residual connections (skip connections) are used to connect the input of each sub-layer to its output, facilitating gradient flow and mitigating the vanishing gradient problem during training.

- **Encoder-Decoder Architecture (Optional)**: In tasks such as machine translation or text generation, transformers may utilize an encoder-decoder architecture, where the input sequence is encoded by an encoder stack, and the decoder generates an output sequence autoregressively based on the encoded representation and previously generated tokens.

- **Output Layer**: The final output of the transformer is typically passed through a linear transformation followed by a softmax function to produce the predicted probabilities for each token in the vocabulary (in language modeling tasks) or the output classes (in classification tasks).

- **Training and Optimization**: Transformers are trained using stochastic gradient descent (SGD) or variants such as Adam, with backpropagation used to compute gradients and update model parameters. Training objectives may include maximum likelihood estimation (MLE), masked language modeling (MLM), next sentence prediction (NSP), or task-specific objectives depending on the task and dataset.

Chapter 3

Proposed work

3.1 Objective:

Detecting mental disorders in social media through Emotional patterns - the case of Depression.

3.2 Use cases:

- **Early Intervention**: Identify users displaying signs of depression and offer them immediate access to mental health resources and support services.
- **Public Health Surveillance**: Monitor changes in emotional patterns on social media to track trends in depression prevalence and inform public health interventions.
- **Personalized Recommendations**: Offer personalized recommendations for self-care activities, therapy, or support groups to individuals at risk for depression based on their emotional patterns.
- **Crisis Response**: Develop early warning systems to detect individuals in acute distress or crisis on social media and provide timely intervention from mental health professionals

To achieve our objective: we aim to develop a machine learning-based approach for detecting depression from social media posts. Using various machine learning algorithm ex- SVM, Random Forest, Logistic Regression, Naive Baise and some Deep learning model like Bert also we have used transformer-based models ex- GPT2.0 , we will analyze textual data from platforms like Twitter and reddit to capture contextual information and linguistic patterns indicative of depressive symptoms. By fine-tuning pre-trained transformer models on a task-specific dataset, we seek to enhance the accuracy and efficiency of depression detection, ultimately facilitating early intervention and support for individuals at risk of mental health challenges.

Former studies focused on the detection of depression and anorexia have mainly considered linguistic and sentiment analysis .the use of sentiments, i.e., polarity, was the preamble for the later use of emotions for the same task . This line of thought exposed the potential of using emotions as features, such as "anger", "surprise" or "joy", instead of linguistic features or general sentiments like positive and negative. we are using[13] novel representation that was built using information extracted from emotions lexicons combined with word embeddings as a way to represent the information contained in users' documents. Then, using a clustering algorithm,

we created sub-groups of emotions, that conveniently we named as sub-emotions. These discovered sub emotions provided a more flexible and fine-grained representation of users and a better performance for the detection of depression. the idea behind this representation was to capture the presence of sub-emotions in users' posts. the idea behind is that users suffering from depression would show a distribution of emotions different from healthy users. now we can mask the document using these sub emotions and this masked document can be used to analyze their emotional pattern. Now we represent each user by a histogram of sub-emotions, which are discovered by clustering the embeddings of words inside coarse emotions.

3.3 From text to Fine grained emotions:

Emotions are pervasive among humans and had widely been studied in different fields like psychology and neuro- science . In particular, in psychology the correlation between emotions and mental disorders has been established. The proposed method for the detection of depression considers representations of documents based on their expressed fine-grained emotions. In order to construct these representations, first, we generate groups of fine- grained emotions (referred as sub-emotions from here on) for each general emotion that belong to the EmoLEX lexicon [14]. This lexical resource indicates the association of words with eight emotions: Anger, Fear, Anticipation, Trust, Sur- prise, Sadness, Joy, and Disgust, and two sentiments: Negative and Positive.Then, we mask the text and represent each document using the sub-emotions labels instead of the original words. The following sections described in detail each step of this procedure.

3.3.1 Generating the SubEmotions:

We represent the set of emotions within EmoLex in a formal way as E = E1,E2,...E10. where Ei = t1,t2,....tn is the set of words associated with emotion Ei. We compute a vector for each word in the lexical resource using Wikipedia pre- trained sub-word embeddings of size 300 from FastText [15]. After computing the vectors for each word (from each coarse emotion), we cluster them using the Affinity Propagation (AP) algorithm[16], which is a graph based clustering algorithm similar to k-means, but that does not require to establish the number of clusters in advance. After the clustering, each centroid represents a different subemotion. That is, now each emotion is modeled as a set of sub-emotions, Ei = s1,s2,s3,...sk, where each Sj is a subset of the words from Ei. This process creates a set S with all computed sub-emotions.

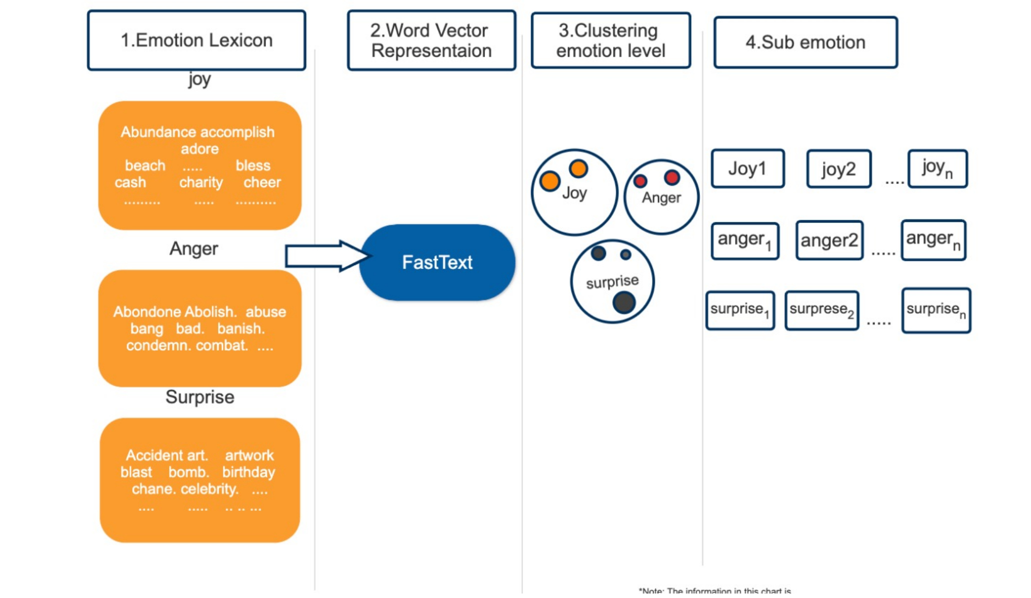
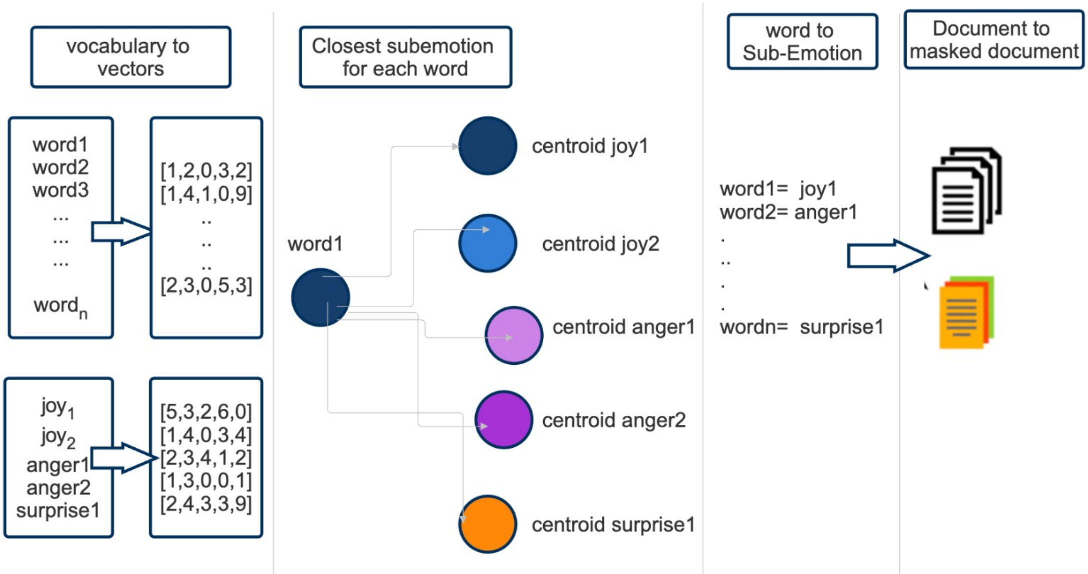


Figure 1.Generating of SubEmotion

3.3.2 Converting text to Subemotion:

Once obtained these vectors, for each word t in a text sample we measure its cosine similarity with all sub-emotions vectors S, and substitute it by a label X(t) related to its closest subemotion. That is, X(t) = Sj : max(sim(t,Sj)) for all sj.



Chapter 4

Experimental setup and Implementation

4.1 Data collection:

Data collection encompasses a diverse range of posts from both Reddit and Twitter, allowing for a comprehensive analysis of emotional patterns across different social media platforms.The dataset comprises a substantial volume of posts from Reddit and Twitter, providing a rich source of textual data for detecting emotional patterns associated with depression. Data collection from Reddit and Twitter ensures user anonymity, preserving privacy while enabling the analysis of authentic and unfiltered expressions of emotions related to mental health.

We have two datasets 1 with 7727 posts, dataset 2 with 6634 posts.

4.2 Data cleaning:

This process involves identifying and rectifying inaccuracies, inconsistencies, and missing values within the dataset. Techniques such as removing duplicate entries, correcting spelling errors, and handling missing data are employed to ensure the integrity and reliability of the dataset. Additionally, outliers and noise may be addressed through statistical methods or domain-specific knowledge. By conducting thorough data cleaning, researchers can mitigate biases, improve the robustness of their analyses, and ensure that subsequent insights and conclusions drawn from the data are accurate and reliable.

4.3 Data preprocessing :

It involves a series of operations aimed at transforming and standardizing the data to make it suitable for further processing. This may include tasks such as cleaning the data to remove noise, handling missing values, and addressing inconsistencies. Additionally, data preprocessing often involves feature scaling to normalize the range of numeric features, encoding categorical variables into a format suitable for machine learning algorithms, and splitting the data into training and testing sets.

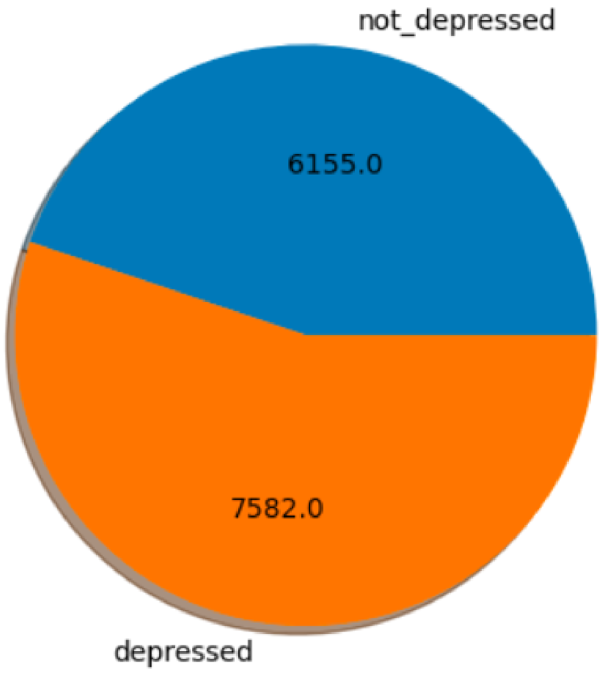


Figure 1. Final depression dataset:Graphical Visualization

4.4 Implementation :

There are multiple steps involved in this process let's discuss sequentially

Emotion based representation of data :

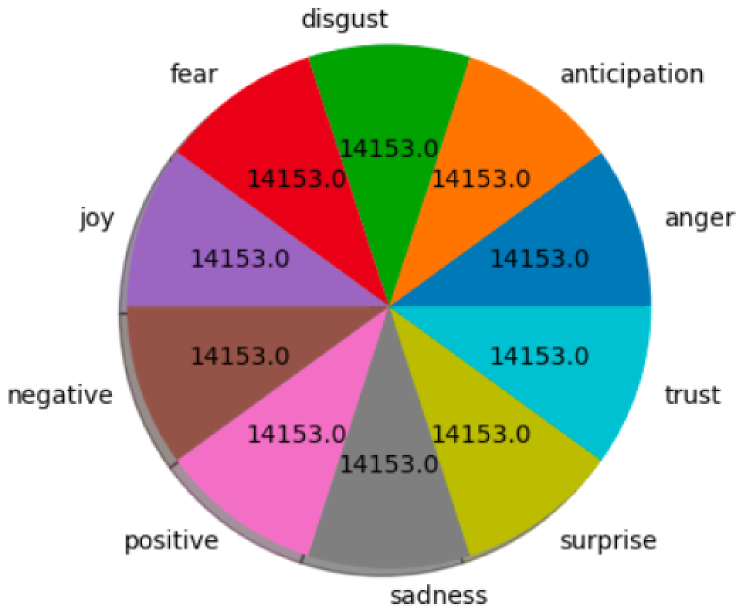


Figure 2. Word Dictionary for Emotion

After generating these sub emotions we will use cosine similarity to mask the words here is the snapshot of output

```
0       joy_1327 anger_370 disgust_952 disgust_938 dis...
1       disgust_741 disgust_937 fear_173 anger_370 dis...
2       anger_617 anger_1130 disgust_1286 positive_360...
3       anger_370 anticipation_701 positive_305 disgus...
4       anger_1130 trust_45 anticipation_57 anger_350 ...
```

Data will look something like this, here every word is masked with sub emotion This will allow us to better capture the emotional pattern.

4.4.1 Used machine learning algorithm

- **Naive bayes**:

  The Naive Bayes algorithm is a simple yet powerful probabilistic classification algorithm based on Bayes' theorem with a strong assumption of independence between the features. Despite its simplicity, Naive Bayes is widely used in various applications such as text classification, spam filtering, and sentiment analysis due to its efficiency and effectiveness.

Assumption of Feature Independence: The "naive" assumption in Naive Bayes is that all features are independent of each other given the class label. This means that the presence of a particular feature in a class is independent of the presence of any other feature.
Types of Naive Bayes: There are different variants of Naive Bayes, such as Gaussian Naive Bayes (for continuous features), Multinomial Naive Bayes (for discrete features), and Bernoulli Naive Bayes (for binary features).
Classification Decision: Once the probabilities for each class are computed, Naive Bayes selects the class with the highest probability as the predicted class for the input.

- **Random forest** :

  Random Forest is a versatile and powerful ensemble learning algorithm used for classification and regression tasks. It operates by constructing multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. Random Forest improves upon traditional decision trees by introducing randomness in the feature selection process and by aggregating predictions from multiple trees, which helps to reduce overfitting and improve generalization performance. It is highly scalable, handles high-dimensional data well, and is less prone to overfitting compared to individual decision trees.

- **Support vector machine**: Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the hyperplane that best separates the data into different classes or predicts the target variable in the case of regression.

Hyperplane: In SVM, a hyperplane is a decision boundary that separates the data points into different classes. For a binary classification problem, the hyperplane is a (n-1) dimensional subspace in an n dimensional feature space, where n is the number of features.

The goal of SVM is to find the hyperplane that maximizes the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors.

- **Logistic regression**: Logistic Regression is a simple yet effective linear classification algorithm used for binary classification tasks. It models the probability that a given input belongs to a particular class using the logistic function. The algorithm computes a weighted sum of the input features and applies the logistic function to produce a probability score between 0 and 1. If the probability score exceeds a predefined threshold (usually 0.5), the input is classified into the positive class; otherwise, it is classified into the negative class.

- **Gpt2.0**: GPT-2 (Generative Pre-trained Transformer 2) is a state-of-the-art language model architecture based on the Transformer architecture. It consists of a multi-layer bidirectional Transformer encoder-decoder architecture with self-attention mechanisms, enabling it to generate coherent and contextually relevant text. GPT-2 utilizes unsupervised pre-training on large-scale text corpora followed by fine-tuning on specific downstream tasks, allowing it to adapt to various natural language processing tasks such as text generation, translation, summarization, and question answering.

Chapter 5

Results analysis

In this chapter we will discuss about the results obtained from used models, in our project we have used number of classification algorithms we will describe the outputs of all of them and comparison among them.

Accuracy: Accuracy measures the proportion of correctly classified instances among the total number of instances.

Recall (Sensitivity or True Positive Rate): Recall measures the proportion of true positives (correctly predicted positive instances) among all actual positive instances.

F1 Score: F1 score is the harmonic mean of precision and recall. It balances both precision and recall. It is calculated as:

Support: Support refers to the number of occurrences of each class in the actual dataset. It is used to determine the importance of each class in the dataset. Support is helpful in understanding whether the evaluation metrics are reliable, especially for imbalanced datasets.

5.1 Result Of Model for Unmasked Data

Table 1: Accuracy, Precision, and F1 Scores of various models Without masked data

MODELS	Accuracy	Precision	F1 score
Naive Bayes	0.85	0.89	0.85
Random Forest	0.96	0.96	0.96
SVM	0.95	0.93	0.95
Logistic Regression	0.96	0.94	0.96
GPT2.0	0.98	0.97	0.98

5.2 Result Of Model for Masked Data

Table 2: Accuracy, Precision, and F1 Scores of various models With masked data

MODELS	Accuracy	Precision	F1 score
Naive Bayes	0.89	0.89	0.88
Random Forest	0.96	0.96	0.96
SVM	0.78	0.78	0.78
Logistic Regression	0.97	0.97	0.97
GPT2.0	0.61	0.61	0.61

5.3 Graph Obtain For GPT

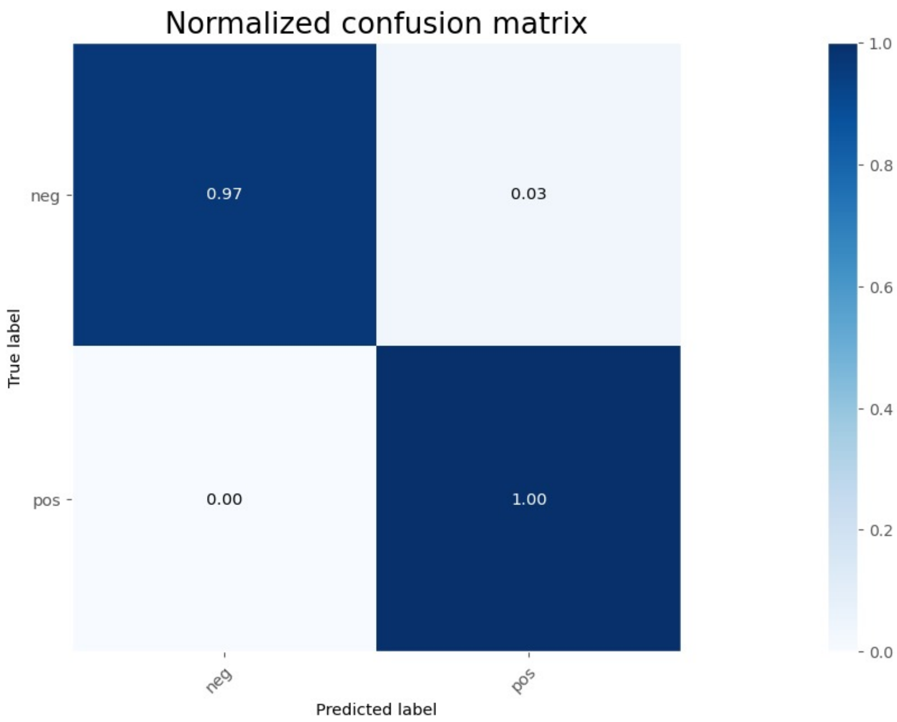


Figure 1. Normalized Confusion Matrix.

5.4 Comaparison between large language model(gpt 2.0) and support vector machine

- **Capturing Contextual Dependencies**:

  Transformers are designed to capture complex dependencies and contextual information in sequential data, such as text. In the context of depression detection, this means that the model can analyze social media posts or textual data from individuals and understand the nuanced relationships between words and phrases that may indicate depressive symptoms. This contextual understanding can enhance the model's ability to detect patterns associated with depression more accurately compared to SVM, which may struggle to capture long-range dependencies.

- **Learning from Large-Scale Data**: Transformers are often pre-trained on large-scale datasets using self-supervised learning objectives, allowing them to learn rich representations of language from vast amounts of text data. This pre-training enables the model to leverage knowledge acquired from diverse sources and domains, potentially improving its performance in depression detection tasks. In contrast, SVM typically relies on handcrafted features and may struggle to generalize well to new data without extensive feature engineering.

- **Adaptability to Various Input Lengths**: Transformers can process input sequences of variable lengths, making them suitable for analyzing social media posts or textual data, which may vary significantly in length. The model's self-attention mechanism allows it to attend to relevant parts of the input sequence adaptively, regardless of its length. This flexibility can be particularly advantageous in depression detection, where individuals may express their emotions and experiences differently across different lengths of text.

- **Transfer Learning and Fine-Tuning**: Pre-trained transformer models, such as BERT or RoBERTa, can be fine-tuned on task-specific datasets for depression detection. This transfer learning approach allows the model to adapt its learned representations to the target task more efficiently, potentially reducing the need for labeled data and improving performance. In contrast, SVM typically requires handcrafted features and may require more labeled data to achieve comparable performance.

- **Handling Non-linear Relationships**: Transformers employ multi-layer neural networks with non-linear activation functions, allowing them to capture complex, non-linear relationships between input features. This flexibility enables the model to learn intricate patterns and interactions in the data, which may be crucial for detecting subtle indicators of depression. SVM, on the other hand, relies on linear decision boundaries and may struggle to model non-linear relationships effectively without kernel tricks.

Overall, integrating the transformer-based model into a depression detection project can enhance performance by leveraging its ability to capture contextual dependencies, learn from large-scale data, adapt to variable input lengths, leverage transfer learning, and model non-linear relationships effectively. However, it's essential to consider factors such as computational resources, data availability, and interpretability when choosing the appropriate model for the task.

Chapter 6

Conclusion and Future work

6.1 Conclusion

In this work, we showed that representations based on fine- grained emotions can capture more specific topics and issues that are expressed in social media documents by users that unfortunately experience depression or anorexia. That is, the automatically extracted sub-emotions present useful information that helps the detection of these two mental disorders.this study has demonstrated the potential of utilizing social media data for detecting depression, offering a promising avenue for early intervention and support. By leveraging machine learning techniques, particularly in natural language processing, we've shown that it's feasible to develop models capable of accurately identifying patterns indicative of depression in social media posts.

6.2 Future work

Future work in this area could explore several avenues to enhance the effectiveness of depression detection using social media posts. This includes investigating the integration of multimodal data sources, such as text, images, and user interactions, to capture a more comprehensive understanding of individuals' mental states. Additionally, research could focus on developing personalized and context-aware models that account for individual differences and social contexts. Furthermore, efforts to address ethical considerations, such as privacy concerns and algorithmic biases, are essential to ensure the responsible deployment of these models in real-world settings. Finally, longitudinal studies are needed to assess the long-term impact of social media-based interventions on mental health outcomes and inform evidence-based practices for supporting individuals at risk of depression.

[8] [11] [12] [5] [3] [7] [13] [10] [6] [4] [9] [2] [1]

References

[1] Ap ://lazyprogrammer.me/mlcompendium/clustering/affinity.html: :text=affinity

[2] Word vectors for 157 languages ://fasttext.cc/docs/en/crawl-vectors.html.

[3] Mario Ezra Aragón, Adrián Pastor López Monroy, Luis Carlos González-Gurrola, and Manuel Montes. Detecting depression in social media using fine-grained emotions. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 1481–1486, 2019.

[4] Mario Ezra Aragón, Adrian Pastor López-Monroy, Luis Carlos González-Gurrola, and Manuel Montes-y Gómez. Detecting mental disorders in social media through emotional patterns - the case of anorexia and depression. *IEEE Transactions on Affective Computing*, 14(1):211–222, 2023.

[5] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In *Companion proceedings of the the web conference 2018*, pages 1653–1660, 2018.

[6] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. What about mood swings: Identifying depression on twitter with temporal measures of emotions. pages 1653–1660, 2018.

[7] Scott Counts Eric J. Horvitz, Munmun de Choudhury. Social media as a measurement tool of depression in populations ://dl.acm.org/doi/10.1145/2464464.2464480.

[8] Ronald C Kessler, Evelyn J Bromet, Peter de Jonge, Victoria Shahly, and Marsha Wilcox. The burden of depressive illness. *Public health perspectives on depressive disorders. Baltimore: John Hopkins University Press*, page 40, 2017.

[9] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.

[10] Daniel Preoţiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H. Andrew Schwartz, and Lyle Ungar. The role of personality, age, and gender in tweeting about mental illness. pages 21–30, June 5 2015.

[11] Freire Ana Ramírez-Cifuentes, Diana. Upf's participation at the clef erisk 2018: early risk prediction on the internet
https://repositori.upf.edu/handle/10230/35589.

[12] Briefing Room. Fact sheet: President biden to announce strategy to address our national mental health crisis, as part of unity agenda in his first state of the union. 2022.

[13] Fumio Kishino Yuichi Itoh Sho Tsugawa, Yusuke Kikuchi. Recognizing depression from twitter activity ://dl.acm.org/doi/10.1145/2702123.2702280.