Predictive Analysis Of Social Media Sentiments And Its Impact On Mental Health

By

Kashmira Ghag

This dissertation is submitted for the degree of

Masters in Business Analytics

(MSc. BA)

At

Dublin Business School (DBS)

Supervisor - Paul McEvoy

January 2025

DECLARATION

I, Kashmira Baliram Ghag, hereby declare that the applied project submitted to Dublin Business School in partial fulfilment of the requirements for the MSc in Business Analytics (MSc. BA) is the result of my independent research and investigation unless stated otherwise and properly acknowledged through references. I further confirm that this work was not submitted, in whole or in part, for any other academic degree or certification.

Signed: Kashmira Baliram Ghag

Student Number: 20023193

Date: 06/01/25

ACKNOWLEDGEMENTS

I want to express my profound gratitude to my supervisor, Paul McEvoy, for his indispensable role in the success of my thesis. His direction, critical input, and steadfast support were instrumental and reassuring during my thesis development. His insights and encouragement have been crucial in assuring the success of our research. Their contributions and encouragement have been critical in ensuring the effective completion of this research.

I am also deeply grateful to Dublin Business School for providing a nurturing academic environment and access to the necessary resources and facilities for this endeavour. Their support has been instrumental in the successful completion of this research.

I want to thank my family and friends for their ongoing encouragement, compassion, and patience throughout this journey. Their confidence in my ability has served as a constant source of motivation.

Lastly, I wish to acknowledge the contributions of the authors and researchers whose works have provided a foundation for my independent research. Their academic efforts have been a guiding light in shaping this thesis.

Thank you all for your invaluable support and inspiration.

ABSTRACT

The study will utilise predictive analysis to understand better the connection between social media emotions and mental health. Every day, social media networks generate massive volumes of data that represent the feelings and perceptions of users. The current study seeks to completely analyse the mental wellness of the population by examining attitudes toward user-generated content on social media. The project aims to uncover trends and patterns in the emotional tone of posts and conversations that can be used to predict mental health concerns. The study will focus on identifying early warning signs of psychological distress, creating the potential long-term impacts of social media interactions on mental well-being, and providing insights into how these platforms affect overall psychological health. The findings of this study are intended to contribute to a broader discussion regarding the role of social media in modern society, addressing both its benefits and possible drawbacks. Finally, the study aims to give practical tools for mental health physicians, policymakers, and social media platforms to understand better and address the complex link between social media use and its effects on mental health.

Keywords: Social Media, Sentiment Analysis, Mental Health, Psychological Well-being, Social Networks, Emotional Tone, Stress, Anxiety, Depression, Online Interactions, Digital Technology, Mental Health Trends, Social Media Impact.

Table of Contents

TABLE OF FIGURES

CHAPTER 1 - INTRODUCTION

1.1 OVERVIEW

Mental health is growing as a major concern in today's world, with many people suffering different kinds of emotions like anxiety, sadness, and even thoughts about suicide. While a variety of factors contribute to these issues, such as family issues, past trauma, or one's own struggles, research shows that social media has an immense effect on mental health, particularly among a younger audience.

Social media platforms such as Twitter, Facebook, Snapchat and Instagram have swiftly become part of everyday life, connecting individuals worldwide and allowing them to share information, thoughts, and emotions. Although these platforms also help us to stay informed and engage with people from everywhere, they also have drawbacks. People frequently use these platforms to insult, harass, criticise, and pass on negativity, often without thinking of the impact on the mental health of others. For example, cyberbullying and cruel comments can leave users feeling lonely, worried, or even depressed.

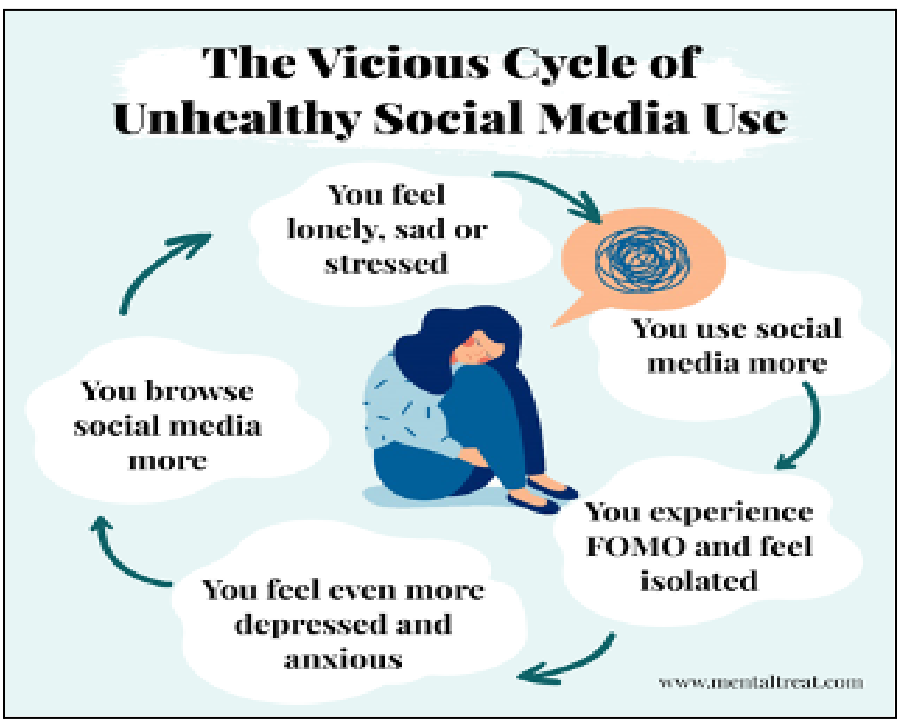

Figure 1 - The Cycle of Social Media

The statistics show a worrying trend. As per(Mir, Cui & Sun, n.d.; National Center for Health Research n.d.), 13% of 12-17-year-olds report depression and 32% report anxiety; mental illness is a concern for adolescent health. It is a concern for young adults as well since 33.7% of 18-25-year-olds report having some form of mental illness. Depression is greatly increasing among young girls. Some researchers suggest that the increase in mental illness is, at least in part, connected to the rise of social media use among adolescents and young adults.

While social media is not the only factor that impacts mental health, it is unique to our time and has become a critical area to study. This study focuses on how social media affects the individual's mental health, aiming to understand how the content people interact which impacts their emotional well-being. This study uses machine learning models to analyse the sentiments of individuals based on the posts, comments, and tags they share on social media. By studying different machine learning models, we can determine which models most effectively identify human emotions(sentiments) and understand how these emotions relate to mental health.

Ultimately, this research could help us predict how social media interactions impact mental health and offer insights into ways to support people who might be struggling. Through this study, we aim to work on creating a safer, better and healthier online environment for everyone while striving to mitigate the adverse effects of social media.

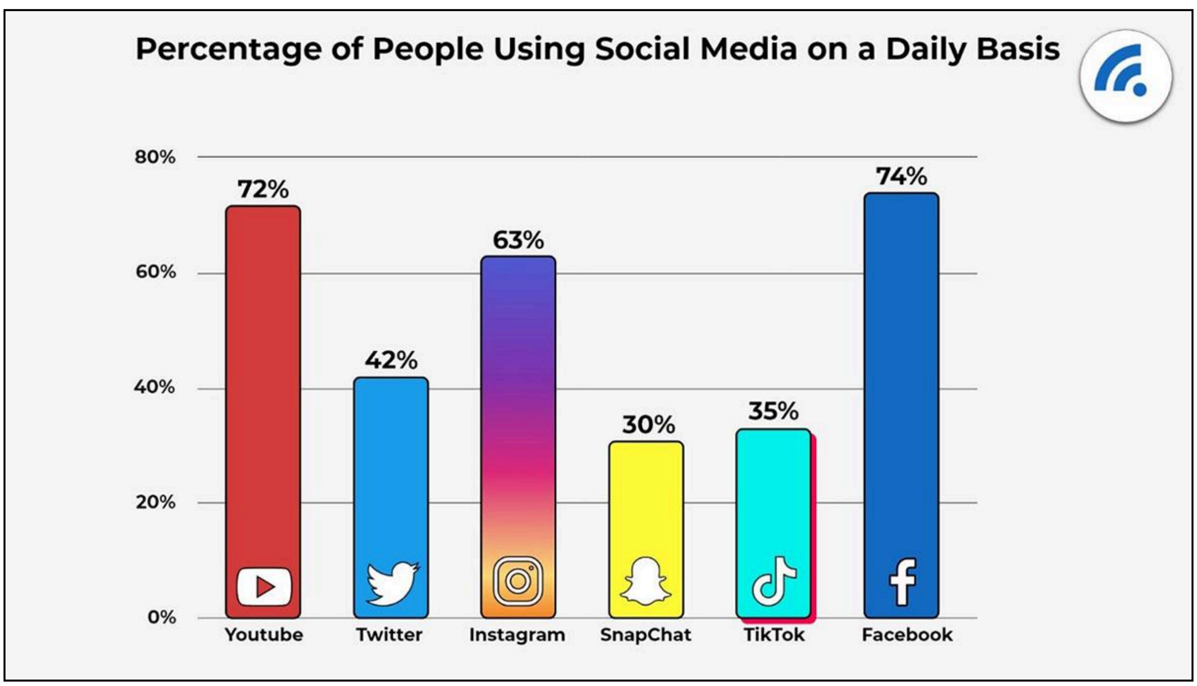

Figure 2: Percentage of People Using Social Media Daily

1.2 PROBLEM STATEMENT

The extreme use of social media has changed the way of communication by allowing users to communicate their opinions and emotions publicly. However, the overuse of a few websites, such as Twitter, Facebook, and Instagram, has raised concerns related to mental health problems such as anxiety, depression, and loneliness. However, the precise mechanisms through which social media emotions influence mental health are not fully understood. To recommend actions on social media and their impact on mental health, there is a need to have robust and sound predictive models. It can be used in diagnosing the first signs that can indicate the development of mental disorders, as well as the impact that SM interactions have on psychological health. To this end, this study uses Machine learning and Natural language Processing to forecast the effect of positive and negative sentiments on Mental health in Social media.

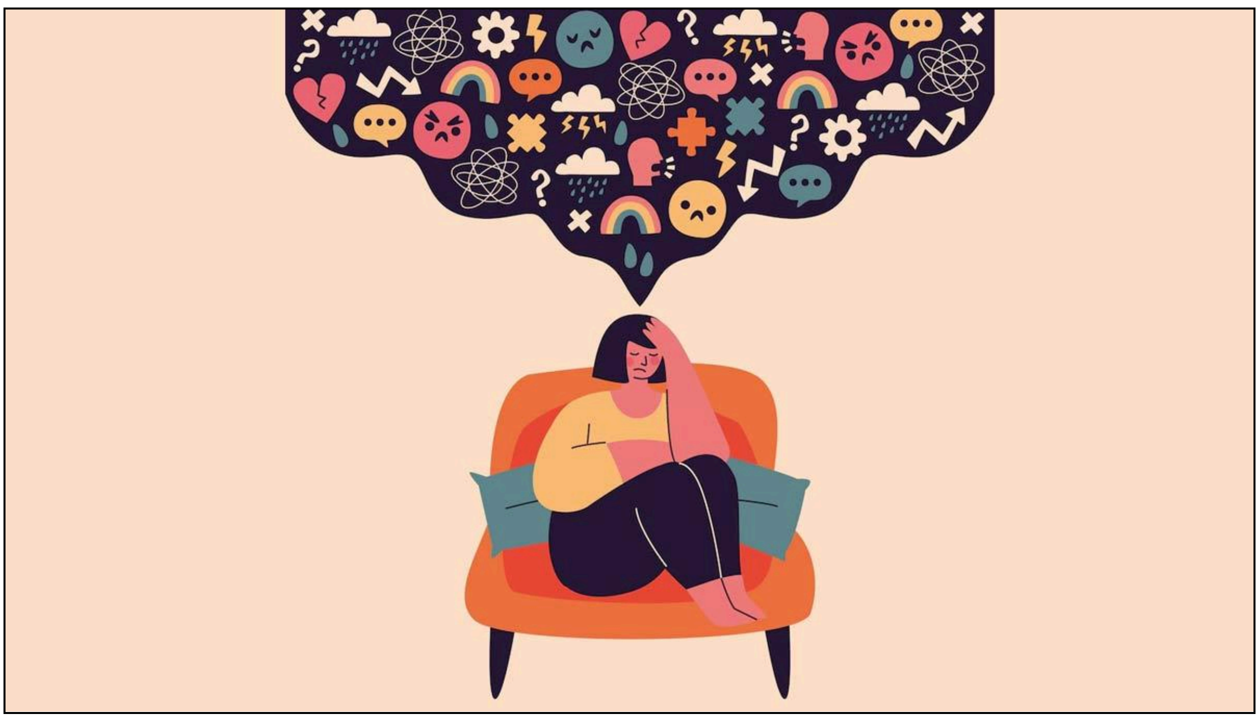

Figure 3: Impact of Social Media

1.3 RESEARCH OBJECTIVE

This study focuses mainly on sentiment analysis to gain a better and more comprehensive understanding of how social media influences mental health. While most studies analyse sentiment data from a single platform (e.g., Twitter or Reddit), this research will recognise that people use different platforms for different purposes, such as Instagram for sharing images, Twitter for expressing opinions, WhatsApp for private chats, and Reddit for seeking advice each of which influences the emotional content they share. This study intends to improve sentiment analysis, emotional pattern identification, and tracking of mental health trends by integrating data across platforms and implementing machine learning algorithms.

Machine learning will be critical in analysing data from many platforms, noticing emotional trends, and delivering insights to assist early interventions and improve people's mental health.

The primary objectives are categorised below to help us better understand our goal, and achieving each objective will add to our success.

- **Analyse the emotional impact of social media comments**

  The objective is to understand how a single social media comment from any platform might affect a person's emotional state. The study's goal is to uncover how different kinds of comments—whether positive, neutral, or negative—affect mental well-being and may contribute to major mental health concerns.

- **Utilise machine learning for sentiment analysis**

  Our primary goal is to implement machine learning techniques in this study to detect and analyse emotions expressed in social media comments accurately. The target is to implement various sentiment analysis algorithms and select the most effective algorithm. The study additionally seeks to identify patterns of emotional distress, providing insights into how social media interactions affect the mental health of people.

- **Identify early signs of mental health difficulties.**

  The purpose of this research is to identify potential warning signs of mental health issues by analyzing how individuals express themselves in social media comments. These insights can aid in predicting how negative interactions might escalate into more severe mental health issues, enabling timely interventions.

- **Assess and compare the performance of machine learning models**

  This objective tests multiple machine learning models to determine their accuracy and efficiency in sentiment analysis. The study will identify the most suitable model for predicting emotional outcomes based on social media interactions.

- **To develop actionable insights for mental health interventions**

  Based on the findings, The study will provide techniques for reducing the adverse impact of social media comments on mental health. The goal is to offer recommendations for creating supportive online environments and developing tools to promote emotional well-being.

1.4 RESEARCH QUESTION

1. On which features can sentiment analysis be optimized to effectively assess its impact on mental health, focusing on specific issues such as depression, anxiety, or stress?
2. Which evaluation criteria can be focused to measure the effectiveness and reliability of sentiment analysis models in identifying mental health-related sentiments from social media comments?
3. Can simpler machine learning models reliably predict mental health states from social media comments, and how do they compare to more complex approaches in terms of accuracy and efficiency?

1.5 DISSERTATION OUTLINE

This section will help you understand the workflow of this study. It explains the sequence or structure we have followed to achieve the results.

Chapter 1 - Introduction

The first chapter is a starting point, providing an outline of the research objectives and environment. It begins with an overview of the research laying out the flow of the framework for the study, followed by the Problem Statement, which explains the underlying difficulties that motivate the research. Next, the Research Objectives are explicitly defined to clarify the study's goals. Finally, the chapter finishes with the Research Questions, which lead the inquiry.

Chapter 2 - Background

This chapter evaluates the background and literature review while reviewing the relevant theories of the study. It reviews existing literature, highlighting important findings and gaps that this study aims to address. That means that there are very large segments of this chapter that have to be understood before the treatment and analysis that comes next.

Chapter 3 - Methodology

This chapter outlines the research methods used throughout the course of the study. It explains the method step by step, beginning with data collection and ending with data augmentation to ensure a solid dataset. Provides a detailed explaination, how techniques such as Data Splitting and Data Augmentation improve the quality and diversity of the dataset. Specific processes are defined to ensure that the research can be understood easily and can be applied in real time

Chapter 4: Implementation.

This part focuses on the hands-on implementation of the research idea. It offers an overview of the Implementation workflow, it breaks down what each step is and what needs to be done during the actual implementation. The chapter discuss about the coding and computing processes in an organized manner, indicating how data is processed and appraised to obtain the required goals.

Chapter 5: Result Analysis

This chapter gives a thorough understanding of the research outcomes. The report summarizes implementation outcomes and evaluates them to research objectives and questions. Ensuring the ethical aspects of the study was aligned with the discovery process, this chapter also links the importance of responsible research, with which the landscapes of the new ethical research concept are discussed.

Chapter 6: Conclusion and Future Scope

The final chapter showcases the study's findings and results. It presents a desired result that is coherent with the pursued research objectives and appraises how well they were fulfilled. The chapter further reviews The Future Scope of the research, delivering the potential issues on the not thoroughly researched topics and the precise ways to improve the method that will benefit its application.

This planned strategy assures a logical flow and full coverage of all parts of the research, from beginning to end, while also providing readers with a clear and detailed roadmap to follow.

CHAPTER 2 - BACKGROUND

In this chapter, we look at how social media comments affect mental health and urge for the use of sentiment analysis as an appropriate method for evaluating the emotional impact of online interactions. As a result, our focus in this chapter includes an in-depth examination of core theories, a review of relevant literature, the study of sentiment analysis tools, and a discussion of identifying emotional states using machine learning techniques.The study involves using models like Support Vector Classifier, Logistic Regression, and Naive Bayes Classifier for sentiment prediction, as well as comparing how efficient they are in understanding the psychological effects of comments on people. This detailed examination provides a strong framework for our research, supporting our aim to understand how online sentiments lead to mental health difficulties.

2.1 Sentiment Analysis in Social Media

Social Media Sentiments

Social Sentiment is the also known as sentiment analysis or opinion mining, is a approach for understanding and analyzing people's emotions, views, and attitudes toward a particular subject, product, or person. Basically, this is achieved through language analysis and the names and phrases that are in text data, such a social media posting, reviewing, or commenting.

Sentiment analysis, which extracts certain words or phrases using opinion mining, evaluates if the text is good, negative, or neutral. It is an essential component of any advanced analytics.

This method has been invented using a combination of natural language processing and machine learning. It has been previously trained to interpret text and assign calculated scores to social media postings and other types of online material.

Sentiment Analysis Methods: Key Terms and Definitions

- **Analysis at the phrase level.**

  This relates to evaluating sentiment based on text phrases rather than the full message. This content-based analysis can be divided into sentences.

- **Programming Languages**

  This coding languages are used by developers to trigger specific tasks and reactions when people engage with a website or app. Some languages are uniquely suited to AI environments.

- **Parts of speech tagging**

  These tags are used to get more specific while evaluating text. They filter out words depending on their part of speech in a sentence, such as an adjective or negative - both of which are vital in detecting human emotion through expression.

- **Semi-supervised sentiment classification.**

  In, semi-supervised classification uses unlabeled data from online posts or material to interpret the opinions, sentiments, and viewpoints stated.

- **Sentiment Analysis Model**

  This model analyzes a collection of words to identify how someone else feels based on the classifications you "provide" it. These models, for example, can predict whether people are happy, sad, or frustrated in a message provided the developer assigns the classifications based on specific criteria.

- **Polarity**

  Results are normally divided into positive, neutral, or negative words or phrases; but, in some cases, you can extract data that is merely divided into neutral or "polarity" replies since they clearly distinguish between two categories.

- **Sentiment scores**

  This scores vary from -1 to 1, with -1 representing negativity, 0 for neutrality, and 1 for positivity.

- **Rule-based analysis**

  Rule-based systems use an initial set of rules and data particular to conduct certain actions In terms of natural language processing, such as giving a sentiment score during text analysis.

- **Sentiment lexicons**

  These are the word categories linked with various types of moods or emotions; for example, the word "great" would fall into the positive group.
- **Bag of words approach**

  In NLP, this method allows you to look at words based on their structure. Instead of looking at the syntax of a sentence, it can be used to take out words that match tags and assist determine how frequently a word or mood type appears.

How Sentiment Analysis using AI Works

AI integrates natural language processing (NLP) with a natural language API and machine learning algorithms to automatically detect emotion in written content. NLP translates human words into data that machines can understand.

Following text processing, machine learning is used to categorize it. As the software acquires knowledge continuously, data patterns are found in order to create meaningful predictions. It has been trained to classify sentiment in text into multiple categories using relationship (one way).

Why Are Sentiment Analysis Tools Important?

For example, one unfavorable comment about your brand may trigger an internet storm involving both opinion and emotion. Regardless, it can be tough to recover from, even when handled properly by the best advertising teams. This is because a bad tweet, for instance may go viral in a matter of minutes.

Using sentiment analysis tools to track how your audience feels allows the company to better regulate the emotions and conversations surrounding your brand. This way, company won't be responding reactively by attempting to splash flames on short notice, and you'll be able to manage your brand perception carefully on a daily basis.

2.2 Impact of Social Media on Mental Health

Unfortunately, for most people out there, scrolling through Instagram, Facebook, TikTok among other social networks is a daily norm. According to the stats, social media usage in 2023 was approximately 4.9 billion people. According to the survey, a regular user spends 145 minutes on social networks per day. When you are bored, seeing posts from friends or family members can be places or countries away feel like you are closer to them. But the flip side of social media usage is not very good.

According to the (UC Davis Health, 2024) research, social media can have huge detrimental effects o the general health outcomes by causing anxiety, depression, loneliness and even FOMO. These problems are familiar to teens and young people, in particular.

Social media brings endorphins, described as the feel-good chemical or dopamine, and stimulates the brain's reward area similarly to drugs. Thus, when we post something, we get endorphins from our friends' and family's 'likes.' But when we do not receive that injection of self-esteem or validation, it can affect the fulfilling of the self and adequacy of the same.

Filters: Social networks pay much attention to esthetics and sexuality. Most of the current social sites including Snapchat, instagram, tik tok and among others enables one to remove the filter on an image. While filters can be entertaining, when it comes to our physical appearance and our ability to change something just for the sake of concealing imperfections are at hand, the social reality is rather far from that. It can also overpower your subconscience and make you feel uncomfortable and hate the way you look each time you are exposed to these images.

FOMO (Fear of Missing Out): A recent revelation is that, employing social media intensifies feelings of FOMO, amongst many users. That is, making a point of logging in to your friend's or family's social media platforms may make you develop the impression that they are more joyful or living better lives than you are.

Let's mention that social networks have been considered to be a ''highlight reel" where only the best moments in the user's life are shown. Nonetheless, the prospect of gaining exposure to others' highlight reels also leads to the augmentation of comparable feelings of dissatisfaction with one's own ordinary, day-to-day. Which can lead to low self-esteems, anxiety and also make us want to engage in the use of social media more. It also leads to the need to always log in to the sites to see what is current in the world in case one may lag behind.

Cyberbullying: Pew research in the same year revealed that forty—four percent of all internet users in the United States reported they had been harassed online. Cyber bullying is when a child chooses to hurt, tease or harass another child, through the use of an electronic gadget. They can lead serious problems with self-esteem and mental health. We know that school children are vulnerable targets for cyberbullying, as are young people using social media to share painful rumours or lies or abuse.

Figure 4: Dangers of Social Media

2.3 Machine Learning in Sentiment Prediction

Sentiment analysis has also become common due to the increase in the use of different sources of social media interaction because it enables analysis of interaction and provides patterns in communication. SA has to employ NLP techniques to search and analyze emotions, feelings, and opinions hidden in texts or dialogue some of which include handling big and multilingual data; recognizing irony and emojis; distinguishing between positive and negatives sentiments; and selecting the right ML algorithm. Supervised, unsupervised, semi-supervised, deep analysis, or machine learning analysis helps in getting insights from the interaction of users.

How Sentiment Analysis Works?

- **Text Preprocessing**: It is necessary to pre-process the text, that is, to delete all the unnecessary symbols, such as puntuation marks and infrequent words, so-called stop-words.
- **Tokenization**: To analyze text divide the text into more manageable segments of words or tokens.
- **Feature Extraction**: They are basically related to the frequent occurrences of certain words, sequences of n words, or POS tags.
- **Sentiment Classification**: Unfortunately, it is nearly impossible to achieve high-precision comparisons of sentiments; however, one can either employ ML models or opt for ready-made systems for this task.
- **Post-processing:** Gather and improve sentiment analysis overall outcomes.
- **Evaluation**: Assess the model's accuracy, precision, recall, and F1-score.

Sentiment Analysis Types

- **Document-Level**: Determines overall sentiment in an entire text.
- **Sentence-Level**: Analyzes sentiment in individual sentences.
- **Aspect-Based:** Focuses on sentiments about specific features or aspects.
- **Entity-Level**: Targets sentiment related to specific entities like brands or products.
- **Comparative**: Compares sentiments about multiple entities or aspects.

The example shown below is a schematic representation of sentiment analysis on the reviews of three fragrances of perfumes—Lavender, Rose, and Lemon.

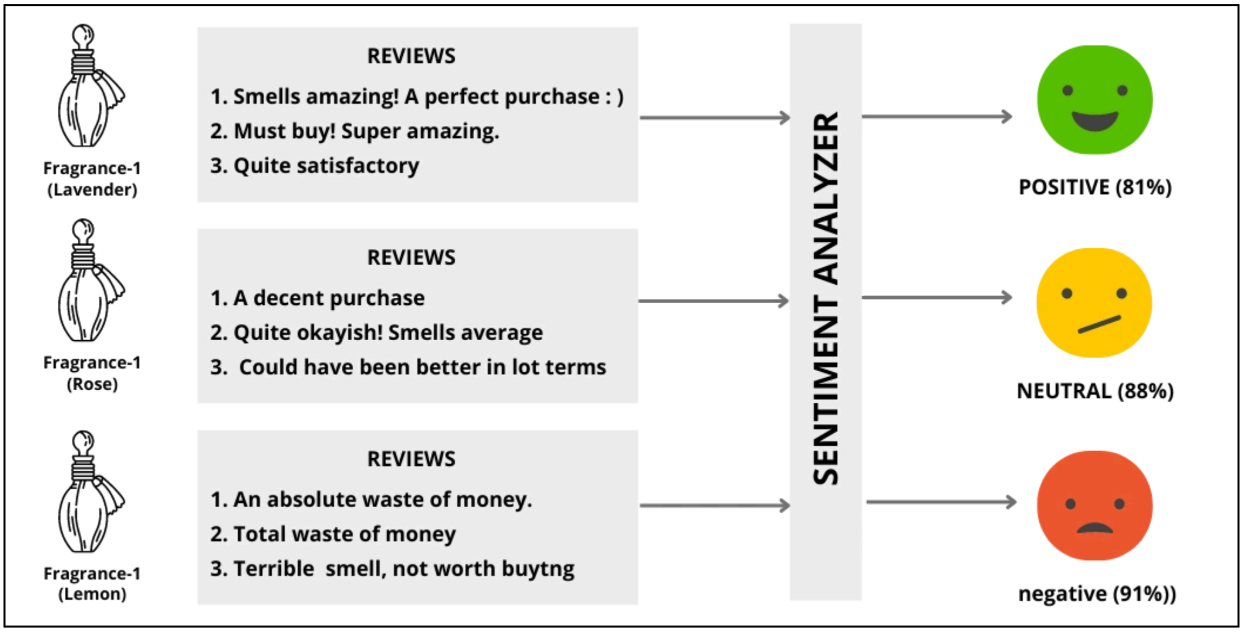

Figure 5: Example of Sentiment Analysis

From these results, it is clearly visible that:

- Fragrance-1 (Lavender) has highly positive reviews by the customers which indicates company can escalate its prices given its popularity.
- Fragrance-2 (Rose) happens to have a neutral outlook amongst the customer which means, company should not change its pricing.
- Fragrance-3 (Lemon) has an overall negative sentiment associated with it thus, company should consider offering a discount on it to balance the scales.

This example conveys how sentiment analysis can help us gain insights into our products/services and help organizations make better decisions.

Use Cases of Sentiment Analysis

- **Social Media Monitoring**: Analyze public opinions about brands or events.
- **Product Reviews**: Assess customer feedback to improve products or services.
- **Stock Prediction**: Use sentiment in news headlines to predict stock trends.

Sentiment Analysis in Python:

Python offers several tools for SA:

1. **TextBlob**: Calculates polarity (sentiment strength) and subjectivity.
2. **VADER**: Ideal for social media text.
3. **Bag of Words**: Converts text to numerical data for ML models.
4. **LSTM models** use deep learning to predict sequences.
5. **Transformers** use advanced deep learning models like with BERT to achieve great accuracy.

Sentiment Analysis is a useful tool for businesses and organizations to gain actionable insights from text data, leading to better decision-making and strategy.

2.4 Text Classification

What is Text Classification?

Text classification is a way to use Machine Learning to sort text into specific categories automatically. It can handle anything from articles and research papers to customer queries and social media posts. This helps businesses organize unstructured data, which makes up the majority of data they collect.

Traditionally, people had to sort this data manually, which was time-consuming and expensive. Now, automated tools use technologies like Natural Language Processing and Machine Learning to analyze and categorize text quickly and accurately.

For example, a tool could recognize positive terms like "user-friendly" in a tweet and tag it as positive feedback.

Why is Text Classification Important?

Text classification helps businesses get more value from their unstructured data by organizing it efficiently. This saves time and allows for better decision-making. Examples of its use include:

- Sorting app crash reports by the type of issue (e.g., "loading time" or "screen freezes").
- Analyzing customer feedback to identify problems, new feature ideas, or trends.

- Monitoring brand mentions in real-time and responding quickly.
- Reducing errors caused by manual work.

How Does Text Classification Work?

There are two methods:

1. Manual: A person reads and categorizes text, which is accurate but slow and costly.
2. Automated: Uses AI, NLP, and Machine Learning to quickly and reliably categorize text.

Automated methods consist of rule-based systems, which sort text according to predetermined rules. For example, if an article includes terms such as "blockchain" or "IT landscape," it is classified as "Industry." Setting up these standards takes time, but it saves effort in the long term.

Key Benefits of Text Classification:

- Identifying Problems: Identify problems that customers have with the product, at first glance.
- Understanding Customers: It is possible to segment audiences by using the language they use so as to address them appropriately.
- Real-Time Insights: Get updated with trends or feedback in a single pull.
- Error Reduction: On the advantage of machines, it should be acknowledged that when trained properly, the machines do not deviate from the set path.

What is Machine Learning-Based Text Classification?

Machine Learning (ML) text classification organizes text into categories by learning patterns from training data instead of relying on human-made rules. It uses algorithms to identify relationships in text and assign tags (categories) based on the input.

For example, if we tag some text as "positive" during training, the system learns to recognize similar patterns and applies the same tag to new text.

How It Works?

1. Feature Extraction: Text is converted into a mathematical form called a vector to train the system. A common method is the "bag of words" approach, where each word's frequency in a set list (lexicon) is counted.
   - Example: For the lexicon "feature, sun, flowers, love," the sentence "I love this feature" becomes: (1, 0, 0, 1).
2. Training the Model: The ML algorithm learns from these vectors paired with their tags (categories). After seeing enough examples, it can make accurate predictions for new text.

Advantages of ML-Based Systems:

- Faster and more accurate than rule-based methods.
- Can adapt to new categories with updated training data.
- Easy to scale and manage.

Common Algorithms Used:

- Support Vector Machines (SVM): Effective for separating text into categories.
- Deep Learning: Advanced models that learn complex patterns.
- Naive Bayes Classifiers: Simple and efficient for many text tasks.
- Combination Systems: Mix rule-based and ML techniques for better results.

Hybrid Systems:

Combining rule-based and machine learning technologies can increase accuracy.
For example:

- Create specific guidelines to correct faults caused by the ML classifier.
- Reduce the workload related to labelling fresh data.
- In simple terms,machine learning-based text classification systems are strong tools that provide flexibility, scalability, and precision.

2.5 RELATED WORK

The combination of solving social media sentiment analysis and the issue of mental health has attracted a lot of interest these years due to the fact that social media platforms plays an increasingly important role in people's lives. This section bring review from different literatures to demonstrate how sentiment analysis perform on social media for identification and prevention of mental health problem.

One of the primary methodologies employed in sentiment analysis is Natural Language Processing (NLP), which has been validated in various contexts, including social media. Tools such as the Valence Aware Dictionary and Sentiment Reasoning (VADER) are specifically designed for analyzing sentiments in social media text, providing quick and accurate assessments of emotional polarity (Huerta et al., 2021). Analyzing large datasets from platforms like Twitter has enabled researchers to draw correlations between social media usage and mental health outcomes, particularly during significant events such as the COVID-19 pandemic (Valdez et al., 2020). For example, research has revealed that large social networks data can be used to contain real time information about people's mood and their psychological state during such situation (Valdez et al., 2020).

Social Media Features and Mental Health Impacts (Beyari H., 2023)

Chatting is a key component of social media that allows users to communicate effectively.

Beyari H's (2023) research focuses on trying to establish the relationship between social media uses and mental health difficulties among young people in Saudi Arabia. Issues have been raised about the possible consequences of social networks for psychological well-being, such as stress, anxiety, and depression, particularly among young people who use them frequently. Beyari H's (2023) study aimed to identify which aspects (likes, comments, followers, and media sharing) have significant negative consequences and investigate the root cause of these effects.

The paper does not make use of machine learning techniques. Instead, it applies the Analytical Hierarchical Process (AHP), a structured decision-making approach that uses pairwise comparisons to determine the relative importance of various social media elements regarding their impact on mental health.

Sentiment Analysis in Social Media Using AI (IntechOpen)

The informal, common language used on social media makes sentiment analysis more difficult. The IntechOpen book by (Victor Rajan K,2024)tackles these issues using machine learning techniques like K-Nearest Neighbors and Support Vector Machines (SVM). These strategies outperform typical NLP tools in classifying sentiment polarity. The use of advanced AI models is very useful for capturing nuanced sentiments and providing actionable insights to enterprises and decision-makers. However, limitations remain in the management of multimodal input, dynamic language evolution, and contextual interpretation. This highlights the importance of adaptable and context-aware models that account for the different linguistic and cultural complications of social media users.

Predicting Depression Through Social Media Data

Some significant information are found out from the social media networks particularly from the twitter regarding the early signs of depression. The research by (De Choudhury *et al.*, 2021) investigates linguistic and behavioural characteristics, discovering that user activity patterns, language use, and ego networks are all key indicators of depression. Machine learning is found suitable for identifying early signs of deterioration and for use in early intervention. However, the factors that make this study limited include; specificity of the platform, absence of historical data and ethnic population. Applying the biochemical information found in social media to build quantifiable therapeutic interventions is needed to translate online observations into functional mental health therapies.

Emotional Tone Prediction in Online Mental Health Communities (Springer, 2023)

Some of the powerful techniques used in analysis are the Recurrent Neural Networks and Gated Recurrent Units (GRUs), are critical for comprehending the emotional tone dynamics in online mental health discussions. The study (Kanaparthi, Patle and Naik, 2023) shows that user interactions on platforms such as Reddit frequently result in favourable emotional tone alterations, with models obtaining a 12.1% increase in accuracy over baselines. The study highlights the usefulness of feature-rich predictive models by using sentiment analysis techniques such as TextBlob and VADER, as well as including DistilBERT embeddings. However, they still have very limited ability in capturing dynamic emotional states, a small amount of features, and require external validation for their application in real-world.

A deep learning model for detecting mental illness from user content on social media

Research by (Kim et al., 2020) developed a deep learning model to detect mental illness by analyzing user-generated content on social media platforms, particularly Reddit. It was trained to sort out the posts into categories related to the certain particular mental disorders such as depression, anxiety, bipolar disorder, schizophrenia and autism. This way, with the help of textual data provided in mental health communities, the model was supposed to predict users' mental states from their postings effectively. The researchers suggested that such a model might be useful as an adjunct in identifying signs of decreased well-being among social media users, and managing the process of their treatment.

After looking at the independent studies, it is possible to notice that some limitations are still left unfilled in the collective studies. Several researchers have also investigated cross-level interactions or effects of implementing multiple features for sentiment analysis, and where this is done, this can sometimes prove counterproductive in that, in their quest to include several variables which they believe will enhance their analysis and produce better results, they end up including factors that actually serve more as a nuisance and results in an exceedingly complex model thus their goal is not achieved. This study focuses on one critical feature: interactions that comprise the user comments or statements with focus on the ways these impacts the mental health states.

Compared to other similar research, there is a unique aspect to our dataset. Despite the fact that it collects information from several applications, the cross-platform analysis is not specifically highlighted. Instead, it provides a structured and straightforward framework, categorizing mental health statuses into six key labels: They include Normal, Depression, Suicidal, Anxiety Stress, Bipolar and Personality Disorder. These categories seem to encompass all the major types of mental disorders that are most widespread among users of the current social networks and, therefore, seem to be the most suitable for further examination.

In this context, by prioritising textual data, which is a core to sentiment analysis, and orientation on these decisive mental health states, this research should provide significant findings. Moreover, this study does not complicate it by incorporating highly complex machine learning models because doing so often results in loss of accuracy. This kind of approach — focused on a rather limited set of measures and simple models — guarantees the stable and efficient investigation of the connection between comments on social networks and the level of psychological distress.

Figure 6: Different Social Medias

CHAPTER 3 - METHODOLOGY

3.1 Data Collection

Data gathering is one of the most important and first phases in developing a machine learning model. While there are several sources and datasets accessible for sentiment analysis, locating relevant and correct data proved to be the most time-consuming element of this study.

Sentiment analysis datasets focus mainly on social media platforms like Twitter and Reddit. However, for this study, we intended to discover a basic and broad dataset, including a wide spectrum of mental health conditions.

The resulting dataset is a thoroughly curated collection of mental health conditions, labeled with textual remarks. It combines raw data from many sources, which has been cleansed and collated to create a reliable resource for constructing chatbots and conducting sentiment analysis.

This dataset, available on Kaggle, combines information from various Kaggle datasets. It includes the following features:

- **unique_id:** A unique identifier for each entry.
- **Statement**: The textual data or post.
- **Mental Health Status**: The tagged mental health status associated with the statement.

The dataset contains 51,074 unique entries. Both columns are essential and interlinked, where the "Statement" represents social media sentiments expressed by individuals, and the "Mental Health Status" indicates the impact of these sentiments.The statuses are categorized into six states: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar & Personality Disorder. The data is gathered from multiple social media platforms which involves the discussion on Reddit, and Tweets. Each entry is tagged with a specific mental health status, making this dataset an invaluable resource for:

1. Developing intelligent mental health chatbots.
2. Involved in carrying out extensive and intensive sentiment analysis.
3. Investigation, research and studies on patterns concerning mental health.

4. Conducting research and studies on mental health trends.

This dataset is ideal for training machine learning models that aim to understand and predict mental health conditions based on textual data.

In conclusion, The collected dataset is a versatile source to build up for the further research for mental health as well as for sentiment analysis.

It enables the creation of machines that are capable of better identify and handle mental health difficulties by combining varied data sources and precisely categorizing mental health conditions.

3.2 Data Preprocessing

Exploratory data analysis is an initial step that prevents the usage of improper data and data preprocessing also contributes a lot to the success of the whole procedure. In the following part, we describe the detailed approach to data preprocessing performed in this study, which includes both sentiment analysis of social media data and mental health states. The data is preprocessed with relevant Python libraries needed in the code.

3.3 Data Cleaning and Normalization

Especially when it comes to model building data cleaning is a crucial step. Cleaning of data varies according to the type of the data set, and if it is textual data then Cleaning of data can be done by using Text Cleaning Methods.

There are different approaches to apply on text data and we should be careful while applying and selecting the steps. The steps of the textual data processing depend on the specific use cases.

Unlike some other text processing techniques where we may eliminate emojis or emoticons as their primary function maybe to express emotions, we have to eliminate unwanted links, tags and stop words.

Text Cleaning

Most common methods of text cleaning process are as follows:

- Lowecasing the data
- Removing Puncuatations
- Removing Numbers
- Removing extra space
- Replacing the repetitions of punctations
- Removing Emojis
- Removing emoticons
- Removing Contractions

The other techniques of text cleaning are

- **Tokenization**: This is a process of segmenting a flow of conversations up into words, phrases, symbols, or other other meaningful elements called tokens. The list of tokens become input for further process.NLTK Library has word_tokenize and sent_tokenize to easily break a stream of text into a list of words or sentences, respectively.
- **Word Stemming/Lemmatization**: The aim of both processes is the same, reducing the inflectional forms of each word into a common base or root. Lemmatization is closely related to stemming. The distinction is that, a stemmer works on individual words, without the benefit of any surrounding text, and therefore cannot tell when two words with exactly the same spelling are spelled differently for different parts of speech. Nevertheless they are far simpler to program as well as run quicker and the difference in precision may not be significant in certain programs.

3.4 Feature Extraction

The two most common approaches for extracting features from text, or transforming text input (strings) into numeric features so that a machine learning model may be trained, are Bag of Words (also known as CountVectorizer) and Tf-IDF.

- **Bag of Words**

  A bag of words (BoW) is a technique which is used to translate text data into numeric features. It requires formation of known word list in the corpus and formation of point vector for all documents, which show frequency of usage of each of them.
- **TF-IDF Vectorization**

  TF-IDF is termed as Term Frequency-inverse Document Frequency and it is another variety of representation of the text by the numerical features. TF-IDF overcome limitations of the Bag of Words (BoW) model.

The TF-IDF model share similarities with its bag of words model but it has an additional factor, which is the frequency of the words in the document and its inverse document frequency. By this the TF-IDF model has better chances of picking important words than the bag of words model.

Converts text data into numerical features using the TfidfVectorizer:

  - Minimum document frequency (min_df=0.0004) excludes rare terms.
  - N-grams (ngram_range=(1,3)) capture context by including unigrams, bigrams, and trigrams.

3.5 Encoding

Label encode the target variable— The data set contain Categorical data of the string format for the target variables, and, therefore the target needs to be label encode. This step involves the mapping of categorical labels (status) to binary integers (1 for mental health conditions and 0 for normal).

3.6 Data Augmentation

The process of data augmentation in machine learning is the process of synthesizing additional data in order to expand the training set to permit better learning. It is specially useful for the case of situations where some classes have far less samples than the other classes. In text classification tasks, data augmentation can include methods such as synonym replacement, back-translation, or contextual word insertion to expose models to more linguistic variability.For this study, the dataset display a large class imbalance. For example, the "Normal" class consisted of 16,343 samples, but the "Personality Disorder" class has only 1,077. Other classes showed varied degree of imbalance emphasizing on data balancing techniques. For this, Synthetic Minority Oversampling Technique (SMOTE) was applied.

SMOTE for Balancing the Dataset

SMOTE is a technique that generates synthetic samples to balance imbalanced datasets. It works by:

1. Picking some random sample from the minority class.
2. Identifying its nearest neighbors in feature space.
3. Generating synthetic samples along the line segments connecting the sample and its neighbors.

This method increases the representation of the minority class, making the dataset more balanced and improving the model's ability to generalize. In this data set, therefore, the use of SMOTE was able to make the underrepresented statuses such as the “Personality Disorder” and others acquire enough representation so as to ensure fair and more credible model means.

3.7 Data Splitting

Data splitting is required when dividing the given data into two or more subsets to train, test and evaluate a model.

In our scenario, we have used the following components

StratifiedKFold

A cross-validation technique that divides the dataset into n_splits folds while maintaining the number of target class labels in each fold.

This ensures that the class distributions in the training and test sets indicate the entire dataset, which is essential when working with imbalanced datasets.

Parameters used in the code :

- n_splits=2: The dataset is split into 2 folds, meaning the training and testing process will run twice, with each fold taking a turn as the test set.

- shuffle=True: Randomly shuffles the data before splitting, ensuring more randomness in the training and testing sets.
- random_state=1: Ensures reproducibility by setting a fixed seed for shuffling.
- **kf. split(data_tfidf, Y)**: Splits the dataset data_tfidf (features) and Y (labels) into training and testing indices for each fold.

By using the **Stratified K-Fold Cross-Validation,** it helps in:

- Evaluate Model Robustness: Training and testing the model on several folds ensures that the model's performance is not heavily dependent on a certain subset of data.
- Handle Imbalanced Classes: Stratification ensures that each fold has almost the same class distribution as the entire dataset, which is vital for imbalanced datasets.
- Reproducibility: The random_state produces consistent results across runs.

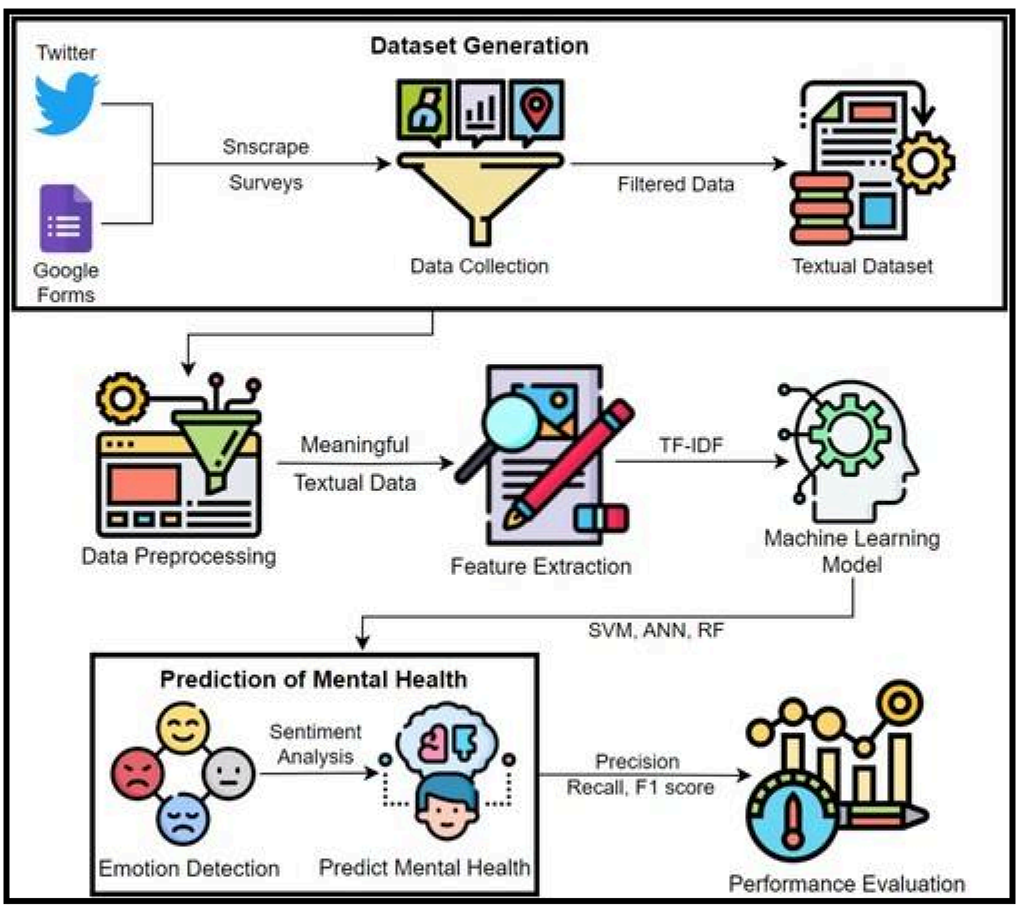

Figure 7: Sentiment Analysis Process

3.8 Machine Learning Models For Sentiment Analysis

1. Support Vector Classifier

A Support Vector Classifier (SVC) is an effective supervised learning technique for binary and multi-class classification. It works by identifying a hyperplane that divides data points into distinct groups and maximizing the margin between them for better precision, even with unseen data.

Key Components of SVC:

- **Hyperplane**: A decision boundary that separates the classes in the feature space.
- **Support vectors** are the data points closer to the hyperplane, determining the margin by optimizing the margin process to find the optimal hyperplane.
- **Kernel Trick**: Maps non-linear data into a higher-dimensional space where classes can be linearly separated.

Application of SVC:

Text Classification: Used for sentiment analysis, spam detection, document classification, and topic categorization based on textual data.

To utilize Support Vector Classifier effectively, the following steps are typically followed:

- **Data Preprocessing**: Cleans and scales the data.
- **Model Training**: Divide the data and train using optimization approaches.
- **Hyperparameter tuning** involves adjusting parameters such as kernel type and regularization.
- **Model evaluation** measures performance in terms of accuracy, precision, recall, and other factors.
- **Model Deployment**: Use the model to anticipate and track its performance.

SVC is an effective classification method with high flexibility and performance across various areas, including image recognition and medical diagnostics.

2. Naive Bayes Classifier

The Naive Bayes classifier(NBC)is a probabilistic machine learning model widely used for text classification tasks. Despite its seemingly simplistic name, its effectiveness stems from its strong theoretical foundation and ability to handle high-dimensional text data efficiently. It's particularly effective with high-dimensional data and can handle large datasets efficiently.The method is easy to use, fast and more importantly efficient on small datasets and hence often used especially when the resources in math are limited.

Bayes Theorem

A theorem signifies a mathematical idea that helps us to measure the likelihood of an event occurring based on previous knowledge or information. In the context of text classification, Bayes' Theorem calculates the probability that a piece of text belongs to a particular category or class based on the occurrence of certain words or phrases within the text.

Naivety Assumption: The "naive" aspect lies in its assumption that word occurrences are independent of each other within a class. While this assumption rarely holds perfectly true, it surprisingly leads to surprisingly strong performance in many real-world scenarios.Flexibility: NB works well with multinomial and Bernoulli word representations, adapting to different text characteristics. Multinomial captures word frequency within a document, while Bernoulli considers mere presence or absence. NB requires minimal feature engineering and training time, making it ideal for applications requiring fast predictions and quick adaptation to new data.

3. Logistic Regression

Logistic regression is a statistical framework that uses the logistic (sigmoid) function in order to predict binary events. In su model, the dependent variable can have values of 0 or 1, and there are independent variables which impact the prediction. Logistic regression is a typical method for predicting outcomes with two possible classes.

The logistic model estimates the relationship between one dependent variable and independent variables. It uses the logarithm of odds (value labeled "1") as dependent variables, which is a linear combination of one or more independent variables i.e. predictors. The logistic function converts log odds to probability. The unit of measurement for log odds is called a 'logit ', derived from the 'logistic unit '. While logistic regression is not inherently a classifier, It can be used for a variety of applications, including binary text classification. It empowers you to model the probability of output in terms of input, giving you a versatile tool in your data science arsenal.

In binary classification, inputs can be classified into two, a probability greater than that value as one category and those with a probability less than that value as the other. This is achieved by the maximum likelihood estimate method.

Steps to perform logistic regression

1. Collect labelled data, which consists of features (an object's factors) and labels (the object's kind or category).
2. Convert labeled data to encoded data: Because calculations are typically performed using numbers, the data is encoded in numerical format.
3. Create a feature set. A set of features is formed by grouping relevant traits.
4. Divide the data into train and test sets: Typically, 80% are utilized for training and 20% for testing, although this can be changed as needed.
5. Use the training data to train the classifier.
6. Evaluate the trained classifier using the testing data.

3.9 Tools and Technologies

For this project, various Python libraries were utilized to implement Sentiment Analysis effectively using Google Colab, some of the libraries are mentioned below

- Text Preprocessing:
  - re: For regular expressions to clean text.
  - nltk: For natural language processing tasks like tokenization, stopword removal, and lemmatization.
  - BeautifulSoup: To clean HTML content.

- lxml - for parsing and manipulating XML and HTML documents.
- Data Manipulation:
  - numpy and pandas: For handling and analyzing structured data.
- Feature Extraction:
  - TfidfVectorizer (from sklearn): To convert text data into numerical features using TF-IDF.
- Model Development:
  - StratifiedKFold: For creating balanced train-test splits.
  - LinearSVC and MultinomialNB: For building classification models.
  - SMOTE: To handle imbalanced datasets by oversampling minority classes.
- Evaluation:
  - metrics and confusion_matrix (from sklearn): For performance evaluation.
  - ConfusionMatrixDisplay: For visualizing the confusion matrix.
- Visualization:
  - matplotlib: To create plots for analysis and results.
- Model Deployment:
  - joblib: For saving and loading trained models.

These Python libraries are used in the code, along with additional libraries that are also imported.

CHAPTER 4 - IMPLEMENTATION

Implementation is an important phase of any machine learning project since it takes place between the gaining of knowledge and actual application. This approach involves a process of converting all the research aims and goals into systematic series of procedures for instance data acquisition and pre-processing, feature engineering and selection, modeling methods selection, model training and assessment. The implementation of this chapter starts with data preprocessing, including imbalance checking and correcting, applied to the dataset by using SMOTE technique and the feature extraction using TF-IDF to transform textual information to numerical form. Each stage is carefully planned to match the overall goal of creating an accurate and dependable system for mental health sentiment analysis.

1. Data Loading and Initial Analysis

The dataset was imported, and the CSV file was loaded using Python libraries. A general introduction of the dataset was done using exploratory functions including dataset.shape and dataset.info() to establish basic figures like the number of rows and columns of the given dataset besides other basic metadata. This gives an initial insight into the formation and contents of the dataset.

2. Handling Missing Values

A check for null values is conducted to analyse if there are any missing values that may have a negative impact on the target models. If null values exist, they need to be deleted to make sure the quality of the data. After removal, the data count is printed again to confirm that all missing values have been successfully handled.

3. Data Visualization

For the better understanding the distribution of mental health statuses within the dataset:

- A **pie chart** is created to visualize the percentage distribution of each status.

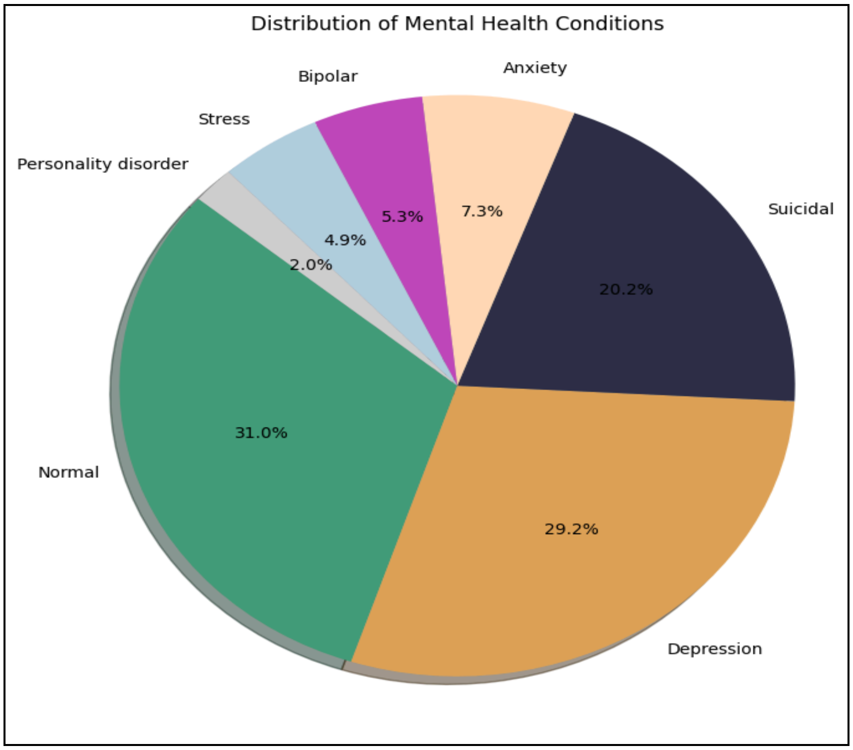

Figure 8: Distribution of Mental Health Conditions

- A bar chart is generated to represent the statistics for the encoded values. These visualizations provided insight into class imbalances and the dataset's composition.

4. Data Mapping

Mapping was performed for the target variable (mental health status), which is categorical. Binary encoding is used to transform the data from categorical to numerical.

- The "Normal" status was mapped to 0, representing the negative class.
- All other mental health statuses were mapped to 1, representing the positive class.
- This binary mapping was chosen because the focus is to identify mental health issues (positives) rather than normal cases (negatives).

Also to minimize prediction errors:

- False negatives (mentally ill individuals incorrectly classified as "Normal") were given priority in order to reduce their occurrence.

- False positives (normal people falsely labeled as having mental health problems) were also monitored closely to maintain balance.

5. Library Installation and Importation

All necessary libraries for data preprocessing, model training, and evaluation were installed and imported. These included libraries for handling text data, implementing machine learning models, and evaluating their performance.

6. Text Cleaning

Text cleaning is an essential step for text classification. Different processes were applied to the statement column in the dataset(which consisted of comments) such as

- Stopword Removal: Eliminating commonly used words that do not contribute significantly to the sentiment or meaning of the text.
- Special Character Removal: Removing symbols, punctuation, and other non-alphanumeric characters.
- Word Lemmatization: Reducing words to their base or root form to ensure consistency.
- After cleaning, the processed data was saved into a new CSV file to maintain a record of the transformed dataset.

Other cleaning processes were also performed to obtain a clean and proper text.

7. Feature Extraction

The next step in Text Mining process formula is known as TF-IDF (Term Frequency-Inverse Document Frequency). This works successfully to transform most of the textual data into numerical features, which weighs the importance of words according to the dataset. The resulting numerical representation is used for text classification.

8. Model Implementation

Once data preprocessing, encoding, and feature extraction were complete, machine learning models are implemented. The models used for the sentiment analysis are:

- Support Vector Classifier (SVC)
- Multinomial Naive Bayes (MNB)
- Logistic Regression

These models are chose due to their suitability for text classification tasks, particularly for sentiment analysis.

9. Evaluation Metrics

As a result, two key metrics for evaluation, Precision and Recall, have been selected to assess the performance of the suggested models. The reason for choosing these models is because:

- Precision: Important to minimize false positives, avoiding normal individuals being misclassified as having mental health issues.
- Recall: It is important to minimize false negatives, ensuring individuals with mental health issues are accurately identified.

Because of the uneven structure of the dataset, accuracy was not prioritized, and the confusion matrix is also not considered in order to prioritize false negatives and false positives.

10. Model Evaluation

Data has been split using the Stratified K-Fold approach using two folds. Even though additional folds were tested, the findings stated no substantial improvements, therefore two folds have been maintained.

Both models SVC and Multinomial Naive Bayes performed training and testing, resulting the desired precision and recall scores. By frequently fine-tuning both the SVC and Multinomial Naive Bayes models, followed by training and testing, the implementation shows consistent performance in diagnosing mental health statuses based on textual data (statements representing user comments).

A slightly different methodology was used for logistic regression, which included creating a pipeline. This approach was able to merge SMOTE and Logistic Regression into one pipeline, in a more seamless manner. The pipeline ensures:

- Class Balancing: During the pipeline execution, SMOTE was used for oversampling minority classes and generate synthetic examples from the training data.
- Classification: The continuous data was they divided into balanced datasets using logistic regression.

This strategy turned out to be a very good solution to address both the preprocessing and the training of the model, and was very satisfactory in balancing the, in essence, unbalanced data set. The pipeline technique greatly reduced the work flow and ensured that class balance was implemented correctly at each stage of the evaluation.

CHAPTER 5 - RESULTS AND DISCUSSION

This chapter concentrates on the findings of our sentiment analysis, particularly the impact of social media comments on mental health. The study carefully assesses the results produced using the stated approach, examining the effects of sentiment analysis on identifying mental health issues. The discussion also explores the more significant implications for mental health support systems and risk-mitigation techniques for improper social media content.

Social Media Sentiment Analysis

Our sentiment analysis model correctly identified mental health states from social media comments with a precision & recall of more than 80%. This proves that the used model helps to identify sentiments and predict the mental state of users from comments on their posts. The training approach have shown concrete improvements in the training structure with the help of general benchmark criteria for performance, encouraging improvements to always be possible.

Key findings

Model Accuracy: Overall precision & recall of more than 80% was achieved, which indicates reliable classification of sentiment capabilities.

Confusion Matrix Analysis: False positives and negatives were carefully evaluated to improve model consistency and reliability.

5.1 RESULTS

Results for SVC Model

Below is the confusion matrix for iteration 1, displaying the count of false positives and false negatives, which are primary evaluation metrics in this study.

Output of Iteration 1

Precision: 0.9343336934694713

Recall: 0.9337955203413073

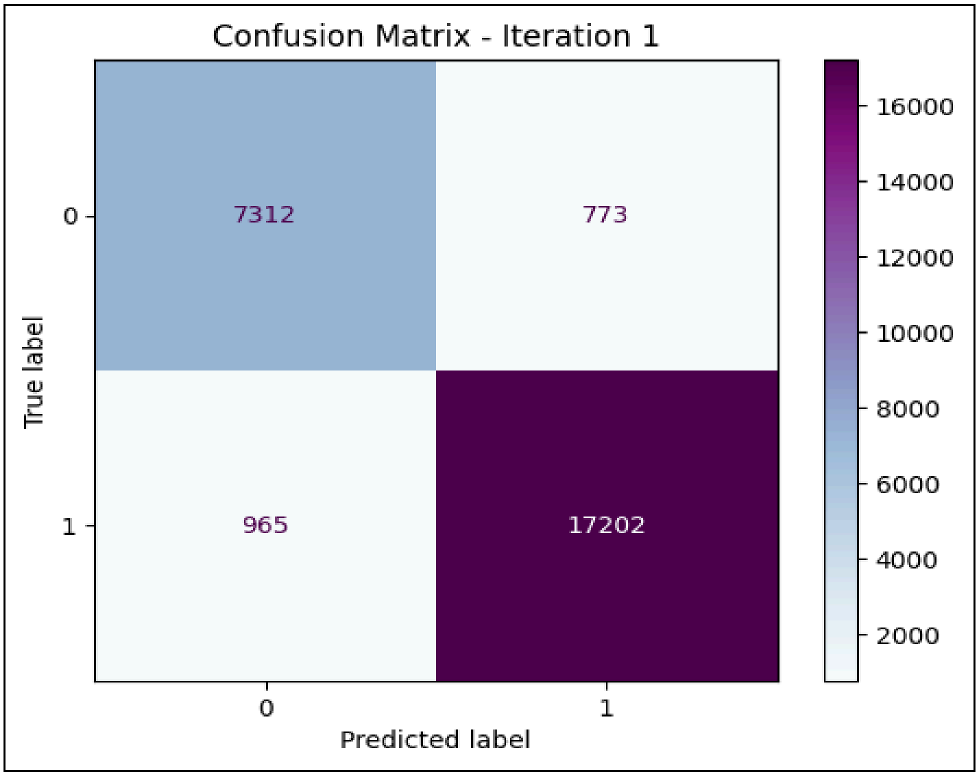

Figure 9: Iteration 1 Confusion Matrix for SVC

The second iteration of the SVC

Output of Iteration 2

Precision: 0.9348210266824994

Recall: 0.934557367057748

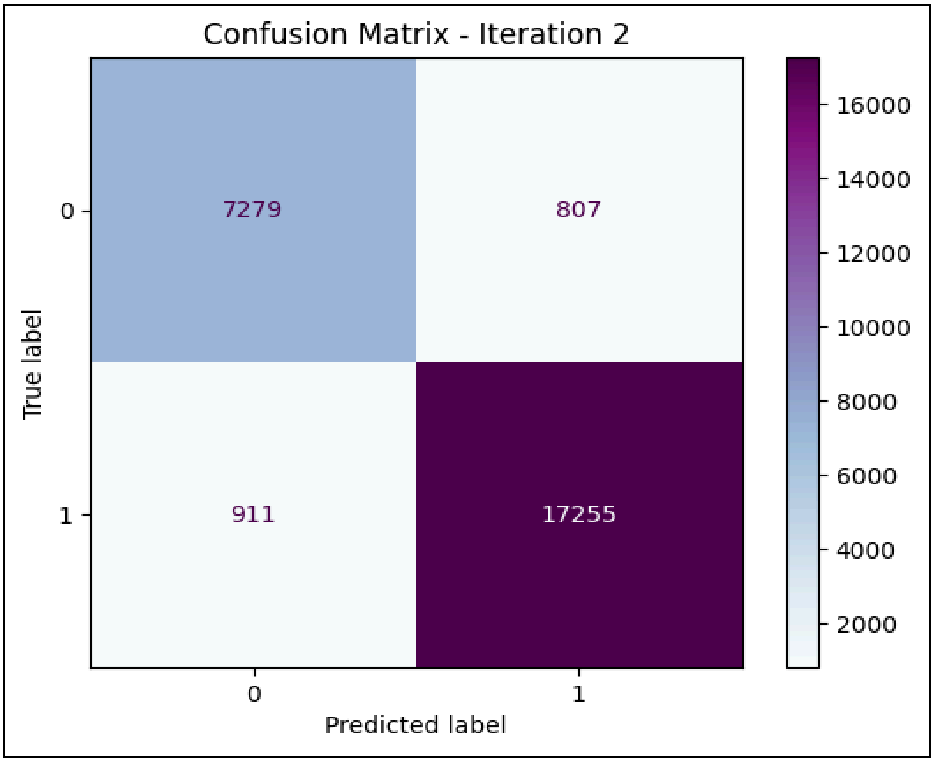

Figure 10: Iteration 2 Confusion Matrix for SVC

The results from both iterations are nearly the same, with a precision and recall score of 93%, indicating remarkable efficiency in evaluating social media emotions. Such a high score shows that the model is extremely good at predicting sentiments, providing significant insights to help implement preventative strategies for related mental health conditions.

Results for NBC Model

In NBC, after implementing the SMOTE technique on training data, the model showed a significant increase in precision and recall.

Below are the 2 iterations with their Precision, Recall and confusion matrix score

Output of Iteration 1

Precision: 0.8883756539152784

Recall: 0.8897988724668596

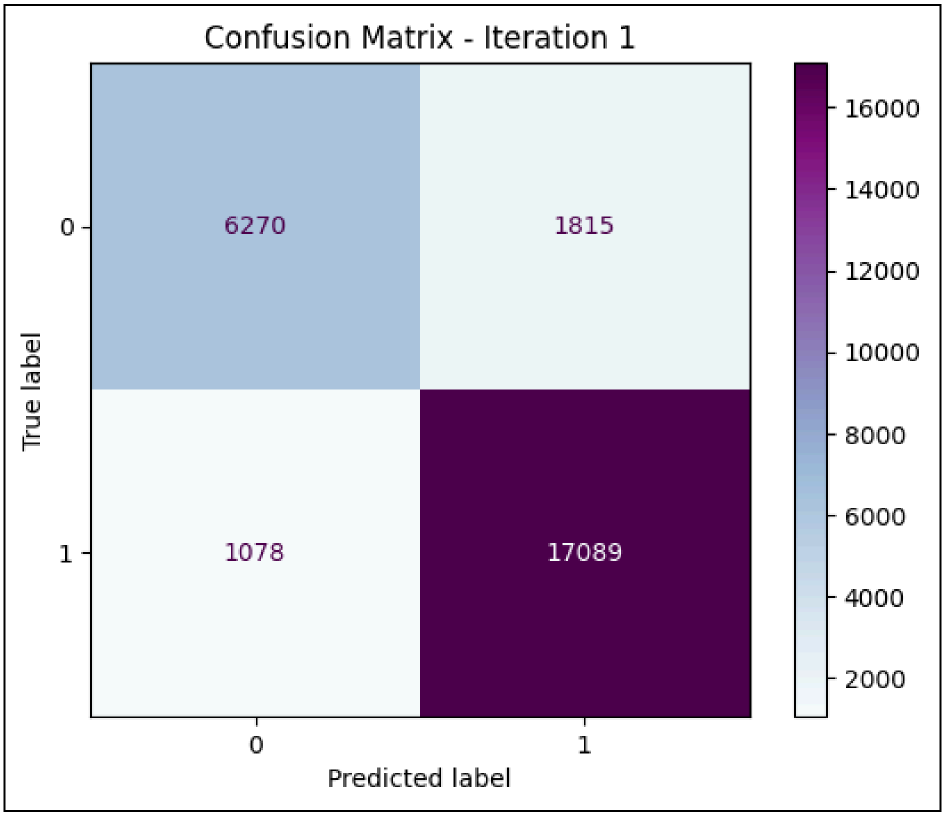

Figure 11: Iteration 1 Confusion Matrix for NBC

Output of Iteration 2

Precision: 0.8912236133101724

Recall: 0.892541520646046

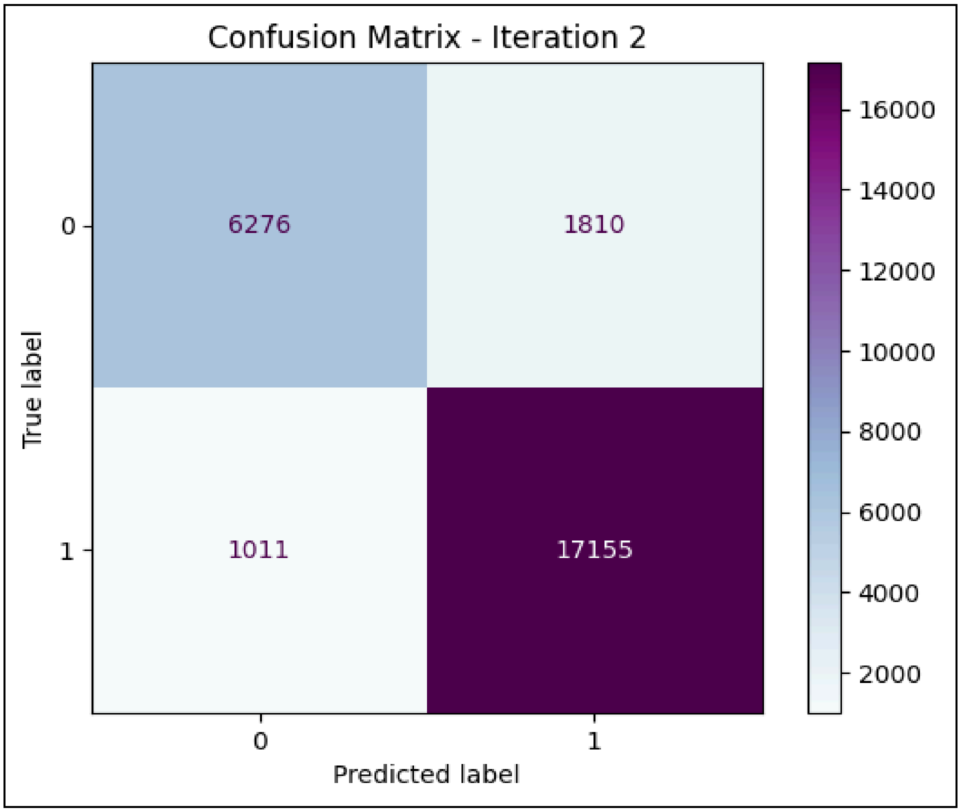

Figure 12: Iteration 2 Confusion Matrix for NBC

The Naive Bayes classifier's results vary slightly throughout both iterations, with precision and recall scores ranging from 88.8% to 89.2%. These outcomes highlight the model's reliable and consistent performance in evaluating social media emotions.

Result of Logistic Regression

In the case of Logistic Regression, both precision and recall achieved identical 93% scores, the same as the SVC model. This consistency suggests that this model is also effective at sentiment prediction, strengthening its applicability for evaluating social media sentiments.

The confusion matrix provides a detailed analysis of the model's classification performance. It shows strengths in accurately predicting feelings while exposing places for improvement, particularly in eliminating false negatives.

5.2 Discussion of Findings

5.2.1 Interpretation of Findings

The high accuracy rates of all models mean that a significant number of expected mental health issues were correctly identified, minimizing false positives. Likewise, high recall values also show efficient identification of most possible mental health issues hence low false negatives.

The evaluation of the confusion matrices showed that misclassifications occurred mostly between closely related categories, such as "Anxiety" and "Stress." Considering the states' common linguistic traits, this overlap is to be anticipated. The findings confirm that textual data alone can provide reliable insights into mental health when analyzed using appropriate models.

5.2.2 Relevance to Mental Health

Unlike many prior studies that focus on one or two emotions, this research stands out by categorizing six distinct mental health states: Normal, Depression, Suicidal, Anxiety, Stress, Bipolar & Personality Disorder. This granularity offers a more nuanced understanding of social media's impact on mental health.

Furthermore, the comments in the dataset suggest that textual comments give prominence to direct expressions on the emotional states of users. Having limited the feature set to text comments for analysis, avoided the potential noise from overcomplicating feature selection, demonstrating that simplicity can yield effective results.

5.2.3 Practical Implications

The findings have significant real-world applications:

- **Mental Health Monitoring**: The models can be used as foundation for more tools that track and forecast mental health problems in real time and thus provide timely support.
- **Support Systems**: Information derived from this research may be useful in the design of appropriate support structures and tools related to particular mental disorders.

CHAPTER 6 - CONCLUSION & FUTURE SCOPE

This chapter concludes all of the work done to get what was intended. The study looks at the predictive value of three different models for social sentiment analysis: SVC, Logistic Regression, and Naive Bayes. The results of these models show strong performance, making them suitable for analyzing and understanding social media sentiments efficiently. Both SVC and Logistic Regression achieved excellent results, with similar precision and recall scores of 93%, proving their accuracy and dependability in sentiment prediction tasks. Although slightly behind, Naive Bayes maintained good and consistent performance, with precision and recall scores ranging from 88.8% to 89.2%. These findings demonstrate the applicability of such models in tackling hard problems, for example determining people's state of mind in social media which is invaluable in gaining insightful understanding of societal needs and wellbeing. The capacity of these models to reliably anticipate mindsets makes them important tools in the larger context of mental health analysis and intervention efforts.

6.1 Future Scope

Although the results are respectable, improvements can be made. To begin with, additional works may consider the aspect of hyper-parameter tuning and optimization to enhance the performance of all three models and quite possibly bridge the gap between Naive Bayes and the other classifiers. Aggregated approaches, such as combining many models, could be investigated to take advantage of different algorithms and thereby increasing the accuracy of predictions.

6.2 Challenges and Limitations

- **Class Imbalance:**
  Looking at the data distribution, dataset exhibited an imbalance with the number of samples in the categories such as 'Suicidal' and 'Bipolar' as it was very limited. To address this problem, approaches, such as SMOTE, were used; however, the problem of overfitting in the minority classes persisted.

- **Lack of Metadata**:

  These features consist of platform identification or temporal regulation; therefore, it is impossible to examine the cross-platform or temporal variance of emotional expression.

- **Bias in Data**:

  This paper identified one major weakness inherent in the data collection process as well as the manual labeling that can result in bias that may affect model performance in the Minority class.

Integration of much more diverse and larger datasets covering a wider variety of social media networks, languages, and cultural settings is another important aspect to be addressed.

This would enhance the models generalisation capability to make the models more available and useful in all groups.

Advanced deep learning algorithm with the help of transformer-based model like BERT or GPT can help in large improvement of both precision as well as recall in cases of computing subtle and complex sentiments. These are far fetched in the context and semantic understanding bracket, which results to more intelligence of user emotions.

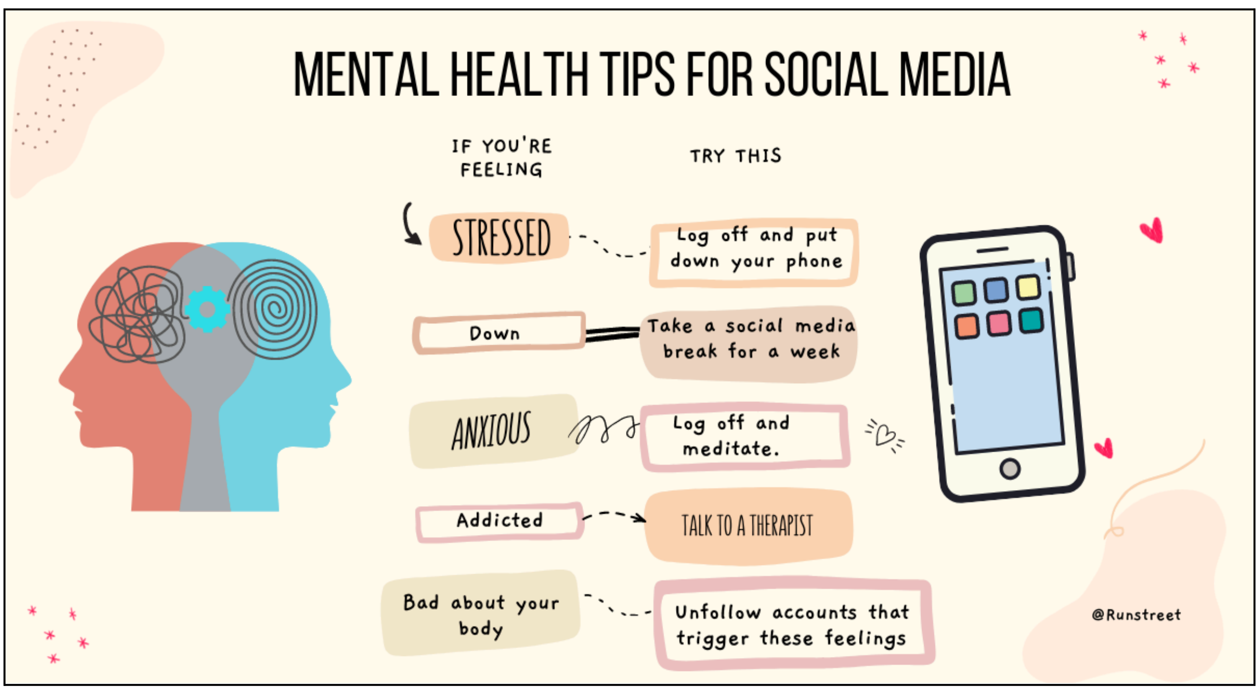

Figure 13 : Tips to Avoid Social Media

Another interesting area of interest is real-time sentiment analysis. Applying these models to real-time data flows allows experts to track and analyze sentiment dynamics in various situations, thus providing relevant solutions when needed most.This ability would be very useful in other competitions like crisis response management, where detecting negative attitudes would need an active approach of breaking the effect.

Ultimately, if these models are used in systems monitoring mental health, such as smartphone applications or social media, it could help in facilitating timely alerts to those at risk of suffering from mental illnesses.

This integration could open the way for specific interventions and greater social programs to improve emotional well-being in the digital era.

In solving these possible avenues, the study could contribute towards building more robust, wide-ranging and preemptive strategies for comprehending and addressing social media attitudes, thereby improving individuals' mental health and well-being in an ever-connected world because social media is so integrated into their lives.

REFERENCES

Center for Research, n.d. **How social media affects mental health**. [online] Available at: https://www.center4research.org/social-media-affects-mental-health/

Klipfolio, n.d. **Social sentiment KPI examples**. [online] Available at: https://www.klipfolio.com/resources/kpi-examples/social-media/social-sentiment

Meltwater, n.d. **How to analyse sentiment with media intelligence**. [online] Available at: https://www.meltwater.com/en/blog/analyse-sentiment-with-media-intelligence

Kaggle, n.d. **Sentiment analysis for mental health dataset**. [online] Available at: https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data

Beyari, H., 2023. The relationship between social media and the increase in mental health problems. *International Journal of Environmental Research and Public Health*, 20(3), p.2383.

Huerta, D., Hawkins, J., Brownstein, J., & Hswen, Y. (2021). Exploring discussions of health and risk and public sentiment in massachusetts during covid-19 pandemic mandate implementation: a twitter analysis. SSM - Population Health, 15, 100851. https://doi.org/10.1016/j.ssmph.2021.100851

Valdez, D., Thij, M., Bathina, K., Rutter, L., & Bollen, J. (2020). Social media insights into us mental health during the covid-19 pandemic: longitudinal analysis of twitter data. Journal of Medical Internet Research, 22(12), e21418. https://doi.org/10.2196/21418

IntechOpen, n.d. **Chapter on text classification and machine learning**. [online] Available at: https://www.intechopen.com/chapters/88300

AAAI, 2023. **Exploring advanced social media sentiment analysis techniques**. [online] Available at: https://ojs.aaai.org/index.php/ICWSM/article/view/14432

SpringerLink, 2023. **Deep learning methods for media analysis**. [online] Available at: https://link.springer.com/content/pdf/10.1007/s11042-023-15316-x.pdf

Nature, 2020. **A study on media's role in influencing mental health**. *Scientific Reports*. [online] Available at: https://www.nature.com/articles/s41598-020-68764-y?utm_source=chatgpt.com

DataCamp, n.d. **Text classification tutorial with Python**. [online] Available at: https://www.datacamp.com/tutorial/text-classification-python

UMA, 2024. **Support vector classifier (SVC): A comprehensive guide**. [online] Available at: https://lp2m.uma.ac.id/2024/01/20/support-vector-classifier-svc-a-comprehensive-guide/

Medium, 2024. **Simple guide to text classification: NLP using SVM and Naive Bayes with Python**. [online] Available at: https://medium.com/@bedigunjit/simple-guide-to-text-classification-nlp-using-svm-and-naive-bayes-with-python-421db3a72d34

UC Davis Health, 2024. **Social media's impact on our mental health and tips to use it safely**. [online] Available at: https://health.ucdavis.edu/blog/cultivating-health/social-medias-impact-our-mental-health-and-tips-to-use-it-safely/2024/05#:~:text=Mental%20health%20impacts,chemical%22%20linked%20to%20pleasurable%20activities.

Levity, n.d. **Blog on text classification techniques**. [online] Available at: https://levity.ai/blog/text-classification

SpotIntelligence, 2023. **Logistic regression in text classification using Python**. [online] Available at: https://spotintelligence.com/2023/02/22/logistic-regression-text-classification-python/

Analytics Vidhya, 2022. **Text cleaning methods in NLP**. [online] Available at: https://www.analyticsvidhya.com/blog/2022/01/text-cleaning-methods-in-nlp/

Khoros Staff (2024) 'How to perform social media sentiment analysis', *Khoros Blog*, 17 January. Available at: https://khoros.com/blog/social-media-sentiment-analysis

Go, S. (2024) 'What Is Sentiment Analysis Marketing? Tips, Tools, and Techniques', *SEMrush Blog*, 22 March. Available at: https://www.semrush.com/blog/sentiment-analysis-marketing/

APPENDIX A : CODE SNIPPETS

This appendix briefly discusses the technology behind the system with snippets of code that powers the performances of the mental health assessment system that is based on sentiment analysis. The included snippets show the techniques used by models to accomplish the goals of the study, in terms of identifying the effects of social media comments on mental health.

1. Implementation of Support Vector Classifier (SVC) Model

```
# Implementing Support Vector Classifier
svc_clf = LinearSVC() # kernel = 'linear' and C = 1

# Running cross-validation
from sklearn.model_selection import StratifiedKFold
from sklearn import metrics
import numpy as np

kf = StratifiedKFold(n_splits=2, shuffle=True, random_state=1)  # 5-fold cross-validation
precisions = []  # For precision
recalls = []  # For recall
iteration = 0
smote = SMOTE(random_state = 101)
for train_index, test_index in kf.split(data_tfidf, Y):
    iteration += 1
    print("Iteration ", iteration)

    # Splitting data
    X_train, Y_train = data_tfidf[train_index], Y[train_index]
    X_test, Y_test = data_tfidf[test_index], Y[test_index]

    # Fitting the model
    svc_clf.fit(X_train, Y_train)

    # Making predictions
    Y_pred = svc_clf.predict(X_test)

    # Calculating metrics
    precision = metrics.precision_score(Y_test, Y_pred, average='weighted')
    recall = metrics.recall_score(Y_test, Y_pred, average='weighted')

    # Printing metrics for each iteration
    print(f"Precision: {precision}")
    print(f"Recall: {recall}")

    # Appending metrics
    precisions.append(precision)
    recalls.append(recall)

    # Calculating and displaying confusion matrix
    cm = confusion_matrix(Y_test, Y_pred)
    print("Confusion Matrix:\n", cm)

    # Visualizing confusion matrix
    disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=svc_clf.classes_)
    disp.plot(cmap='BuPu')
    plt.title(f"Confusion Matrix - Iteration {iteration}")
    plt.show()

# Calculating mean metrics
svc_mean_precision = np.mean(precisions)
svc_mean_recall = np.mean(recalls)

# Printing mean metrics
print("Mean cross-validation precision: ", svc_mean_precision)
print("Mean cross-validation recall: ", svc_mean_recall)
```

2. Implementation of Naive Bayes Classifier (NBC) Model

```python
from sklearn.metrics import precision_score, recall_score, confusion_matrix, ConfusionMatrixDisplay
# Initialize Naive Bayes Classifier
nbc_clf = MultinomialNB(alpha=0.3)  # Tuned alpha value after experimentation

# Running cross-validation
kf = StratifiedKFold(n_splits=2, shuffle=True, random_state=1)  # 2-fold cross-validation
precisions = []  # For precision
recalls = []  # For recall
iteration = 0
smote = SMOTE(random_state=101)  # SMOTE for oversampling minority class

for train_index, test_index in kf.split(data_tfidf, Y):
    iteration += 1
    print("Iteration ", iteration)

    # Splitting data into training and testing sets
    X_train, Y_train = data_tfidf[train_index], Y[train_index]
    X_test, Y_test = data_tfidf[test_index], Y[test_index]

    # Applying SMOTE on the training set
    X_train_resampled, Y_train_resampled = smote.fit_resample(X_train, Y_train)
    print(f"Training data size before SMOTE: {X_train.shape}, after SMOTE: {X_train_resampled.shape}")

    # Fitting the Naive Bayes Classifier on resampled data
    nbc_clf.fit(X_train_resampled, Y_train_resampled)

    # Making predictions on the test set
    Y_pred = nbc_clf.predict(X_test)

    # Calculating metrics
    precision = precision_score(Y_test, Y_pred, average='weighted')
    recall = recall_score(Y_test, Y_pred, average='weighted')

    # Printing metrics for each iteration
    print(f"Precision: {precision}")
    print(f"Recall: {recall}")

    # Appending metrics to calculate mean later
    precisions.append(precision)
    recalls.append(recall)

    # Calculating and displaying confusion matrix
    cm = confusion_matrix(Y_test, Y_pred)
    print("Confusion Matrix:\n", cm)

    # Visualizing the confusion matrix
    disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=nbc_clf.classes_)
    disp.plot(cmap='BuPu')  # Lighter purple palette
    plt.title(f"Confusion Matrix - Iteration {iteration}")
    plt.show()

# Calculating mean metrics across all folds
nbc_mean_precision = np.mean(precisions)
nbc_mean_recall = np.mean(recalls)

# Printing mean metrics
print("Mean cross-validation precision: ", nbc_mean_precision)
print("Mean cross-validation recall: ", nbc_mean_recall)
```

3. Implementation of Logistic Regression Model

```
[ ]  import numpy as np
     import pandas as pd
     from sklearn.feature_extraction.text import TfidfVectorizer
     from sklearn.metrics import precision_score, recall_score
     from sklearn.linear_model import LogisticRegression
     from imblearn.pipeline import Pipeline
     from imblearn.over_sampling import SMOTE
     from sklearn.linear_model import SGDClassifier
     from sklearn.model_selection import train_test_split
     import warnings
     warnings.filterwarnings("ignore", category=FutureWarning)

     # Assuming the dataset has 'cleaned_tweet' (features) and 'status' (labels) columns
     X = dataset['cleaned_tweet']  # Use cleaned tweets for features
     Y = dataset['status']  # Target sentiment column

     # Convert text data to numerical data using TF-IDF
     tfidf_vectorizer = TfidfVectorizer(min_df=0.001, max_features=4000, ngram_range=(1, 2))
     X_tfidf = tfidf_vectorizer.fit_transform(X)

     # Split data into train and test sets
     X_train, X_test, Y_train, Y_test = train_test_split(X_tfidf, Y, test_size=0.2, random_state=1)

     # Define the pipeline with SMOTE and Logistic Regression
     pipeline = Pipeline([
         ('balancing', SMOTE(random_state=101)),
         ('classification', LogisticRegression(max_iter=500, random_state=1))
     ])

     # Fit the model
     pipeline.fit(X_train, Y_train)

     # Evaluate the model on test data
     Y_pred = pipeline.predict(X_test)

     # Calculate Precision and Recall
     precision = precision_score(Y_test, Y_pred, average='weighted')
     recall = recall_score(Y_test, Y_pred, average='weighted')

     # Print Precision and Recall
     print(f"\nPrecision (Weighted): {precision:.4f}")
     print(f"Recall (Weighted): {recall:.4f}")
```

These are the code snippets of the three models employed for the sentiment analysis. Both the models showed the specific performance characteristics and helped to achieve the desired result. These models presented the desired results in terms of prediction and were particularly useful in measuring sentiments. The models selected for this study has been put to test with this successful implementation.

APPENDIX B: Data Sources and Ethical Considerations

This appendix discusses the datasets used in this research project in light of the ethical considerations as well as the privacy measures that have been taken to enhance responsible research.

Data Sources Used in the Project

- **Social Media Comment Dataset**: A set of comments which are originally in form of text obtained from several social media platforms. It is evident that the sentiment and mental health states that involve in this dataset are Normal, Depression, Suicidal, Anxiety, Stress, Bipolar and Personality Disorder; thus, it has adequate coverage with core issues.
- **Mental Health Status Annotations**: The dataset presents notes which have tags regarding particular mental health disorders as the classes. These annotations were developed with the help of domain knowledge, and then checked for their correctness and reliability.

Ethical Considerations and Data Privacy Measures

In conducting this research, we adhered to strict ethical guidelines to protect the privacy and confidentiality of individuals whose data were analyze

- **Data Anonymization**: All datasets used were deidentified and there were no patient identifiable information (PII) available in any of the datasets used. This approach shares information consistent with enactments of ethics in managing sensitive information.
- **Informed Consent:** We used publicly available datasets and worked under the stated term of service provided by the various sites. Since it was impossible to obtain direct informed consent from the users their data was gathered using methods that complied with ethical research standards.
- **Data Security**: Some steps were taken while analyzing and using quantitative data such as limiting access and encrypting data to avoid unauthorized access.
- **Ethical Approval**: That research approach was presented and approved for IRB by the relevant university to the standards of ethical treatment of human subjects in experimental studies.

Following these ethical concerns and data protection measures, we wanted to execute the present study in the least harm way to people's rights to privacy and self-protection.