The Potential of AI in Detecting Depression Using Social Media

Aafreen Singh, Manya
Department of Electronics and Communication Engineering,
Thapar Institute of Engineering & Technology, Patiala, Punjab, India.
asingh22_be20@thapar.edu , mmanya_be20@thapar.edu

***Abstract*— The use of AI techniques in predicting depression from social media is a promising area of study, as it has the potential to revolutionize mental health diagnosis and treatment. By leveraging the vast amount of data available on social media platforms, AI techniques can provide valuable insights into individuals' mental health status, helping mental health professionals identify those who may benefit from professional intervention. However, it is important to consider ethical considerations when implementing these techniques in healthcare and psychiatric diagnosis. The use of AI in psychiatric diagnosis should be carefully considered, and privacy and fairness concerns should be addressed to ensure that these technologies are used ethically and effectively.**

This research paper explores the use of artificial intelligence (AI) techniques in predicting depression from Twitter data through sentiment analysis. The paper utilizes the TextBlob library and multiple classification algorithms, such as Random Forest Classifier, KNN Classifier, Decision Tree, and Naïve Bayes, to analyze a labeled dataset of 20,000 English tweets collected from the Twitter API. The tweets are labeled as either depressed or non-depressed based on the user who posted them.

This study intends to assess the accuracy of using AI techniques to predict depression based on data from social media and to explore the ethical concerns related to using AI for psychiatric diagnosis. The study's results reveal that AI techniques, especially the Decision Tree followed closely by Random Forest Classifier, demonstrate high accuracy in predicting depression from Twitter data.
The use of AI techniques to identify depression from social media data is promising, potentially reducing barriers to accessing mental health services. While ethical considerations exist, further research is needed to explore their potential use in identifying and treating mental health disorders.

Keywords: AI, Sentiment Analysis, Natural language processing*, *TextBlob, Naïve Bayes*, *Random Forest Classifier, KNN Classifier, Decision Tree

I. INTRODUCTION

Mental health disorders are a major health concern, with millions of individuals suffering from conditions such as depression, anxiety, and bipolar disorder. These disorders significantly impact individuals' lives, affecting their daily activities, relationships, and overall well-being. Traditional diagnostic methods for mental health disorders, such as clinical interviews and self-report questionnaires, can be time-consuming, expensive, and may not always be accurate. The need for more efficient and reliable diagnostic methods has led to an interest in using technology and social media platforms like Twitter as potential sources of data for mental health diagnosis.

Social media is an integral part of people's lives. Keeping ethical considerations in mind, we can use the wealth of information that it provides to gain insights into their mental health. Twitter, in particular, has emerged as a popular platform for individuals to share their thoughts and feelings publicly. Sentiment analysis of Twitter data using artificial intelligence (AI) techniques has shown promise in predicting depression, as individuals may reveal their emotional state and personal experiences on social media. Sentiment analysis involves the use of NLP techniques to analyze text data and identify the sentiment or emotional tone of the text.

Several recent studies have explored using sentiment analysis of Social Media data for predicting depression. These studies have shown promising results, suggesting that social media data can be a valuable tool for identifying individuals who may be at risk of developing depression. However, ethical implications are also associated with the use of AI in psychiatric diagnosis. This research paper aims to investigate the potential of sentiment analysis of Twitter data for predicting depression and examine the ethical implications of using AI in psychiatric diagnosis.

In our research, we will describe the methodology employed for **sentiment analysis using TextBlob** and various classifying algorithms like **Decision Tree, Random Forest Classifier,** and **KNN Classifier**. We will then compare the various algorithms using various evaluation metrics.

Overall, the objective of this research paper is to contribute to the growing body of literature on the use of social media data for mental health diagnosis. Our findings will have important implications for mental health diagnosis and highlight the need for responsible use of AI in psychiatric diagnosis.

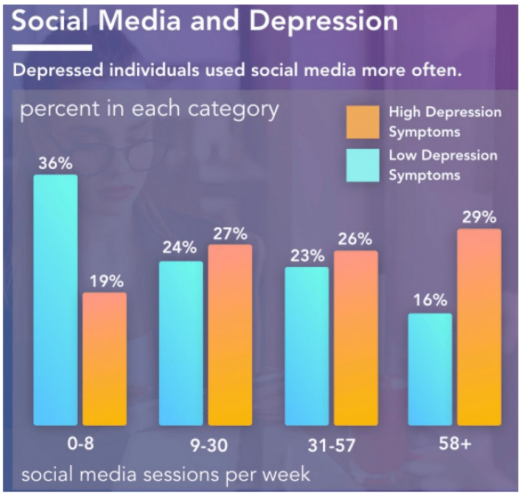

Figure 1. Social Media and Depression. Source: Clearvue Health

Figure 1 shows Social Media usage of individuals with different levels of symptoms of depression. This suggests that social media can be an effective tool for detecting depression symptoms in users.

I. Literature Study

In the past years, interest has grown in using computational linguistics techniques to analyze social media posts. These methods have been effectively applied to study various human behavior aspects, including mental health. Previous research has demonstrated the efficacy of using these techniques to identify and predict depression by analyzing language patterns in social media posts. Consequently, computational linguistics has emerged as a tool for research related to mental health and can potentially be used for transforming how mental health is studied and treated.

Our research aimed to build on the methods used in previous research by studying various papers related to mental health, sentiment analysis, and other techniques employed here. By reviewing the work of other researchers, we aimed to identify methods used to collect data from social media platforms as well as various computational linguistics techniques employed in sentiment analysis, machine learning algorithms, and natural language processing.

By incorporating the best practices and ideas from these studies, we sought to compare the existing methods for detecting depression in social media posts.

1. Their post tell the truth: Detecting social media users mental health issues with sentiment analysis

This study [1] addresses the cultural stigma surrounding mental health disorders and the need for a space for individuals to express their thoughts. Social media is identified as a potential medium for individuals with mental health disorders. The study aims to identify mental health disorders via words or tweet narration with specific keywords. **5537 clean tweet data** were collected from the Indonesian population using **rapidminer sentiment analysis** and categorized into categories (Positive, Negative, and Neutral). The results indicate that Twitter is an effective tool used to identify symptoms of mental health disorders and is considered a platform for individuals with mental health issues to express themselves.

2. I feel you: Mixed-methods study of social support of loneliness on twitter

This research [2] examines features associated with the social support provided for expressions of loneliness on Twitter. Analyzing 4 million tweets, it finds that users with more extensive networks and positive language receive more feedback, while swearing leads to fewer responses. Emotional support is the most common, often including elements of invisible support. However, some replies are considered online bullying, highlighting the need for intervention. Results suggest Twitter can be a space for loneliness expression and social support.

3. Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations

[3] uses deep learning for detection of optimism and pessimism in tweets about COVID-19. In this research, several network architectures are compared. A **transformer embedding** is used for the purpose of extracting semantic features. The models based on **bidirectional long- and short-term memory networks** perform the best**.** The approach evaluated experimental results on four periods of the COVID-19 pandemic.

4. The great methods bake-off: Comparing performance of machine learning algorithms

The study [4] compares the performance of various algorithms, including newer approaches, in developing risk assessments. A sample of **over 250,000 youth assessments** was used to observe predictions by varying sample size and base rate. The study found that while newer machine learning approaches showed promise, sample size was the most critical factor in determining algorithm performance. The study suggests that agencies and providers should prioritize transparency and consider sample size when adopting or developing risk assessment tools.

5. Comparing the performance of machine learning algorithms using estimated accuracy,

Different data mining algorithms are compared to find the one with the most accuracy. The approach involves trial and error to choose the best algorithm and using test metrics for comparison. Results show that **Decision tree** is the best model for prediction.[5]

6. Live Sentiment Analysis Using Multiple Machine Learning and Text Processing Algorithms

This study [6] focuses on analyzing the sentiment of trending tweets on Twitter using multiple algorithms. The aim is to improve accuracy of sentiment analysis, which can be challenging due to the huge amount of unstructured data on the platform. **Naive Bayes, Support-Vector Machine, Lexicon Approach, and Textblob** are used to achieve this goal. The study hypothesizes that combining these methods would result in more accurate sentiment analysis.

7. ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification

The study [7] uses an ensemble model which combines deep learning and machine learning models for improvement of the classification accuracy of unstructured data collected from social media platforms. Data preprocessing is done using **TextBlob** and **VADER**, and the study compares the accuracy of **Word2Vec, TF, and TF-IDF** features. The **tree classifier** trained with TF-IDF features data shows the best performance. The proposed model achieves high accuracy scores, with TextBlob scoring 0.97 and VADER scoring 0.95 in accuracy using Word2Vec features. The results suggest that the ensemble model with a voting criterion outperforms other models. Sentiment analysis tweets related to COVID-19 reveals predominantly negative sentiments expressed by people.

8. The climate change Twitter dataset, Expert Systems with Applications

This study [8] presents a comprehensive dataset of over 15 million tweets spanning 13 years about climate change and human opinions. The dataset includes seven dimensions of information, such as geolocation, user gender, climate change stance and sentiment, aggressiveness, and deviations from historic temperature, among others. The dimensions were generated using various machine learning algorithms such as SVM, LSTM, CNN, BERT, RNN, Naive Bayes.

9. DEPTWEET: A typology for social media texts to detect depression severities

There is a lack of standard typology and inadequate data for data-driven mental health research to build a typology for social media texts. The typology emulates the DSM-5 and PHQ-9 procedures to detect the severity of depression from tweets, with a new dataset of 40191 tweets labeled by expert annotators. The tweets are labeled categorically as ‘depressed’ or ‘non-depressed’ with varying severity. The dataset quality is validated with associated confidence scores. The best results are achieved using attention-based models such as **DistilBERT** and **BERT**. [9]

10. Sentiments comparison on Twitter about LGBT

This paper [10] discusses sentiment analysis of tweets related to the LGBT topic, which has gained popularity and controversy in society. The study collects tweets from fifty US states and applies pre-processing techniques used for classification of sentiments into categories (positive, negative, and neutral). Five algorithms: **Naive Bayes, TextBlob PatternAnalyzer, Linear Support Vector Machine, XGBoost,** and **Logistic Regression**, are used with and without pre-processing. The study finds that Logistic Regression without pre-processing yields the highest F1-score. The sentiment analysis of US tweets on the LGBT topic shows that most tweets have a neutral sentiment.

11. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry.

The study [11] uses a hybrid approach for sentiment analysis in the airline industry using **lexicon-based methods**. Deep learning models are used to improve accuracy. The study evaluates the usage of **TextBlob** for classification in comparison with other sentiment analysis methods. **CNN, LSTM, CNN-LSTM, and GRU** are compared with other machine learning algorithms. The efficiency of BoW and TF-IDF are also investigated. Results indicate that models show better performance when trained using sentiments assigned using TextBlob compared to the original ones in the dataset. **LSTM-GRU** performs the best out of all models and it achieves the highest accuracy. The **extra tree classifier** with TF-IDF and **support vector classifier** with BoW show the best accuracy. The study concludes that TextBlob-based annotation can be useful for humans to improve accuracy, but cannot replace humans as error-proneness and subjectivity are present.

12. Depression Detection by Analyzing Social Media Post of User

This research [12] focuses on employing machine learning and NLP to detect depression through social media posts. Depressed users show different posting patterns, and the proposed system can classify users as depressed or non-depressed. The system uses **NLP** and the **BERT** algorithm for efficient detection. The approach can aid in early detection and treatment of depression and other mental illnesses by analyzing social media data.

13. *Real-time Acoustic based Depression Detection using Machine Learning Techniques*

This study [13] provides an overview of three primary models that have been developed to predict depression in individuals: a) utilizing machine learning classifiers and WEKA, b) utilizing imaging and machine learning methods, and c) utilizing risk factors.

A wide variety of methodologies and techniques are available for detecting the level of Depression in social media posts, which are continuing to grow.

In our study, we use TextBlob for conducting sentiment analysis and various classification algorithms for predicting Depression. The framework consists of various steps, including data cleaning and pre-processing, and comparing various machine learning classifiers, followed by experimental results.

III. Proposed Model

We intend to utilize social media as a tool to gain insights into individuals' behavior attributes, thinking style, mood, opinions, and socialization patterns, in order to screen for and predict depression levels. We will work on extracting information from social media posts and utilizing machine learning classifiers to predict the depression levels of users.

We use a dataset of 20,000 labeled English tweets collected via the Twitter API, with labels indicating whether the users who posted them are depressed or non-depressed. The tweets are preprocessed using NLP techniques such as stop words removal, stemming and lemmatization. Data is cleaned by removing punctuation and numbers, and converting the text to lowercase (case normalization)

Further, we employ the TextBlob library and various classifying algorithms, including Naïve Bayes, Decision Tree, KNN Classifier, and Random Forest Classifier to analyze the preprocessed data and classify the tweets into depressed and non-depressed categories.

We will utilize multiple evaluation metrics such as Accuracy, Precision, Recall, F1 score and Confusion Matrix to compare the efficacy of different classifying algorithms.

In our research, we will follow the approach illustrated in Figure 2.

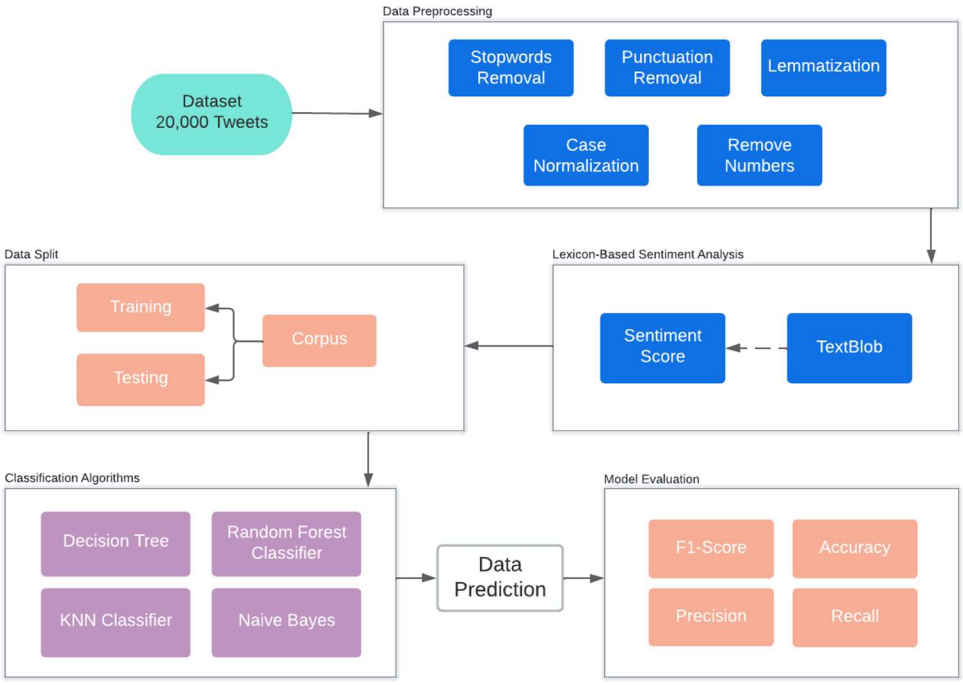

Figure 2. Architecture of the proposed approach

Data Preprocessing:

To prepare our data for efficient sentiment analysis, we employed various Python libraries, including **Pandas** and **Numpy** for preprocessing, and **Matplotlib** and **Seaborn** for data visualization. Our data was pruned down only to include the necessary columns.

We also applied text-cleaning techniques, including the removal of punctuation, numbers, stop words, and normalization of the case. To accomplish this, we utilized the **TextBlob** and **NLTK** (Natural Language Toolkit) libraries, which are widely used for natural language processing (NLP) tasks.

TextBlob is particularly useful as it offers a more user-friendly API for common NLP tasks like sentiment analysis, part-of-speech tagging, and text classification. **NLTK**, on the other hand, provides an extensive range of functionalities for tasks such as tokenization, stemming, lemmatization, parsing, and more. It also includes numerous corpora and datasets for language modeling and machine learning.

Additionally, we employed Lemmatization to group similar words and Word Tokenization to tokenize each word.

Sentiment Analysis:

Sentiment analysis is a NLP technique used to determine the sentiment (emotional tone) of a given piece of text. The goal is to classify the text into categories: positive, negative, or neutral, based on the words and phrases used.

TextBlob for Assigning Sentiments:

Using TextBlob, a Python library, we can conduct sentiment analysis on preprocessed text data. TextBlob utilizes a machine learning algorithm to analyze text and assign a sentiment.

Sentiment is a score ranging from -1 to +1. A score of -1 indicates a very negative sentiment, +1 indicates a very positive sentiment, and 0 indicates a neutral sentiment.

By analyzing tweets using TextBlob's sentiment analysis, we can categorize them as depressed or non-depressed based on their sentiment scores. For example, a tweet with a sentiment score of -0.8 might be classified as depressed, while a tweet with a sentiment score of +0.5 may be classified as non-depressed. The objective of sentiment analysis using TextBlob in this context is to classify each tweet as exhibiting signs of depression or not, based on its sentiment score.

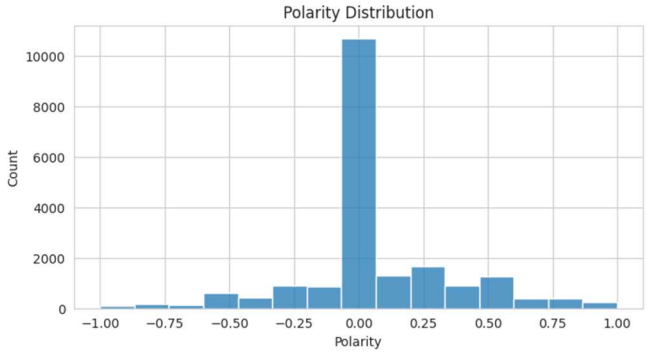

Figure 3. Count vs Polarity

Upon applying sentiment analysis and assigning polarity to each tweet, we can see in Figure 3 that approximately over 50% of the tweets are classified as neutral.

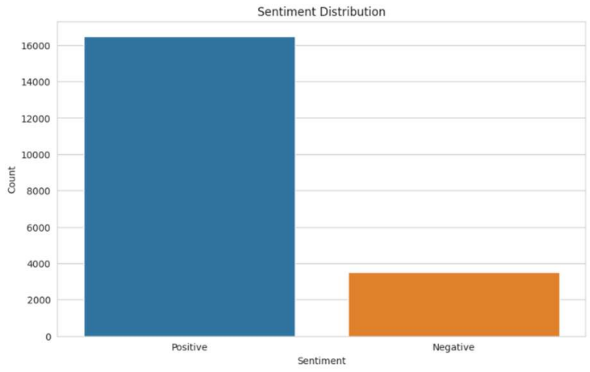

Figure 4. Number of tweets classified as positive and negative

Figure 4 shows that around 16,000 tweets are classified as positive. However, to improve the accuracy of the model, it should

be tested on a dataset containing a higher proportion of negative tweets.

Classification Algorithms:

Classification is a type of predictive modeling technique used in supervised learning. It involves predicting the classes of new observations according to a set of labelled data learned by the model. This algorithm learns from a given dataset and identifies the category or group that new observations belong to. Examples of classes or groups can be binary, such as Yes/No or 0/1, or categorical, such as Spam/Not Spam, or, in our case, depressed/not depressed. Essentially, the classification algorithm maps input data to output labels by learning from a training dataset.

Popular classification algorithms include **Logistic Regression, SVM, Naïve Bayes, Random Forests, and Decision Trees.**

In our research, we use various classification algorithms to train models to predict whether a given social media post is indicative of depression or not. By comparing their performance on our dataset, we can determine the most effective approach to detecting depression in social media posts.

Decision Tree:

It is a popular classification algorithm, represented by a tree-like graph. It is easy to understand and interpret, making it a valuable tool for sentiment analysis.

Decision Tree for Sentiment Analysis:

Decision tree algorithms are used for sentiment analysis by building a model that predicts the sentiment of a given text. In this approach, the text is first preprocessed to extract relevant features such as the presence of specific words or phrases, part-of-speech tags, or other linguistic features that can be indicative of sentiment. These features are used for training the model, which can predict the sentiment of new, unseen text based on the learned patterns.

The decision tree algorithm splits the data based on the features that provide the most information gain or the best separation between classes. For each node, a decision is made based on a specific feature or set of features, and the data is partitioned into subsets that are more homogenous compared to the target variable (sentiment, in this case). This process continues recursively until stopping criterion is met.

Further, when the model is trained, it is used to predict the sentiment of new text based on certain features of the input text. The prediction is made by following the path down the tree till leaf node is reached, which corresponds to a predicted sentiment label (e.g., positive or negative). The decision tree approach can be effective for sentiment analysis, especially when the text contains a small number of features and the relationships between features and sentiment are relatively simple.

Results of Decision Tree algorithm:

The various evaluation metrics of the decision tree on our dataset are shown in the table below.

	Precision	**Recall**	**F1 Score**
Negative	0.80	0.90	0.85
Positive	0.98	0.95	0.96
Accuracy			0.96

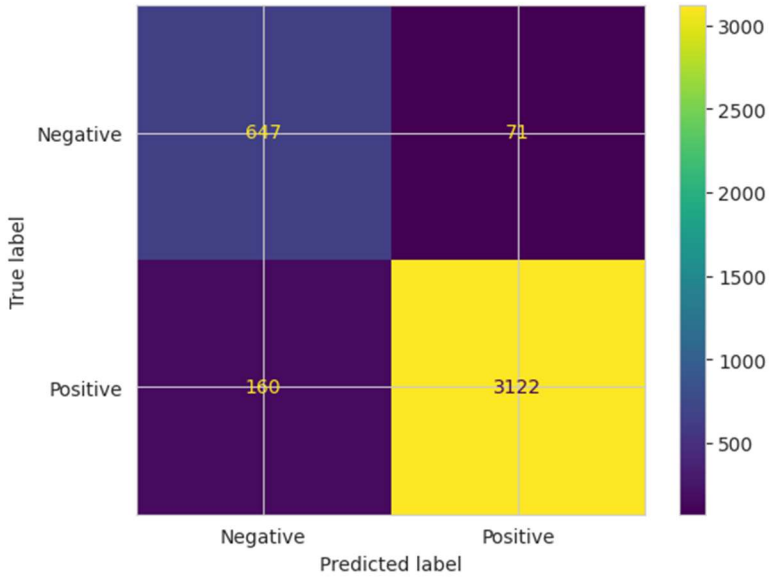

Figure 5 Confusion Matrix of a Decision Tree

Based on our results, the Decision Tree algorithm proves to be an effective method for detecting depression symptoms from Twitter data, achieving an accuracy of 96%. As illustrated in Figure 5, the Confusion matrix indicates that out of a testing dataset of 4,000 tweets, 3,769 tweets (94.225% of the testing set) were correctly classified, while 231 tweets (5.775% of the testing set) were classified incorrectly.

Random Forest Classifier:

It is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes or the mean prediction of each tree in the forest. It provides high accuracy and robustness to noisy data.

Random Forest Classifier for Sentiment Analysis:

To use a Random Forest Classifier for sentiment analysis, we would typically start by preprocessing the text data by stop words removal, stemming or lemmatizing words, and converting the text into numerical feature vectors using methods like bag-of-words or TF-IDF.

Next, we would use the pre-processed text data as input to train a Random Forest Classifier. During training, the algorithm learns to identify patterns in the input data associated with particular sentiment labels, allowing it to classify new, unseen text data according to their predicted sentiment.

Finally, we use the trained Random Forest Classifier to classify the sentiment of unseen data. For instance, given a tweet, the classifier would process the text, convert it into numerical feature vectors, and use the learned patterns to predict the sentiment (positive, negative, or neutral).

It works by constructing a large number of decision trees and combining their predictions to determine the overall sentiment of a tweet.

Results of Random Forest Classifier algorithm:

The various evaluation metrics of the Random Forest Classifier on our dataset are shown in the table below.

	Precision	**Recall**	**F1 Score**
Negative	0.79	0.82	0.81
Positive	0.96	0.95	0.96
Accuracy			0.93

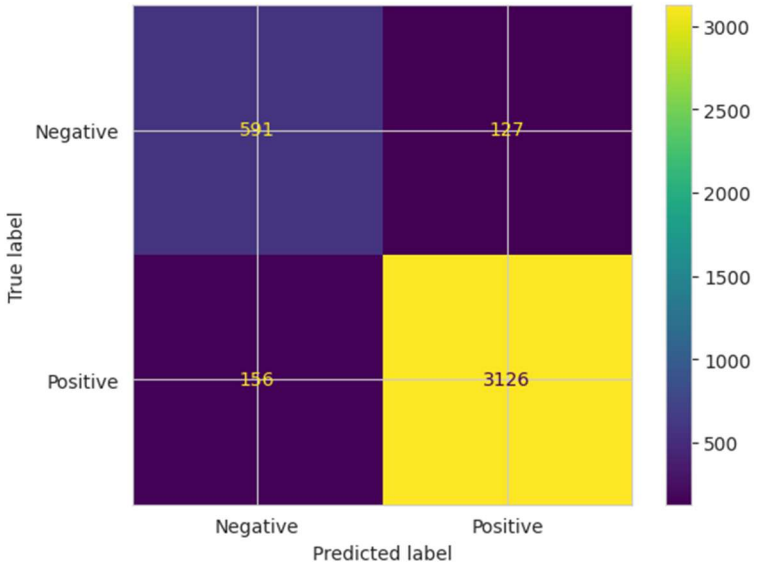

Figure 6 Confusion Matrix of a Random Forest Classifier

Figure 6 displays the Confusion matrix of the Random Forest Classifier model, which achieved an accuracy of 93% in detecting individuals with depression from Twitter data. Out of the testing dataset of 4000 tweets, 3717 tweets (92.925% of the testing set) were accurately classified, while 283 tweets (7.075% of the testing set) were classified incorrectly.

KNN Classifier:

K-nearest neighbor is a non-parametric algorithm that classifies new instances based on the class of their k-nearest neighbors in the training set. It is a machine learning algorithm which can be used for both classification and regression.

KNN Classifier for Sentiment Analysis:

It can be used for sentiment analysis by first training the algorithm on a labeled dataset, where the sentiment of each text sample (e.g., a tweet) is known. The KNN classifier works by identifying the k-nearest neighbors to a given test sample in the feature space (i.e., a vector representation of the text sample) and classifying the sample based on the majority class of its neighbors.

To use KNN for sentiment analysis, the first step is to preprocess the text data, which typically involves removing stop words, stemming or lemmatization, tokenization and converting text to numbers (using BoW or TF-IDF).

Once the data is preprocessed, the KNN classifier can be trained using the labeled training data. In case of sentiment analysis, the labels are usually binary (e.g., positive or negative sentiment). When a new text sample is given to the classifier, it is transformed into the numerical feature representation and the k-nearest neighbors to the sample in the feature space are identified. The majority class of the neighbors is then used to classify the sample as positive or negative.

However, KNN classifier has some limitations in sentiment analysis. For instance, it tends to suffer from the "curse of dimensionality" problem, where the performance deteriorates as the number of features (i.e., the dimensionality of the feature space) increases. Additionally, it may not perform well with imbalanced datasets, where one class has significantly fewer samples than the other.

Results of KNN Classifier algorithm:

The various evaluation metrics of the KNN Classifier on our dataset are shown in the table below.

	Precision	**Recall**	**F1 Score**
Negative	0.82	0.18	0.29
Positive	0.85	0.99	0.91
Accuracy			0.85

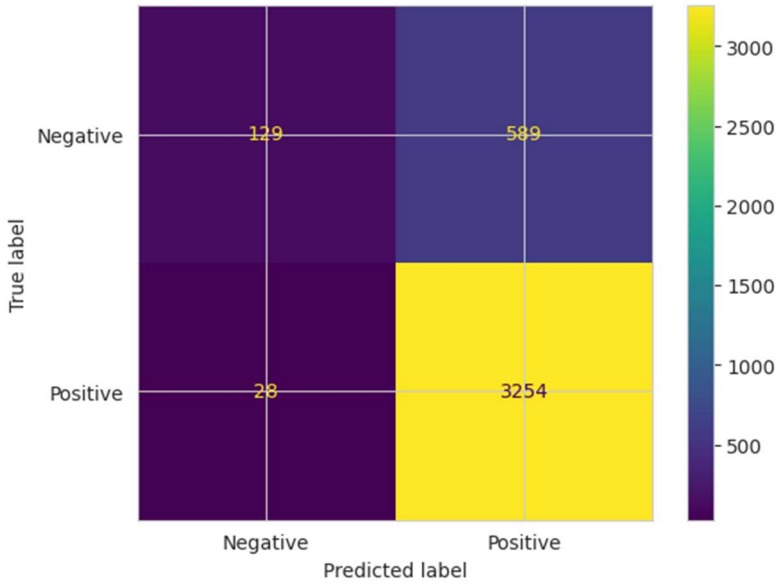

Figure 7 Confusion Matrix of a KNN Classifier

The KNN classifier demonstrated an accuracy of 85% in detecting individuals with depression from Twitter data. The Confusion matrix in Figure 7 depicts that out of the 4000 tweets in our testing data, 3383 tweets (84.575% of the testing set) were correctly classified, while 617 tweets (15.425% of the testing set) were classified incorrectly.

Naïve Bayes:

It is a probabilistic technique commonly used in sentiment analysis. The algorithm uses Bayes' theorem for calculation of probability of a given text belonging to a sentiment category, such as positive or negative.

Naïve Bayes for Sentiment Analysis:

The algorithm learns from a labeled dataset of texts that are classified as either positive, negative, or neutral. It uses the frequency of words in the text to calculate the probability of the text belonging to each category.

The Naive Bayes algorithm assumes that all words in the text are independent of each other, which is the reason why it is called "naive". This allows the algorithm to calculate the probability by multiplying the probabilities of each word in the text belonging to that category.

For example, if we want to classify the sentiment of the following text: "The movie was great, I really enjoyed it". The algorithm would calculate the probability of the text belonging to each category (positive, negative, or neutral) based on the frequency of words in the text. The algorithm multiplies the probabilities of each word to get the probability of the entire text belonging to a category.

If the probability of the text belonging to the positive category is higher than the negative category, the algorithm classifies the text as positive. If the probability of the text belonging to the negative category is higher than the positive category, the algorithm classifies the text as negative. If the probabilities of both categories are equal, the text is classified as neutral.

Overall, this is a simple and effective algorithm for sentiment analysis which can be used on large data with high accuracy. However, its performance may suffer when dealing with sarcasm or figurative language.

Results of Naïve Bayes algorithm:

The various evaluation metrics of Naive Bayes on our dataset are shown in the table below.

	Precision	**Recall**	**F1 Score**
Negative	0.87	0.40	0.55
Positive	0.88	0.99	0.93
Accuracy			0.88

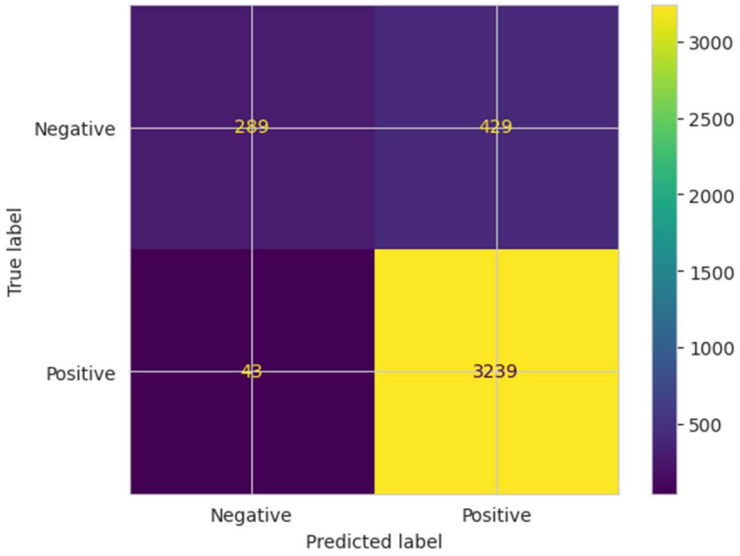

Figure 8 Confusion Matrix of Naive Bayes

The Naïve Bayes algorithm shows an accuracy of 88% in detecting depression from Twitter data. As shown in Figure 8, out of our testing data of 4000 tweets, 3528 tweets (88.2% of the testing set) were classified correctly, while 472 tweets (11.8% of the testing set) were incorrectly classified.

A. Abbreviations and Acronyms

NLTK – Natural Language Toolkit
NLP – Natural Language Processing
ML – Machine Learning
API – Application Programming Interface
BERT – Bidirectional Encoder Representations from Transformers
WEKA – Waikato Environment for Knowledge Analysis
KNN – K-Nearest Neighbors
SVM – Support Vector Machine
SVR – Support Vector Regression
TF-IDF – Term Frequency - Inverse Document Frequency
BoW – Bag of Words
TP – True Positive
TN – True Negative
FP – False Positive
FN – False Negative

IV. Evaluation of the Proposed System

The proposed system for sentiment analysis using machine learning algorithms appears to be a promising approach for identifying signs of depression in social media data. The study utilized four different algorithms, namely Random Forest Classifier, Decision Tree, KNN Classifier, and Naïve Bayes, to evaluate their performance in accurately classifying tweets as depressed or non-depressed based on their sentiment scores.

The results indicated that the Decision Tree algorithm achieved the best performance with an accuracy rate of 93.97%, followed closely by the Random Forest Classifier with an accuracy rate of 92.95%. The Naïve Bayes algorithm also performed reasonably well, achieving an accuracy rate of 88.2%. However, the KNN Classifier exhibited the lowest accuracy rate of 84.58%.

The high accuracy rates achieved by the Decision Tree and Random Forest Classifier algorithms suggest that they may be the most suitable for this task, and may be considered the preferred algorithms for future sentiment analysis tasks. However, further research may be necessary to investigate the performance of these algorithms on larger and more diverse datasets, as well as assess their generalizability to different contexts and domains.

While accuracy is a crucial measure for evaluating the performance of machine learning models, other metrics, such as precision, recall, and F1 score, are also considered which also provide a more comprehensive evaluation.

Moreover, the study could benefit from a more in-depth analysis of the limitations and potential biases of the algorithms used. For example, Decision Trees are known to be susceptible to overfitting, which may limit their generalizability to new data. Similarly, the Naïve Bayes algorithm assumes independence between features, which may not always hold in real-world scenarios.

Overall, the proposed system for sentiment analysis using machine learning algorithms appears to be a promising approach for identifying signs of depression in social media data. However, further research is necessary to investigate its generalizability, potential limitations, and biases.

Evaluation Metrics:

Evaluation metrics are an essential aspect of any machine learning model, as they enable us to assess its performance and effectiveness. The following are some of the evaluation metrics used for assessing and comparing the performance of the classification models used in our research.

1. **Precision**: The ratio of true positives to the total number of predicted positives. It measures how well a model identifies relevant instances and eliminates irrelevant ones.
   Precision = TP / (TP + FP)
2. **Recall**: The ratio of true positives to the total number of actual positives. It measures how well a model can detect all the relevant instances in the data.
   Recall = TP / (TP + FN)
3. **F1 score:** The harmonic mean of precision and recall. It is a measure of a model's accuracy in identifying positive instances while avoiding false positives.
   F1 Score = 2 * (precision * recall) / (precision + recall)
4. **Accuracy**: The ratio of the correctly classified instances to the total number of instances. It is a measure of how well a model can classify instances into their correct classes.
   Accuracy = (TP + TN) / (TP + TN + FP + FN)
5. **Confusion matrix**: A table used to evaluate the performance of a classification model. It contains information about the true positive, true negative, false positive, and false negative instances of a technique.

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

where:
TP: True Positive - correctly classified positive instances
TN: True Negative - correctly classified negative instances
FP: False Positive - incorrectly classified positive instances
FN: False Negative - incorrectly classified negative instances

Algorithm used	**Accuracy** (for negative tweets)	**Precision** (for negative tweets)	**Recall** (for negative tweets)	**F1 score** (for negative tweets)
Decision Tree	93.97%	0.98	0.95	0.96
Random Forest Classifier	92.95%	0.96	0.95	0.96
KNN Classifier	84.58%	0.85	0.99	0.91
Naïve Bayes	88.2%	0.88	0.99	0.93

OBSERVATION

The study findings indicate that AI techniques, specifically Decision Tree and Random Forest Classifier, exhibit high precision in forecasting depression from Twitter data. These techniques have an accuracy rate of 93.97% and 92.95%, respectively, making them highly effective in detecting depression. Additionally, KNN Classifier and Naïve Bayes are also highly accurate, with success rates of 84.58% and 88.2%, respectively.

The usage of sentiment analysis in social media data is a relatively new area of research. However, results show its potential to be a highly efficient method for detecting depression. As social media platforms have become increasingly popular, people have started to share more personal information, including their thoughts and feelings. This makes social media data an invaluable source of information for detecting depression.

The results of our study suggest that the accuracy of the AI techniques employed in the research could be attributed to the fact that depression is usually associated with negative sentiments, which can be quickly identified through sentiment analysis. Therefore, it is highly probable that AI techniques would have a high accuracy rate in detecting depression in social media data.

In conclusion, the study findings reveal that AI techniques, particularly Decision Tree and Random Forest Classifier, are highly effective in predicting Depression from Twitter data, with a success rate of 93.97% and 92.95%, respectively. KNN Classifier and Naïve Bayes also show a high accuracy rate of 84.58% and 88.2%, respectively. These findings demonstrate the potential of using sentiment analysis of social media data for predicting Depression, which could have significant implications for the early detection and treatment of Depression.

Future research can explore how these techniques can be applied to larger and more diverse datasets. Additionally, it would be beneficial to investigate any limitations of the AI techniques used in the study.

Further studies could also examine the applicability of the findings to different contexts and domains, and assess the impact of various parameters and hyperparameters on the performance of the AI models. It may be useful to compare different AI techniques or combinations thereof, and explore any ethical implications and potential societal impacts of their use in various applications.

Moreover, developing interpretability and explainability methods for these AI models could enhance their transparency and trustworthiness.

CONCLUSION

In our research paper, we explored the potential of AI techniques to predict depression using Twitter data. Our findings demonstrate that sentiment analysis of social media data can be an effective tool for identifying individuals who may be at risk for depression.

The Decision tree and Random Forest Classifier proved to be highly accurate in predicting depression using Twitter data. This suggests that AI could be a useful tool in mental health diagnosis.

However, it is important to note that ethical concerns should be addressed when using AI in mental health diagnosis. One of the biggest concerns is the potential for bias present in the data used to train the AI models. For example, if the training data is only taken from a certain population, the AI models may not accurately predict depression in certain groups of individuals. Additionally, there are concerns about privacy and the potential for misuse of personal information.

Future research should focus on developing more accurate and reliable AI models for mental health diagnosis while addressing these ethical concerns. One possible avenue of research is to incorporate additional sources of data beyond social media, such as electronic health records or wearable technology. By combining multiple sources of data, AI models could potentially provide more accurate and comprehensive assessments of mental health.

Furthermore, it is vital to involve mental health professionals in the development and implementation of AI tools for mental health diagnosis. This will help ensure that the tools are used responsibly and ethically, and that they are integrated into existing mental health care systems in a way that benefits patients.

Overall, our research highlights the potential of AI in mental health diagnosis and the need for continued research in this area. While there are ethical concerns that need to be addressed, the benefits of accurate and efficient mental health diagnosis using AI could be significant.

REFERENCES

[1] Haris Herdiansyah, Rusdianto Roestam, Richard Kuhon, Adhi Setyo Santoso, Their post tell the truth: Detecting social media users mental health issues with sentiment analysis, Procedia Computer Science, Volume 216, 2023, Pages 691-697, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.185.

[2] Yelena Mejova, Anya Hommadova Lu, I feel you: Mixed-methods study of social support of loneliness on twitter, Computers in Human Behavior, Volume 136, 2022, 107389, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2022.107389.

[3] Guillermo Blanco, Anália Lourenço, Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations, Information Processing & Management, Volume 59, Issue 3, 2022, 102918, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2022.102918.

[4] Alex Kigerl, Zachary Hamilton, Melissa Kowalski, Xiaohan Mei, The great methods bake-off: Comparing performance of machine learning algorithms, Journal of Criminal Justice, Volume 82, 2022, 101946, ISSN 0047-2352, https://doi.org/10.1016/j.jcrimjus.2022.101946.

[5] Sunil Gupta, Kamal Saluja, Ankur Goyal, Amit Vajpayee, Vipin Tiwari, Comparing the performance of machine learning algorithms using estimated accuracy, Measurement: Sensors, Volume 24, 2022, 100432, ISSN 2665-9174, https://doi.org/10.1016/j.measen.2022.100432.

[6] Andrew Motz, Elizabeth Ranta, Adan Sierra Calderon, Quin Adam, Fadi Alzhouri, Dariush Ebrahimi, Live Sentiment Analysis Using Multiple Machine Learning and Text Processing Algorithms, Procedia Computer Science, Volume 203, 2022, Pages 165-172, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.07.023.

[7] Muhammad Umer, Saima Sadiq, Hanen karamti, Ala' Abdulmajid Eshmawi, Michele Nappi, Muhammad Usman Sana, Imran Ashraf, ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification, Pattern Recognition Letters, Volume 164, 2022, Pages 224-231, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2022.11.012.

[8] Dimitrios Effrosynidis, Alexandros I. Karasakalidis, Georgios Sylaios, Avi Arampatzis, The climate change Twitter dataset, Expert Systems with Applications, Volume 204, 2022, 117541, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.117541.

[9] Mohsinul Kabir, Tasnim Ahmed, Md. Bakhtiar Hasan, Md Tahmid Rahman Laskar, Tarun Kumar Joarder, Hasan Mahmud, Kamrul Hasan, DEPTWEET: A typology for social media texts to detect depression severities, Computers in Human Behavior, Volume 139, 2023, 107503, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2022.107503

[10]Aldinata, Axell Mondrian Soesanto, Vincent Christian Chandra, Derwin Suhartono, Sentiments comparison on Twitter about LGBT, Procedia Computer Science, Volume 216, 2023, Pages 765-773, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.194.

[11]Wajdi Aljedaani, Furqan Rustam, Mohamed Wiem Mkaouer, Abdullatif Ghallab, Vaibhav Rupapara, Patrick Bernard Washington, Ernesto Lee, Imran Ashraf, Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry, Knowledge-Based Systems, Volume 255, 2022, 109780, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2022.109780.

[12] Rutuja K Bhoge, Snehal A Nagare, Swapanali P Mahajan, Prajakta S Kor "Depression Detection by Analyzing Social Media Post of User" ISSN: 2321-9653 https://doi.org/10.22214/ijraset.2022.41874

[13]B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella and S. V. Alluri, "Real-time Acoustic based Depression Detection using Machine Learning Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-6, doi: 10.1109/icETITE47903.2020.394.

[14]https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media