Analyzing Twitter Data for Depression Signs Based on Machine Learning Techniques

Eman Khattak*, Haleem Ullah†, Irfanullah Khan‡, Malik Taimoor Ali§, Latif Jan**

Abstract

This research examines tweets with Natural Language Processing (NLP) and Machine Learning to detect trends associated with depression. The aim of this research is to create a dependable system for the early detection of depression. For this purpose, several classification algorithms are studied such as Decision Tree, K-Nearest Neighbors (KNN), Random Forest, Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Convolutional Neural Networks (CNN). All these models are tested to find out which is most dependable for depression detection. The effectiveness of these models is analyzed to determine which one is most precise in depression detection. Ongoing monitoring of users' text data allows participants to visualize the progression of depression over time, as well as recognize shifts in emotional states through graphs. The approach is made up of three phases: text cleaning, model building, and intensive testing of a given dataset. The data analysis shows that the best performing algorithms in accuracy were Logistic Regression and SVM with 91.8% and 91.9% respectively. It was noted that Logistic Regression performed better in precision and recall metrics which highlights its effectiveness in symptom depression detection.

***Keywords*: Natural Language processing, Machine Learning, K-Nearest Neighbors, Naïve Bayes, Decision Tree.**

Introduction

Depression is a complicated, multifaceted illness that impacts both the mind and the body. Symptoms include chronic discontent, lack of interest in hobbies one used to enjoy, and in hardcore cases, thoughts of self-harm. Depression is identified as an issue in public health worldwide, and is more prevalent in women as compared to men. India is one of the countries in the world that suffers from depression the most, with an estimated percentage of around 6.5% of the population in 2019, according to the WHO (Ahmed et al., 2021). One of the most problematic issues with the management of depression is the inability to identify depression in a

* School of Computer Science, Iqra National University Peshawar, Peshawar 25124, Pakistan, Khattake47@gmail.com

† School of Computer Science, Iqra National University Peshawar, Peshawar 25124, Pakistan, haleemkhankhalil@gmail.com

‡ School of Computer Science, Iqra National University Peshawar, Peshawar 25124, Pakistan, irfi0092@gmail.com

§ School of Computer Science, Iqra National University Peshawar, Peshawar 25124, Pakistan, maliktaimur@hotmail.com

**Corresponding Author: School of Computer Science, Iqra National University Peshawar, Peshawar 25124, Pakistan, latifjan@inu.edu.pk

timely manner and with precision. Conventional methods of diagnosis, such as PHQ-6 are based on self-reported responses, which are riddled with inaccuracies in cases where patients respond with fabrications. This drawback calls for identifying depressive tendencies (Anavi et al., 2015).

Babenko et al. (2015) developed an impressive depression detecting model using the classification algorithms, including K-Nearest Neighbors (KNN), Naive Bayes, Decision Tree, and Random Forest. In addition, the system can track past interactions to track the moods of the users and give graphical displays that will allow the users to view the change in moods over time (Babenko et al., 2014). According to Chicco et al. (2021), early intervention notifications are transmitted via Machine Learning into mental health assessment, thereby lessening significantly the working load of psychologists. The push to take care of the mentally challenged who suffered as a consequence of COVID pandemic that has caused unparalleled depression and anxiety levels across the world has never been as acute as it is now.

Further on, in the study where it examines how the methods of Machine Learning and Deep Learning can be applied to diagnose depression using social media information, as opposed to using the current methods of diagnosis based on key words, as in Dubey (2021). In this study, the accuracy of methods employed in the detection of depression is to be improved by looking at the labeled datasets to train and test on different datasets. The research proves the efficiency of utilizing the Machine Learning classifiers in the identification of depression as a radical change of the digital era in the mental health diagnostics, (Everingham et al., 2015).

The proposed research presents a Machine learning model based on the Natural Language Processing (NLP) to predict depression at its early stages using social media tweets. Multiplicity of classification algorithms, i.e. KNN, Naive Bayes, Decision tree, random forest, Support Vector Machine (SVM) and Logistic regression are all relatively tested to ascertain the most reliable model to identify depression. The proposed solution consists of the series of operations that embrace a pretext of texts, model development, and lastly testing of performance. The results of the experiment demonstrate that SVM and Logistic Regression prove to be the most precise with Logistic Regression being more precise and recall meaning better that it can be used in the detection of depressive symptoms. The structure also enables individuals to track and visualize the trends of emotions between sessions and this is why the system can be applicable to real-life mental health monitoring systems.

Literature Review

There has been a considerable focus on the use of Machine Learning algorithms for the detection of depression among users of social media and social platforms such as Twitter, Facebook, and YouTube (Hameed et al., 2021). The research showed that adding more features increased accuracy and F-measure scores which in turn improved depression detection. Likewise, Vasha et al. (2023) used approximately 10,000 data points obtained from comments on Facebook and YouTube and classified them into "sad" and "not sad" using six distinct Machine Learning classifiers. The study by Lacoste et al. (2007) indicated that the peak of the highest SVM was used to obtain accuracy. Tejaswini et al. (2025) employs Machine Learning for recognition of emotional conditions using internet expressions.

Proposed ways of detecting depression in X (previously Twitter) by Lu (2025) used Machine Learning and Neural Network-based modeling methods to find patterns that result into depressive behavior in user-generated textual data. Mao et al. (2014) have suggested a deep captioning model on Multimodal Recurrent Neural Networks (m-RNN) to automatically produce natural language description of a picture. The model combines visual characteristics obtained with the help of Convolutional Neural Networks (CNNs) with linguistic ones obtained with Recurrent Neural Networks (RNNs). This is because processing visual and written information is done using a common multimodal layer. The m-RNN system allows one to generate sequential word prediction on the basis of image products. The experimental evidence showed a better performance in image captioning than in the traditional methods.

Opelt et al. (2005) observed certain limitations in the predictive accuracy of diagnosing true suicidality, despite the fact that their study validated the accuracy of distinguishing emotions associated with distress. The field was significantly advanced on automated depression detection using different Machine Learning algorithms (Pathak et al., 2022).

The team emphasized the way social media communication and the organization of online networks have the potential to display trends associated with depression (Russell et al., 2008). The explainable hybrid deep learning framework suggested by Zogan et al. (2022), was developed to identify depression on social media based on the extraction of multi-aspect textual features of Twitter data. Their method combines linguistic, semantic, and emotional cues to enhance predictive performance as well as decipherability of depression-related material. On the contrary, Sarwar et al. (2019) concentrated on the augmentation of content-based analysis with the relevant feature extraction methods showing that structured feature representations may enhance computational models to a large extent. Subsequently, an attention-enhanced attention-augmented convolutional neural network. In appraisals, the new architecture surpassed the empirical standards it has proven its superiority (Sukanya et al., 2016).

The NLP has been instrumental in developing machine learning models to detect depression. Tejaswini et al. (2024) described a depression detection model, which is a framework of text analysis (social media) utilizing NLP and a hybrid deep learning model. To recognize effectively depressive behavior patterns, they used a combination of linguistic preprocessing, semantic feature extraction, and deep neural architectures. The hybrid model proposed showed better detection accuracy than the traditional machine learning feature, especially when dealing with low-resource and informal language features, which are frequent in the social media environment. The work notes the usefulness of applying NLP methods in combination with deep learning, which can be used to deal with the complexity and variability of the textual data that pertains to mental health.

Some researchers including Resnik et al. (2015) have worked on the detection of depression through social media data by applying a broad spectrum of Machine Learning and Deep Learning tools. Some of the earliest attempts involved the use of more traditional classifiers, like SVMs, Naive Bayes, Decision Trees, and KNN on services like Twitter, Facebook, YouTube comments, and microblogs, with fair results (as seen in inaccurate predictions) and with drawbacks, including account dependency, un-generalization, and poor performance measures. More current models used Deep Learning networks, such as CNNs, Long Shot-term Memory (LSTM), and Gated Recurrent Units (GRUs), mixed CNN LSTM, attention, and ensemble learning networks, which enhanced detection and feature representation. Nevertheless, they were normally computationally complex, not extensively tested on a wide variety of data, and were often simply concerned with accuracy but did not care about precision, recall and real-time usefulness.

Unlike the existing methods, the proposed work presents a systematic NLP-based Machine Learning model, which systematically test a variety of classifiers, such as KNN, Naive Bayes, Decision Tree, Random Forest, SVM, and Logistic Regression, in a unified experimental condition. In contrast to the previous research, which considers a single or hybrid Deep Learning model, this paper focuses on the comparison of models, interpretability, and robustness. Moreover, the suggested system facilitates the uninterrupted tracking and visualization of the emotional condition of the users over a period of time, which is one of the primary drawbacks of the previous studies and improves its applicability to the real-world context of the early detection of depression and monitoring mental conditions.

Proposed Methodology

Depression detection in Twitter data using Machine Learning approaches is an intensive procedure focused on achieving specific goals within the suggested work is presented in Figure 1. An extensive review of numerous Machine Learning techniques including Decision Tree, KNN, Random Forest, Naive Bayes, SVM, and Logistic Regression, all integrated with Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction, is conducted at the start of the research. An overview of literature on identification of depression is provided discussing the effectiveness and restrictions of previous researchers' efforts. The subsequent sections outline the objectives of the research which aim at refining Machine Learning algorithms, analyzing depression indicators in language, and studying cross-cultural elements.

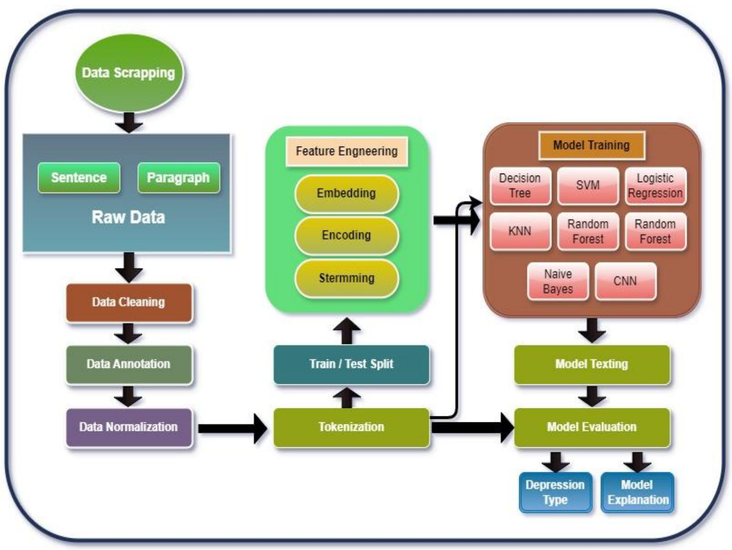

Figure 1: Proposed Methodology for Depression Detection.

Data Sets Description

The use of the compilation of depression and mental health-related data by Kaggle dataset coverage keywords, this study is focused on the coverage and quality of data. Although significant, raw text entails the cleaning or standardization that TF-IDF improves by feature extraction. A stratified split is then done on the data which is then divided into teaching and experiment groups. The choice of some of the best algorithms includes

Logistic. The simplest Regression that is optimized and trained is regression and one put to test on the data, as well as others. These algorithms then these are checked against obtained metrics: accuracy, precision, recall, F1 score and ROC-AUC. The systematic methodology can be explained by the hidden presented in the end of the research advantages and problems of each model, yet simultaneously points out the complexities that are involved in detection of sadness in Twitter data.

Out of the research process, we can affirm that the Logistic regression is the best method, demonstrating its effectiveness in estimating binary classification probabilities. The holistic research method makes a useful contribution to the difficult business of identifying patterns of depression in social media text, which are capable of inspiring new studies and practical advances in the field of mental health diagnostics.

Machine Learning Models

Decision Tree

An efficient model for identifying depressive states in tweets is created by combining the TF-IDF technique with a Decision Tree algorithm. Each node in the Decision Tree, a tree-like structure used in this integrated approach, represents a decision based on attributes that were scored using TF-IDF weighting. Because terms are scarce throughout the corpus and occur in a particular tweet, TFIDF gives them significance. The significance of non-linear interactions is indicated by the Inverse Document Frequency. The application is far more effective at identifying interactions in text. As the algorithm processes and categorizes complex content, it achieves accuracy in capturing subtleties of text; consequently, Anavi et al. (2015) claimed that TF-IDF improves the performance of the Decision Tree algorithm. By giving different weights to phrases that occur frequently in a tweet but infrequently in the text, this method improves the accuracy of the model.

K-Nearest Neighbors

Applications like depression detection in tweet analysis are based on the simple instance-based learning that leads to small and interpenetrated data isolated by the model, known as KNN algorithm. In particular, as suggested (Lacoste et al., 2007). KNN classifies just using a search of the nearest observations, where distances are calculated in an adequately crafted feature space. This localized context allows the microblogging algorithm to capture small changes in mood (Lu, 2025). In practice, KNN calculates the probability of a tweet being in label t based on its distance (also called similarity) to each neighboring tweets as they all use similar set of features. The process is optimized with the help of as weight wording, consideration of the TF-IDF method. The method works in multiple steps by initially matching a small amount of so called "seed" words with discrete frequency profiles that are then used to accumulate term scores which are descriptive of that frequency pattern, which in turn maintains the learner attentive to more considered and potentially signal-rich terms (Li et al. 2022). The ensuing integration of both KNN and TF-IDF geometries and semantic focus provides the model that skillfully combines the neighborhood-based judgment and clinical importance of token selections. The collaboration is required in the social channels where topical slang, emotive coding and user stylistics undergo almost daily development, which is supported by Lowe et al. (2004) in KNN augmented with TF IDF, when presented with tweets, however, highlights cluster successfully of latent melancholy, translating fleeting, individual expressions into coherent diagnostic inferences and thereby navigating the ambient arrogance of online language.

Random Forest

In case of depression detection with the help of tweets, the algorithm of Random Forest is an ensemble-based algorithm that constructs a large number of decision trees when training. The decision trees are constructed individually which brings variety to the learning process (Mao et al., 2014). Predictions of both the trees are then summed up leading to a final prediction based on the mode of the classes. Random forests are better than single decision trees and most other Machine Learning methods due to their superior reliability and accuracy since they make an average across the predictions of the trees. When implementing the Random Forest to detect depression, the algorithm does not only find various patterns in the text data, it does not even over fit, unlike most Machine Learning models (Opelt et al., 2005). Random Forest has better generalization because of its generalization ability which makes it more applicable in identifying even the most hidden depressive symptoms on twitter. The TF-IDF method also randomizes the basic model of text classification of tweets to boost the significance of words that are more important. Moreover, the integration is based on the fact that the algorithm is able not only to identify key features but also to utilize the numerous decision trees (Pathak et al., 2022). Random Forest is easy to use with the promise of reliability and accuracy, which makes it appealing in the diagnosis of depression and especially in a complex environment of social media data. Random Forest with TFIDF is effective with the use of ensemble learning to determine patterns associated with depression with high accuracy.

Naïve Bayes

The method is effective because of its simplifying assumption of independence of features which are easier to compute and also efficient. Naive Bayes is applied mainly to detect the depression in the context to extract the features in tweets, it includes TF-IDF weighting of words (Pham et al., 2018). Naive Bayes method has the ability to support the dynamic and the very diverse nature of social media posts because it correctly calculated the probability of a tweet which falls into a specific category of posts, according to assumption. The usage of this method and TF-IDF also improves the ability of this model to calculate the relevance of the variables which will describe depressive moods. TF-IDF asses for us the strength of the words by calculating the frequency with which they used in a tweet with the compared whole collection, enhanced evaluation of the each and every content of the tweet. Naive Bayes has a good potential and also it is a promising technique in detection of depression in tweets due to its good and greater probabilistic basis and efficiency in computation. It is recognized to as a powerful method when it comes to handling high dimensional data, that's why it can easily identify patterns that are indicative of the depressive conditions, which is characterized by a myriad of variables in a myriad of interrelations, which is the reason is why it is effective (Platt, 1999).

Support Vector Machine

SVM is one of the strongest classification algorithms as this method can easily identify the optimal hyperplane which classify the given information into classes in multi-dimensional data space. SVM is a power method to work with complicated data, this method increases the quality and also performance of the classification of patterns that are related to the depressive states in the social media content like, tweets and posts etc. When we use the SVM in identification of depression, the algorithm very easily identifies the required hyperplane that optimally separates depressive and non-depressive tweets. This identified hyperplane can be used as a decision boundary which is the main reason of accuracy in the classification on the feature space which is a product of the content tweets. The agility of this method to accept the multi-dimensional data sets is better in addressing the volatile and volume nature of data of social media. Moreover, SVM model is accurate and refined with the help of the TF-IDF because it attaches a weight to the terms in accordance with their characterization of the content's social media. TF-IDF increases the performance of the algorithm in high-dimensional data since it also highlights the role of the certain words in representation of an exceedingly difficult linguistic patterns which is connected to the concept of depression. The fact that SVM can categorize, and process complicated data is universally applicable in the precise identification of depression patterns in domain of tweets (Chicco et al., 2021).

Logistic Regression

Logistic Regression is a model for binary classification that will help to identify depression in tweets. The algorithm provides the probability of a sample belonging to a given class. Thus, it can be applied to detect the pattern behind depressive conditions in social media. In the case of Logistic Regression in detecting depression, the algorithm makes a decision of classifying a tweet as either a depressive or non-depressive based on its features. The logistic function takes the linear combination of the features and compresses it into a score in probability space which is much more convenient to classify the observations into the two classes. This is why Logistic Regression can be applied especially well to those questions that involve two potential solutions, such as identifying the signs of depression in tweets. Although it is linear model, it is quite impressive how adaptable the Logistic Regression can be to the variety of content in the social media. The application of TF-IDF technique enhances the model to produce the important phrases and language patterns associated with depression. Logistic Regression also has the benefit of being used in the sensitive process of identifying depression since it offers probabilistic estimates, interpretable values, and understanding of whether a tweet can indicate depressive symptoms. We have used these models in order to examine the dataset and reveal the trends concerning depression in tweets as pointed out by Sarwar et al. (2019). Solving the issues in the model performance during the evaluation process contributes to the optimization of the models, which makes them work at their highest level in the future measurements.

Convolutional Neural Networks

The Deep Learning technique known as CNNs as Figure 2 has recently been applied to the task of diagnosing depression from tweets. With the ever-growing amount of data on social media, it can be rich and complex. This is a challenge for traditional linear models, but not for CNNs which learn hierarchical representations of data to capture intricate patterns. For diagnosis of depression, worrying tweets can be processed through CNNs which help in accurately identifying depressive moods by automating the information extraction process. To capture sadness in texts, translation to a processable format is done and CNNs are then able to detect it. As is common in the processing of tweets, the text is first translated into word embedding's, which convert it into a continuous vector space. Thus, the CNN processes these embeds through several convolutional and pooling layers, which detect local dependencies and important signals indicating potential onset of depression. Through the use of different filters, CNNs are capable of capturing a number of linguistic patterns associated with sadness, be it particular phrases or phrases of words. The ability of CNNs to extract both low-level (single words) and high-level (word and phrase combinations) features is crucial in differentiating between depressive and non-depressive tweets (Miao, 2023). In addition, the use of modern model optimization strategies ensures that a high degree of refinement has been achieved, which, along with the thorough evaluation process, ensures that model performance is at its peak level.

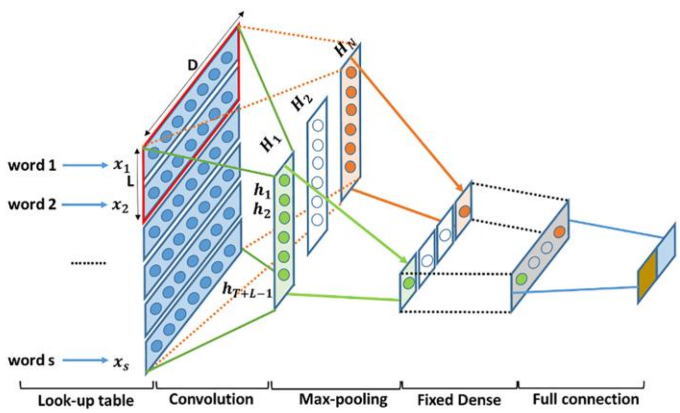

Figure 2: Standard CNN on text classification.

Incorporating CNNs into the depression detection workflow offers a significant benefit because they can process large datasets as well as complex datasets, capturing intricate details that simpler models may overlook. If the models of CNNs can be trained on a large dataset of tweets, it may be possible to detect patterns associated with depression, thus enabling the automated detection of depression from social media content.

Performances Matrix

Using the following formulae, accuracy measures how well a model is able to identify all classes (or pixels) accurately, in both the positive and negative cases:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Sensitivity (SEN) evaluates how well the model can detect positive patterns by measuring the rate of true positive pixels that were correctly detected from the pool of actual positives.

$$SEN = \frac{TP}{TP+FN} \tag{2}$$

Specificity it assesses the model by estimating the proportion of positive patterns, or pixels, out of positively detected patterns and the entire dataset that constitutes true positives.

$$SPE = \frac{TN}{TN+FP} \tag{3}$$

Recall (RE) measures the proportion of ground truth classes/pixels properly predicted by the model.

$$RE = \frac{TP}{TP+FN} \tag{4}$$

Precision (PR) The positive predictive value represents the frequency and proportion of correct class or pixel predictions made by the model.

$$PR = \frac{TP}{TP+FP} \tag{5}$$

F1-Score is the most often used statistic that combines precision and recall, indicating the harmonic mean of the two.

$$F1score = 2\frac{PR*RE}{(PR+RE)} \tag{6}$$

The Dice Similarity coefficient computes the spatial overlap between the actual tumor region as labeled in the ground-truth datasets and the corresponding area modelled and segmented by the algorithm. Zero overlap is where the two regions are completely disjoint and 1 overlap indicates full agreement between the two regions.

$$DS = \frac{TP}{\frac{1}{2}(2TP+FP+FN)} \tag{7}$$

In the present step of model selection, we are testing several classification algorithms to determine their efficacy in detecting depression. The available methods are Decision Tree, KNN, Random Forest, Naive Bayes, SVM, and Logistic Regression. This variety makes it possible for us to explore various modeling techniques and find the most suitable algorithm for accurately identifying signs of depression in the tweets under analysis. Different algorithms help us better understand their relative benefits and drawbacks, which helps us shape the selection process based on established metrics and how well the study achieves its objectives.

Results and Discussion

After training the dataset, evaluation metrics were calculated for the logistic regression model. With an accuracy of 91.8%, it is apparent that the model's forecasts are correct a good portion of the time. The precision score for recognizing positive predictions stands at 92.5% which shows that most of the model's predictions are accurate. Moreover, its recall score measuring capturing positive events currently sits at 90.8%. With these continual assessments, we can learn a lot about the model and its ability to predict and determine sentiment within the dataset. As with all models, these numbers are accompanied by visualizations like the confusion matrix and the ROC curve, the model's performance characteristics are elucidated more in depth. Figure 3 depicts our Logistic Regression model including a confusion matrix and a ROC curve.

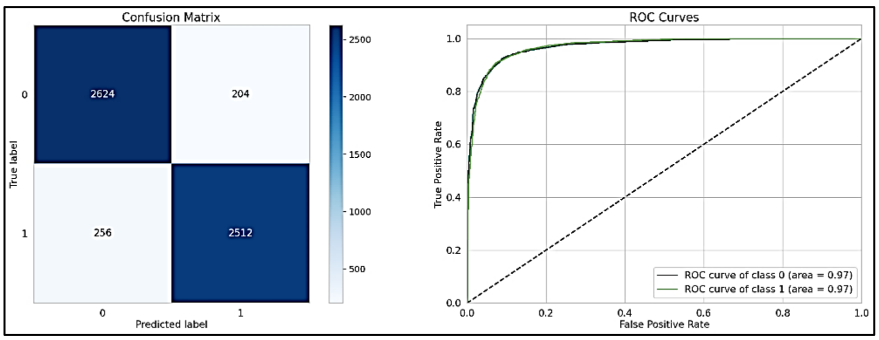

Figure 3: Confusion Matrix and ROC of Logistic Regression.

A Multinomial Naive Bayes classifier was trained on the dataset. As usual, the model is evaluated after training on key metrics. The accuracy of the NB model is currently at 88.8%, meaning that this is how many correct predictions it generates. The precision score, which evaluates the correct positive prediction, stands at 84.6% which indicates that the NB model was able to diagnose positive attitudes reasonably well. At the same time, recall score which evaluates the model's ability to capture positive instances also reported a good value of 94.7%. The combination of these measures is quite important for understanding the NB model's performance in sentiment analysis as well as its overall assessment capabilities on the provided dataset. Confusion matrix along with ROC of Naïve Bayes classifier are shown in Figure 4.

Once the dataset had been trained we measured the metrics of performance of the Decision Tree classifier. Accuracy of the Decision Tree model stands at 82.2 percent and this parameter implies that the percentage of accurate predictions made by the model on the dataset is 82.2 percent. Moreover, the score of precision, which is the accuracy of positive prediction, is also noted to be 82.2%. This number indicates the ability of the model to recognize cases of positive sentiment that exist. Moreover, the recall score is provided as 81.8% which is the capacity of the model to reflect the positive events. These measures are available to the sentiment analysis and the general accuracy measure since the model will be measured on various datasets. The Confusion matrix and ROC of the Decision Tree classifier is presented in Figure 5.

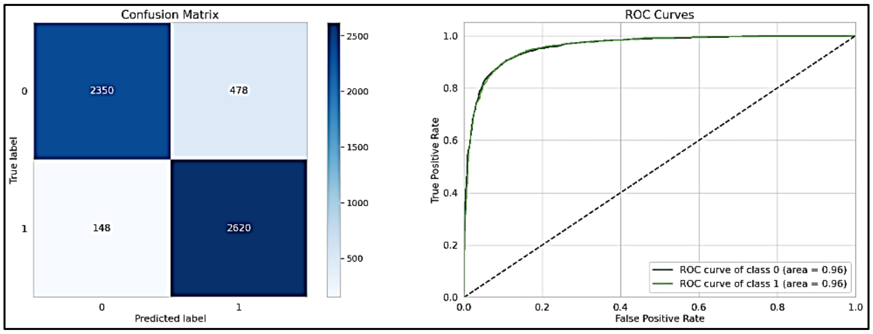

Figure 4: Confusion Matrix and ROC of Naïve Bayes.

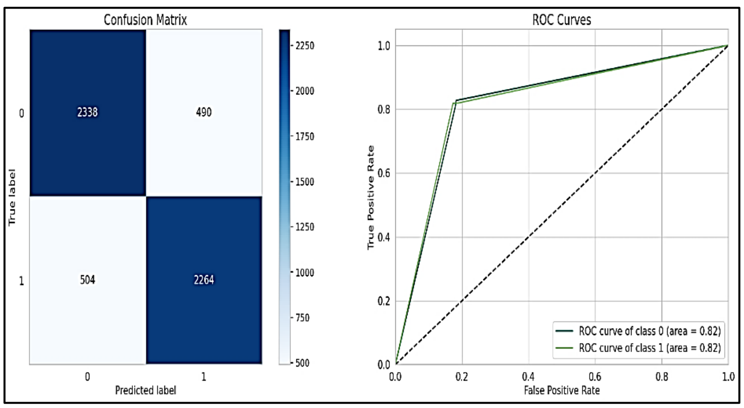

Figure 5: Confusion Matrix and ROC of Decision Tree.

A Random Forest classifier was trained on the dataset, and its performance was evaluated afterward. The model achieves an accuracy metric of 89.2 percent, which is the proportion of accurate predictions across the entire dataset. The precision score which defines the ratio of positive predictions out of total predicted positives is recorded as 88 percent. The accuracy shows how well the model is able to predict the presence of positive sentiment. In addition, the recall score which defines as the ability of the model to capture the said positive prediction stands at 90.4 percent. These metrics give a comprehensive understanding of the performance of Random Forest model in sentiment analysis and evaluation of the model's accuracy in making predictions over the dataset. Confusion matrix and ROC curve of RF classifier is presented in Figure 6.

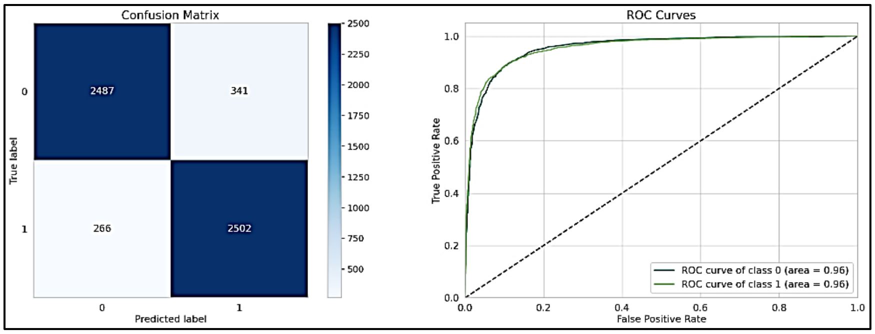

Figure 6: Confusion Matrix and ROC of Random Forest

The SVM model has been trained and evaluated on the specific dataset. It reached a predictive accuracy of 91.9% on the entire dataset, meaning 91.9% of the predictions were correct. Positive predictive accuracy, or precision in this case, is defined as measuring the correctness of the claimed positive outcomes, which in this scenario is at 92.6. In addition, the recall score captures the ability of the model to detect positive instances, which has been calculated at 90.9%. This set of metrics evaluates the performance SVM achieves in analyzing sentiment of text in the provided dataset. The Confusion matrix together with ROC of the SVM classifier is shown in Figure 7.

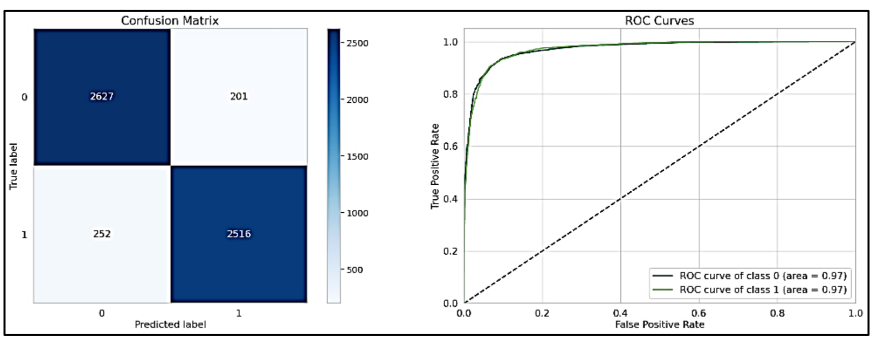

Figure 7: Confusion Matrix and ROC of SVM.

A KNN classifier is set up and trained with the dataset, but it appears that the model may not be suitable for sentiment analysis based on the accuracy, precision, and recall metrics. The supposed accuracy of the model sits at 50.8%, which reflects insufficient overall predictive performance. The precision score for positive predictions stood out as exceptionally high at 83.3%. However, recall rate, which reflects the model's ability to identify positive instances, plummeted to an astonishing 0.5%. The disparity between these two figures suggests the model is overly careful, or conservative, when making predictions, particularly limiting the true positive cases it identifies. Therefore, the KNN model's performance resonated quite poorly for sentiment analysis using the dataset in question. Confusion matrix and ROC curve of KNN classifier are shown in Figure 8.

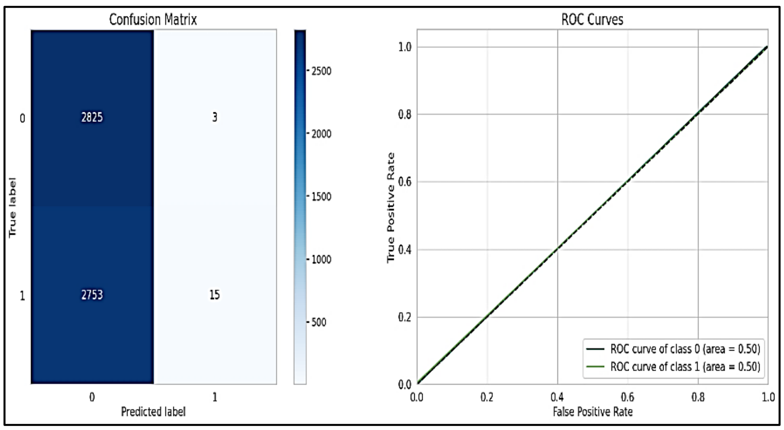

Figure 8: Confusion Matrix and ROC of KNN.

The confusion matrix confirms that the model performed well, yielding numerous correct predictions. It recognized 2473 cases as negative and 2545 cases as positive. Yet, as shown in Figure 9, it also generated 355 false positive predictions (negatives deemed as positives) and 223 false negatives (positives recognized as negatives). This demonstrates reasonably low error rates in both directions, indicating strong performance overall. From the graph of the training and validation accuracy, it can be deduced that even though the training accuracy increases progressively reaching a hundred percent, the validation accuracy stagnates and even declines after the fourth epoch. This indicates that overfitting is taking place as the model attempts to learn the peculiarities of the training data and is unable to generalize to new data.

The graphs for training and validation loss support this explanation. While the training loss decreases after each epoch because the model improves on the training set, the validation loss starts to increase after the fourth epoch. This further indicates overfitting wherein the model becomes incapable to perform well on new data as training improves.

The classifiers' results in a depression detection test are depicted in the Table 1. A different classifier is shown in each row, with accuracy, precision, and recall placed in the columns. The information provided in the table give an exhaustive overview of the strengths and weaknesses of the classifiers as far as depression evaluation is concerned. The depression detection classifiers' efficacy and the related measures shed light on different facets of their fitness for purpose.

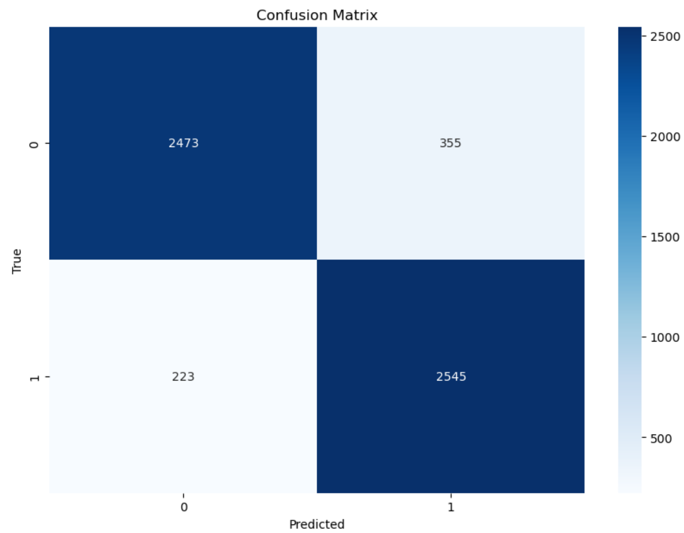

Figure 9: Confusion matrix shows the Overall model performance.

Table 1: Accuracy Comparison.

Classifier	Accuracy%	Precision%	Recall%	F1 score%
Logistic Regression	91.8	92.5	90.8	91.8
Naive Bayes	88.8	84.6	94.7	88.8
Decision Tree	82.2	82.2	81.8	82.2
Random Forest	89.2	88.0	90.4	89.0
SVM	91.9	92.6	90.9	91.9
KNN	50.8	83.3	0.5	0.11
CNN	90	90.5	90.3	90

In the challenge of identifying depression, several Machine Learning classifiers were implemented to determine its presence based on text information. The results generated are particularly impressed by the outcomes of the logistic regression model, which scored an accuracy of 91.8%, precision of 92.5%, and recall of 90.8% and F1 score of 91.8%. Both the logistic regression and Naive Bayes models performed impressively, with the latter scoring 88.8% accuracy, 84.6% precision and 94.7% recall. Even though the decision tree achieved only 82.2% accuracy, precision, and recall were exactly the same at 82.2% and 81.8%, respectively. The random forest classifier also performed moderately well with 89.2% accuracy, 88.0% precision and 90.4% recall. While SVM did not outperform the logistic model, its results were also very good at 91.9% accuracy, 92.6% precision, and 90.9% recall. On the other hand, the KNN showed a dismal accuracy of 50.8%, but had 83.3% precision and 0.5% recall. In summary, these results illustrate the varying capabilities different models have in predicting depression, thus helping to make better choices when trying to find an appropriate classifier for depression detection.

By comparing the findings of this paper with the literature that is already available as summarized above, one can see that the results are very consistent with the previous studies as well as there are some significant improvements. As presented by the previous literature, conventional machine learning methods (SVM, Naive Bayes, decision trees, etc.) have been intensively used to detect depression in social media and text-based data sets with moderate to high reliability depending on the dataset and feature representation. Similar to the studies by Pathak et al. (2022) and Zogan et al. (2022), margin-based models like SVM, and Naïve Bayes, perform well in high dimensional space of text, but the findings of these studies indicate that logistic regression is a better exposed predictor than SVM in terms of balanced accuracy, precision, recall and F1-score. This observation supports results of earlier studies that the simpler linear models can be extremely effective in situations where features have been well developed. In line with the findings of Lu (2025) and Miao (2023), Naive Bayes has high recall, thus is applicable in detecting the cases of depression, but at the expense of higher false positives. Ensemble and tree based models, including random forest and decision trees, show moderate results, as per the results found in previous reports, which might indicate the inability to work with sparse textual representations. The underperformance of KNN is corroborated by the fact that it was rarely used in the literature reviewed by Babenko et al. (2015), is further indication of the fact that this algorithm is not the best tool in terms of detecting depressions. In general, the results, in comparison to the previous work, such as deep and hybrid models, reveal that well-regularized linear and margin-based classifiers can be trained to achieve competitive or superior performance, using lower computational complexity, highlighting the significance of the simplicity and interpretability of the model, as well as balanced evaluation metrics. All in all, the results indicate that less sophisticated, highly regularized models can be useful than the more sophisticated methods, and that mixed consideration measures are vital in mental health implementations.

Conclusion and Future Work

The primarily goal of this research is to detect depression within the Twitter data for that we apply multiple Machine Learning classifiers. The methodology that we adopt has given inputs into the abilities of various classifiers as well as text cleaning, training the models that we use and evaluation of them. Logistic Regression and SVM achieved the highest accuracy with 91.8% and 91.9%, Respectively for this research. Particularly, Logistic Regression showed high precision of 92.5% and recall of 90.8%, Shows us the effectiveness in identifying sadness in Twitter data. These results illuminate the important aspects for the classifier's deployment and also refine understanding when classifier will perform in practical and real-world scenarios. Even though KNN only had a very low accuracy of 50.8%, but its precision in correctly classifying depressive tweets was very good and significantly higher at 83.3%. These different results indicates that different classifiers can be more suited for aspects when it comes to depression detection. Due to this reason we need to use appropriate classifiers based on evaluation criteria in mental health analytics. So, these insights very are critical when implementing programs in real-time settings for the detection of depression. Since social media platforms are central to people's self-expression, the conclusions of this research can be used to enhance the monitoring and support for depression from a technological standpoint.

Based on the study's results, there are some suggestions for enhancing the application of Machine Learning while detection of depression using Twitter data. As a first step, based on research findings it recommended to use of either Logistic Regression or SVM due to their better performance and ease of use. These recommended classifiers are not only precise but also, they have high rates of accuracy, precision, and recall, which is the reason, they can be used effectively in mental health monitoring systems. Also, given the unique characteristics of each classifier, a combination of hybrid or ensemble methods to increase the reliability and precision of multiple models may be utilized.

The exploration of model interpretability remains an important avenue for future inquiries. An understanding of how the Machine Learning classifiers, particularly those based on mental health detection, operate internally will bring confidence and encourage the use of such tools. Also, it is relevant to determine the transferability of models to other groups of populations and cultural environments to ensure their fair and inclusive use. Continuous adjustment of the models and social media algorithms by optimizing the preprocessing methods and feature selection strategies can be used to keep up with the changing language usage. Finally, longitudinal research devoted to the monitoring and analysis of how individual users can express mental illness concerns throughout time may be involved in the process of comprehending the online discussion of mental health concerns. These questions would be useful in moving forward of technologies to monitor and support users with their mental health needs.

References

Ahmed, A. (2021). Pre-trained CNNs models for content based image retrieval. *International Journal of Advanced Computer Science and Applications*, *12*(7).

Anavi, Y., Kogan, I., Gelbart, E., Geva, O., & Greenspan, H. (2015, August). A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. In *37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2940-2943). IEEE.

Babenko, A., & Lempitsky, V. (2015). Aggregating deep convolutional features for image retrieval. *arXiv preprint arXiv:1510.07493*.

Babenko, A., Slesarev, A., Chigorin, A., & Lempitsky, V. (2014, September). Neural codes for image retrieval. In *European conference on computer vision* (pp. 584-599). Cham: Springer International Publishing.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access*, *9*, 78368-78381.

Dubey, S. R. (2021). A decade survey of content-based image retrieval using Deep Learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(5), 2687–2704.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*(1), 98–136.

Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, *8*(1), 1927469.

Kulkarni, S., & Manu, T. M. (2022). Content-based image retrieval: A literature review. *In Proceedings of the 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1580–1587). IEEE.

Lacoste, C., Lim, J. H., Chevallet, J. P., & Le, D. T. H. (2007). Medical image retrieval based on knowledge-assisted text and image indexing. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*(7), 889–900.

Lu, X. (2025). Detecting Depression via Tweets on Twitter Utilizing Machine Learning and Neural Network Models. *Journal of Artificial Intelligence and System Modelling*, *3*(02), 1-14.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., & Yuille, A. (2014). Deep captioning with multimodal recurrent neural networks (m-rnn). *arXiv preprint arXiv:1412.6632*.

Miao, R. (2023). Emotion analysis and opinion monitoring of social network users under deep convolutional neural network. *Journal of Global Information Management (JGIM)*, *31*(1), 1-12.

Opelt, A., & Pinz, A. (2005). Object localization with boosting and weak supervision for generic object recognition. I*n Proceedings of the Scandinavian Conference on Image Analysis* (pp. 862–871). Springer.

Pathak, D., & Raju, U. S. N. (2022). Content-based image retrieval for super-resolution images using feature fusion: Deep Learning and handcrafted features. *Concurrency and Computation: Practice and Experience*, *34*(22), e6851.

Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality* (pp. 99-107).

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1–3), 157–173.

Sarwar, A., Mehmood, Z., Saba, T., Qazi, K. A., Adnan, A., & Jamal, H. (2019). A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine. *Journal of Information Science*, *45*(1), 117–135.

Sukanya, C. M., Gokul, R., & Paul, V. (2016). A survey on object recognition methods. *International Journal of Science, Engineering and Computer Technology*, *6*(1), 48.

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

Tejaswini, V., Sahoo, B., & Babu, K. S. (2025). Major Depressive Disorder Symptoms Detection System Through Text in Social Media Platforms Using Hybrid Deep Learning Models. *IEEE Transactions on Computational Social Systems*.

Vasha, Z. N., Sharma, B., Esha, I. J., Al Nahian, J., & Polin, J. A. (2023). Depression detection in social media comments data using Machine Learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *12*(2), 987-996.

Xie, L., Hong, R., Zhang, B., & Tian, Q. (2015). Image classification and retrieval are one. *In Proceedings of the 5th ACM International Conference on Multimedia Retrieval* (pp. 3–10). ACM.

Zogan, H., Razzak, I., Wang, X., Jameel, S., & Xu, G. (2022). Explainable depression detection with multi-aspect features using a hybrid Deep Learning model on social media. *World Wide Web*, *25*(1), 281-304.