Social Media Sentiment Analysis for Depression Detection using Machine Learning

Prof. Prakash P., Vidya Saurabh Mishra

School of Computer Science and Engineering (SCOPE)

Vellore Institute of Technology, Chennai

prakash.p@vit.ac.in, vidyasaurabh.mishra2022@vitstudent.ac.in

**Abstract—Depression is one of the most common public health concerns. Depression is often undiagnosed in the initial stages. We used a dataset consisting of posts from social media. We have used machine learning models like Logistic Regression and Naïve Bayes for classification. The classification models have been evaluated using evaluation metrics like accuracy, precision, recall and f-1 score.**

***Keywords—sentiment classification; sentiment analysis; Naïve Bayes(NB); Logistic Regression; Natural Language Processing (NLP)***

1. INTRODUCTION

With the advent of social media platforms, individuals are increasingly sharing their thoughts and experiences online. This has led to an abundance of data that can be analyzed to gain insights into various aspects of human behavior, including mental health.

One such application is the detection of depression, a serious and widespread mental health disorder that affects millions of people worldwide. According to W.H.O., the economic loss that India is going to face due to mental health conditions, between 2012-2030, is estimated at USD 1.03 trillion [1]. Sentiment Analysis a process of determining and categorizing whether the sentiment of a given text is positive, negative or neutral.

Depression is often underdiagnosed because In India, mental health issues are even bigger taboo than in the West. The stigmas attached to mental illnesses ensure that people sweep things under the carpet and suffer in silence instead of speaking out and seeking help. However, the ubiquity of social media provides a unique opportunity to detect signs of depression in a non-intrusive manner. By analyzing the sentiment of social media posts, we can potentially identify individuals who are exhibiting signs of depression and may seek professional help.

In this paper, we are going to do the sentiment analysis for depression detection using machine learning techniques. We hypothesize that individuals suffering from depression exhibit certain patterns in their social

media activity and use of language that can be captured through sentiment analysis.

Our approach aims to contribute to the broader understanding of how mental health is reflected in our online interactions.

This paper is organized as follows: **Section II** provides a review of related works in the field of social media sentiment analysis for depression detection. In **Section III**, we describe our proposed methodology. **Section IV** contains the results of our study. Finally in **Section V**, we conclude this study.

1. RELATED WORKS

Various Machine learning techniques have been used for sentiment analysis. Some examples are Support Vector Machine (SVM) [2][3], Multilayer Perceptron (MLP) [4], Logistic Regression (LR) [5], Decision Tree (DT) [6], Naïve Bayes (NB) Classifier [7].

Amolik et al. (2016) used Feature-Vector and classifiers such as NB and SVM. Despite NB exhibiting higher precision but lower recall and accuracy, SVM outperforms in terms of overall accuracy. [8]

Choudhury et al. (2013) observed that social media tweets contain useful signals for characterizing onset of depression[9].

Tsugawa et al.(2015) revealed that frequencies of word usage, along with topic modeling, are useful features for the prediction model. [10]

Nadeem et al. (2016) utilized the bag-of-words approach for depression detection [11]

Davidov et al. [12] employed user-defined hashtags in Twitter posts for sentiment classification. They used punctuation, individual words, and patterns as distinct feature types. These were then used for sentiment classification.

1. METHODOLOGY

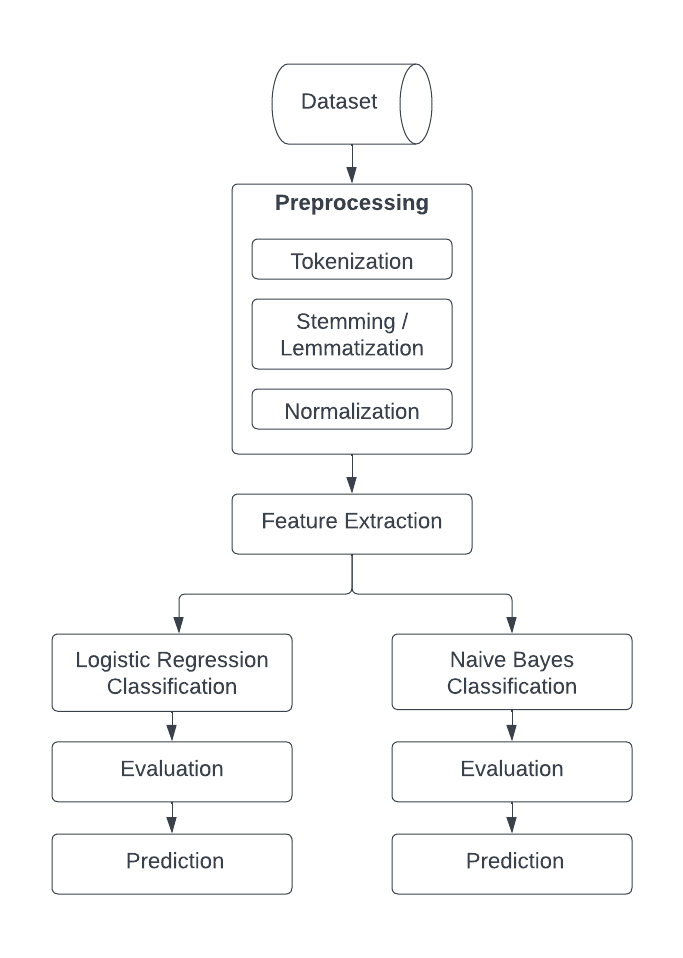
In this paper, the approach we used for sentiment analysis is explained in Fig. 1. The dataset used in this study is the Sentiment140 dataset, which contains 1.6 million tweets. Firstly, we did data preprocessing. Then feature extraction is done using Term Frequency-Inverse Document Frequency (TF-IDF), and classification is done using Logistic Regression and Naive Bayes classifiers.

Fig. 1. *The flowchart of the proposed model*

1. *Data Set Pre-Processing*

The Sentiment140 dataset underwent a series of preprocessing steps to ensure optimal performance of the classifiers. The text in the tweets was converted to lowercase. This is done to ensure uniformity in the dataset, as text data is case-sensitive. The dataset was cleaned to remove any URLs, user mentions, non-alphabetic characters. Tokenization was used to convert the normal text strings into a list of tokens. Tokenization allows the model to understand and analyze the text word by word. Any numbers, stopwords present in the tweets were also removed.

Stemming was performed on the tweets to reduce words to their root form. This was done using the Porter Stemmer provided by NLTK. Stemming improves the performance of the classifiers.

1. *Feature Extraction*

After preprocessing the data, the next step was feature extraction. In this study, the Term Frequency-Inverse Document Frequency (TF-IDF) was used for feature extraction. TF-IDF is a numerical statistic that signifies the importance of a word in a document within a collection or corpus [13]. TF-IDF has been effectively employed in numerous studies focusing on sentiment analysis for depression detection on social media [14].

1. *Classification*

In this study, we have conducted a comparative analysis of machine learning models. We used Logistic regression which is a linear classifier and it is a widely used classifier in Natural Language Processing (NLP) [15]. Logistic regression is based on sigmoid function which maps any real-valued number into a value between 0 and 1. That is why it is a good choice for predicting binary outcomes.

We also used another classifier i.e. Naïve Bayes which assumes conditional independence between pair of features, which is used for sentiment analysis [16]. Naive Bayes classifier is suitable for classification with discrete features.

1. *Evaluation*

Four evaluation measures (accuracy, precision, recall and f-measure) are used to evaluate the performance of classifiers. Their mathematical definitions are given in Eqs. (1), (2), (3) and (4) respectively.

(1)

(2)

(3)

(4)

In Eqs. (1), (2) and (3) TP, TN, FP, FN stand for true positive (TP), true negative (TN) false positive (FP) and false negative (FN) respectively. These metrics are used for comprehensive evaluation of our classification models. The Logistic Regression model and Naive Bayes models are trained and evaluated using these metrics.

1. *Prediction*

In this step, we used our classification models to predict the sentiment of an unseen text. The new input text is first transformed into a numerical format using the same vectorizer that was used during the training phase. This ensures that the new input data has the same feature representation as the training data.  Predictions represent the models’ best guesses for the sentiment of the input text, based on the patterns they learned from the training data.

1. RESULTS AND DISCUSSIONS

We used the training dataset of 1.6 million tweets to train Logistic regression and Naïve-Bayes classifier. The Logistic Regression model was trained with a maximum iteration parameter of 1000. The model achieved an accuracy of 0.78 on both the training and test data sets. The precision, recall, and F1-score indicated balanced performance of the model for both classes. The confusion matrix further confirmed the balanced performance, with similar numbers of true positives and true negatives.

The Naive Bayes model was trained using the Multinomial Naive Bayes algorithm. The model achieved an accuracy of 0.76 on the test data set. The precision and recall indicated a slightly imbalanced performance towards the positive class. The F1-score, which is the harmonic mean of precision and recall, was 0.76, confirming the model’s overall performance.

To evaluate the performance of our models, we used four standard evaluation measures including accuracy, precision, recall and F1 score.

Table I: Classification results (in %)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| Logistic Regression | 78.38 | 77.53 | 80.12 | 78.80 |
| Naive Bayes | 76.48 | 77.80 | 74.33 | 76.02 |

Both models performed well on the binary classification task, with Logistic Regression achieving slightly higher accuracy. The choice between these two models would depend on the specific requirements of the task. For instance, if the cost of false positives is high, the Logistic Regression model would be a better choice due to its higher precision. On the other hand, if the cost of false negatives is high, the Naive Bayes model would be a better choice due to its higher recall. These results are consistent with the findings of previous studies. For example, Ng and Jordan (2002) found that Logistic Regression tends to outperform Naive Bayes when the number of features is large, but Naive Bayes can perform better when the feature independence assumption holds [17].

1. CONCLUSION

This research shows that machine learning techniques can effectively be used to perform sentiment analysis for detecting depression. This research offers new ideas for designing more robust methods that leverage machine learning capabilities for detecting depression in textual datasets. The proposed method can be applied to other sectors such as the healthcare, corporate, leisure, public and private sectors to help them to understand their customers better, identify the relevant risks, and improve their products and services.

REFERENCES

1. World Health Organization. *Global Status Report On Non- Communicable Diseases 2010.* (2011). p. 1. Available online at: http://www.who.int/nmh/publications/ncd\_report\_full\_en.pdf.
2. B. Gokaraju, S. S. Durbha, R. L. King, S. Member, and N. H. Younan, “A machine learning based spatio-temporal data mining approach for detection of harmful algal blooms in the Gulf of Mexico,” IEEE J. Sel. Top. Appl. earth Obs. Remote Sens., vol. 4, no. 3, pp. 710–720, 2011.
3. T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis,” Assoc. Comput. Linguist., vol. 35, no. 3, 2009.
4. Alsagri HS, Ykhlef M (2020) Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features. IEICE Transactions on Information and Systems E103.D (8):1825-1832. doi:10.1587/transinf.2020EDP7023.
5. Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. In: Proceedings of 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, South Korea, 18-20 Oct. 2017. pp 138- 140. doi:10.1109/ICTC.2017.8190959.
6. Sato JR, Moll J, Green S, Deakin JF, Thomaz CE, Zahn R (2015) Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. Psychiatry Res 233 (2):289-291. doi:10.1016/j.pscychresns.2015.07.001.
7. Ojeme B, Mbogho A (2016) Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders. Procedia Computer Science 96:1294-1303. doi:10.1016/j.procs.2016.08.174.
8. Amolik, A., Jivane, N., Bhandari, M., and Venkatesan, M., 2016. Twitter sentiment analysis of movie reviews using machine learning techniques. International Journal of Engineering and Technology, 7(6): 1-7.
9. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Seventh International AAAI Conference on Weblogs and Social Media (2013).
10. Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., and Ohsaki, H. Recognizing depression from twitter activity in Recognizing depression from twitter activity 3187-3196 (ACM, 2015).
11. Nadeem, M. Identifying depression on Twitter. Preprint at arXiv:1607.07384 (2016).
12. Dmitry D, Oren T, Ari R. Enhanced sentiment learning using twitter hashtags and smileys. Coling 2010—23rd International Conference on Computational Linguistics, Proceedings of the Conference. 2. 2010; 241–249.
13. Babu, N.V., Kanaga, E.G.M. Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. *SN COMPUT. SCI.* **3**, 74 (2022). https://doi.org/10.1007/s42979-021-00958-1.
14. Bokolo, B.G.; Liu, Q. Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques. Electronics **2023**, 12, 4396. https://doi.org/10.3390/electronics12214396.
15. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. Applied Logistic Regression; John Wiley & Sons: New York, NY, USA, 2013.
16. Singh, S., Chandra, S.K. (2023). Sentiment Analysis for Depression Detection and Suicide Prevention Using Machine Learning Models. In: Garg, L., *et al.* Key Digital Trends Shaping the Future of Information and Management Science. ISMS 2022. Lecture Notes in Networks and Systems, vol 671. Springer, Cham. https://doi.org/10.1007/978-3-031-31153-6\_36.
17. Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, 2, 841-848.