iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57534

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Mental Health Prediction using Sentimental Analysis

Rishabh Verma[1], Nipun[2], Nitin Rana[3], Dr. Rakesh Kumar Arora[4]
Department of Computer Science & Engineering Dr. Akhilesh Das Gupta Institute of Technology and Professional Studies, Delhi, India

Abstract: ***The objective of this paper is to create a mental health prediction system through text sentimental analysis. The system will employ natural language processing technique to examine expressions and sentiments conveyed in text, aiming to predict the likelihood of an individual experiencing mental health issues such as anxiety, depression, or stress. The proposed system will make use of machine learning models to extract features that will provide with the results of sentiment analysis derived from text input. The integrated features will then be utilized to predict the likelihood of an individual experiencing mental health issues.***
Keywords: Machine Learning, Natural language processing, Text preprocessing, Sentimental Analysis.

I. INTRODUCTION

Mental health encompasses an individual's emotional, psychological, and social well-being, influencing thoughts, emotions, and behaviours. It plays a crucial role in how individuals manage stress, form relationships, and make decisions. The maintenance of optimal mental health is vital for overall well-being, contributing to a life that is fulfilling and productive. Mental health issues encompass a range of conditions that affect thoughts, emotions, and behaviour, often leading to distress and impairment in daily functioning. These conditions are diverse and can impact individuals of all ages, backgrounds, and walks of life.

Sentiment analysis is a computational process that involves evaluating and determining the emotional tone or sentiment expressed in a piece of text. This analysis is particularly valuable for understanding mental health status, public opinion, customer feedback, and overall attitudes towards various subjects. Text sentiment analysis involves leveraging machine learning techniques to automatically classify the sentiment expressed in textual data. Machine Learning based algorithms includes Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, Random Forest and Convolutional Neural Networks (CNNs) and more.

II. LITERATURE REVIEW

Though many researchers explored number of techniques for text sentiment analysis, machine learning based techniques are performing significantly well. This section analyzes some important research works performed by researchers, along with their outcomes.

Sahayak, et al. [1] proposed a machine learning algorithm with existing twitter dataset to sentiment analysis. The concept of sentiment analysis using a proposed approach that automatically classified the Tweets as positive, negative or neutral. They are also using Part Of Speech (POS)-Specific polarity features and a tree kernel. F. Ceci et al. [2] proposed a model for sentiment analysis is proposed which is based on ontology. The authors proposed a model work which intends to mix domain ontology with natural language process techniques to spot the sentiment behind judgments going to offer a description for such polarization. The methodology tests were developed by using two individual areas, digital camera, and movies [3]. J. Serrano-Guerrero et al. and E. Cambria [4]-[5] the methods of sentiment analysis are trained for detection sentiment polarity which can automatically track out sentiments from different documents, blog, sentences or words. R. Upadhyay et al. [6] developed a new method for semantic knowledge extraction from research documents and article using an integration of semantic technology, NLP, and information extraction. S. S. A. Q. Mahlawi [7] proposed a novel method which extracts structured data from emails by using data cleaning, data extraction, and data consolidation. S. Rosenthal et al. [8] proposed that the supervised learning approach is based on label datasets which are trained to provide meaningful outputs. To supervise the learning approach, apply Naive Bayes algorithm, maximum entropy and support vector machine which helps to achieve great success in sentiment analysis. P. Aruna et al. [9] showed that the main goal is to connect on Twitter and search for the tweets that contain a particular keyword and then evaluate the polarity of the tweets as positive and negative. The sentiments of the online tweets are evaluated based on feature selection of score words. S. Ghosh et al, [10] used K-Nearest Neighbor classifier to sentiment analysis and generated an averaged histogram model in the process that deals with text classification in continuous variable approach, containing a generalized feature representation in that particular class.

N. P. M. Vadivukarassi et al. [11] were presenting a survey on sentiment analysis of Twitter data with using different techniques. They used different machine learning algorithms, such as Naive Bayes, Maximum Entropy and Vector Machine support, to sentiment analysis and demonstrate the accuracy of different sizes of features.

III. PROPOSED FRAMEWORK

A. Technology Used

Sentiment analysis is a process of identifying and extracting subjective information from text data. It is used to determine the polarity of the text, i.e., whether the text expresses a positive, negative, or neutral sentiment. Sentiment analysis is a subfield of natural language processing (NLP) and machine learning (ML) that uses various techniques to analyze text data. These techniques include rule-based systems, machine learning algorithms, and deep learning models.

B. Flowchart for the proposed study

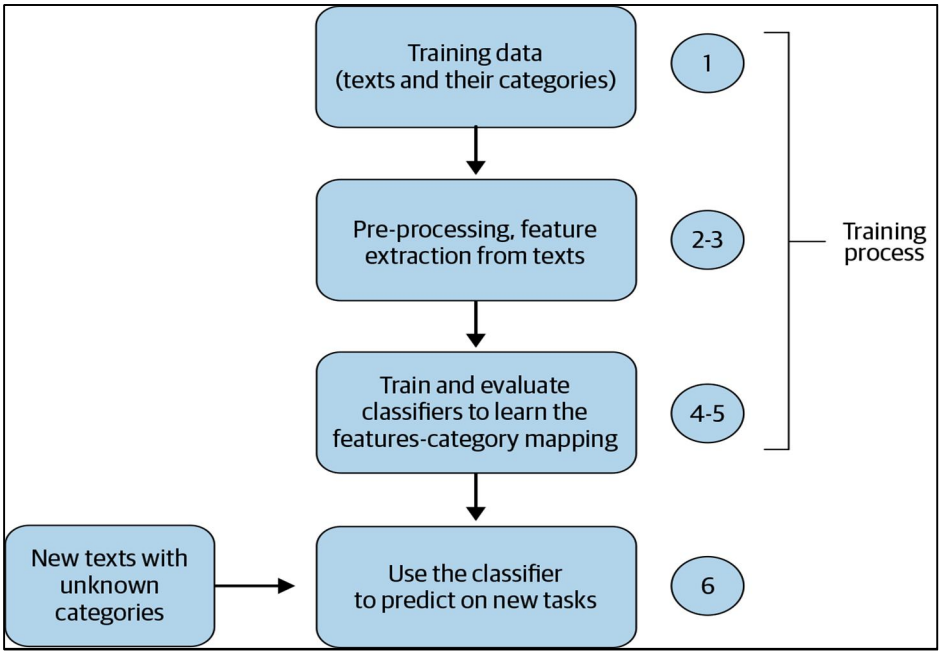


Fig. 1. Flowchart for the proposed study

1) Collect or create a labelled dataset suitable for the task.
2) Split the dataset into two (training and test) or three parts: training, validation (i.e., development), and test sets, then decide on evaluation metric(s).
3) Transform raw text into feature vectors.
4) Train a classifier using the feature vectors and the corresponding labels from the training set.
5) Using the evaluation metric(s) from Step 2, benchmark the model performance on the test set.
6) Deploy the model to serve the real-world use case and monitor its performance.

IV. METHODOLOGY

A. Dataset Used

Collecting a dataset of text inputs from individuals with and without mental health issues. The dataset will be used to train and test the models. Dataset used is from Mental Health Corpus: The Mental Health Corpus is a collection of texts related to people with anxiety, depression, and other mental health issues. The corpus consists of two columns: one containing the comments, and the other containing labels indicating whether the comments are considered poisonous or not. The corpus can be used for a variety of purposes, such as sentiment analysis, toxic language detection, and mental health language analysis. The data in the corpus may be useful for researchers, mental health professionals, and others interested in understanding the language and sentiment surrounding mental health issues.

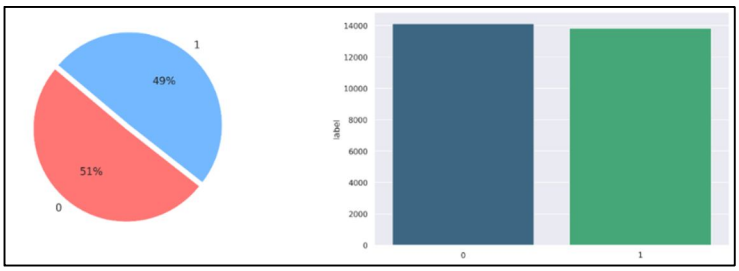


Fig. 2. Pie chart and Bar chart showing amount of positive and negative text data available (1-Negative, 0-positive)

B. NLP

Sentiment analysis of text using Natural Language Processing (NLP) is generally implemented using Machine Learning where the input text "Raw Text" is given and a list of emotion types are given based on which the emotion is determined.

Here's a simplified outline of the process:

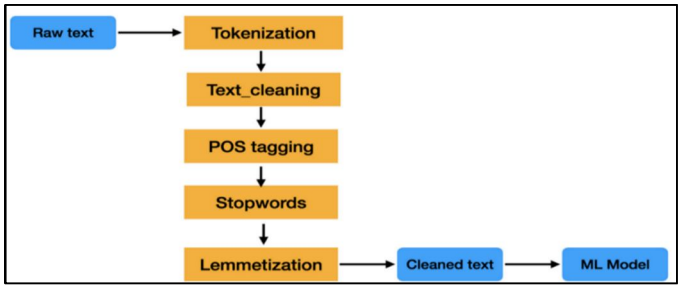


Fig. 3. Flowchart of NLP

1) *Tokenization:* Tokenization is a way of separating a piece of text into smaller units called tokens. This can be done at the sentence level or word level.
2) *Text Cleaning:* This phase deletes words and items from a corpus of text data to help enhance a machine-learning model 's efficiency. Numbers, capitalization, punctuation, stopwords, single quotes will be removed from the text data. Text cleaning process is done using regular expression.
3) *POS tagging:* POS tagging is a task of labelling each word in a sentence with its appropriate part of speech. Parts of speech include nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.
4) *Stopwords:* These are the words that do not add much value to the meaning of the document. Removing of stopwords must be done after tokenization.
5) *Lemmatization:* This reduces the inflected words with properly ensuring that the root word belongs to the language. It helps to get necessary and valid words. Nltk provides WordNet Lemmatizer that uses the WordNet Database to lookup lemmas of words.

V. APPLICATIONS

There are various applications of Mental Health Prediction using Sentiment Analysis such as:

Remote Patient Monitoring: Allows for the remote monitoring of patient's mental health, providing insights to healthcare professionals without the need for in-person visits, Public Health Surveillance: Enables public health agencies to monitor and analyses population-level mental health trends, informing the development of targeted interventions, Mood Tracking Apps: Integration into mobile applications that track users' moods over time, providing individuals with insights into their mental well-being, Employee Well-being Programs: Employers can use sentiment analysis to monitor the emotional well-being of employees, leading to the development of workplace mental health initiatives, Research and Clinical Studies: Provides valuable data for research studies exploring the relationship between language, sentiment, and mental health outcomes, Social Media Monitoring: Analyzing sentiment on social media platforms to identify individuals who may be in need of mental health support, Chatbot and Virtual Assistant Integration: Integrating sentiment analysis into chatbots or virtual assistants to provide more empathetic and personalized responses in mental health support scenarios.

VI. CONCLUSION & FUTURE SCOPE

The model's results depend on various factors, including the quality of the data, the effectiveness of the models. Using sentiment analysis allows for a much better understanding of an individual's mental health. The integrated model considers textual cues, capturing a richer set of features related to emotional well-being. Using appropriate metrics to quantitatively evaluate the performance of the emotion recognition and sentiment analysis models. This may include accuracy, precision, recall, Confusion matrix and ROC curve.

```
Accuracy of the model : 0.889
Precision score of the model : 0.879
Recall score of the model : 0.9
```

Fig. 4. Accuracy, Precision and Recall score

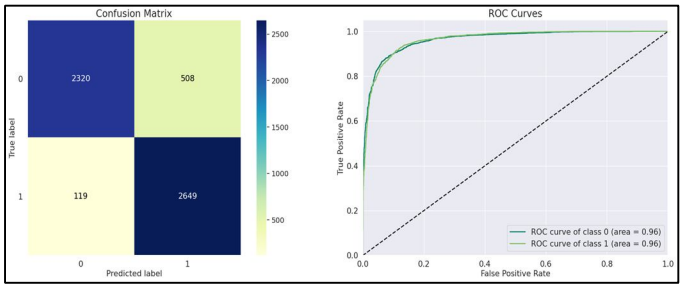


Fig. 5. Confusion matrix and ROC curve for Random Forest Classification Model

The proposed system's potential benefits include its ability to predict mental health conditions accurately and efficiently using sentiment analysis. This approach has the potential to overcome some of the limitations of existing approaches that use only one of these methods. If successfully implemented the proposed system could have a significant impact on mental health diagnosis and treatment particularly in areas where mental health professionals are scarce or inaccessible. The system's ability to provide early and accurate detection of mental health conditions could lead to earlier interventions and improved treatment outcomes for patients.

VII. FUTURE SCOPE

In terms of future scope, there are several areas of improvement for the system. Firstly, the data set used for training the model can be expanded to include a more diverse range of individuals including different age groups, ethnicities, and mental health conditions. Secondly, more advanced machine learning and deep learning techniques can be used in improving the accuracy and speed of modules. Thirdly. the system can be integrated with other mental health support services, such as online counselling or self-help resources to provide a more comprehensive and personalized approach to mental health care. Overall, the mental health prediction using sentiment analysis system has great potential to revolutionize mental health care by providing a more accessible and effective tool for identifying and addressing mental health issues.

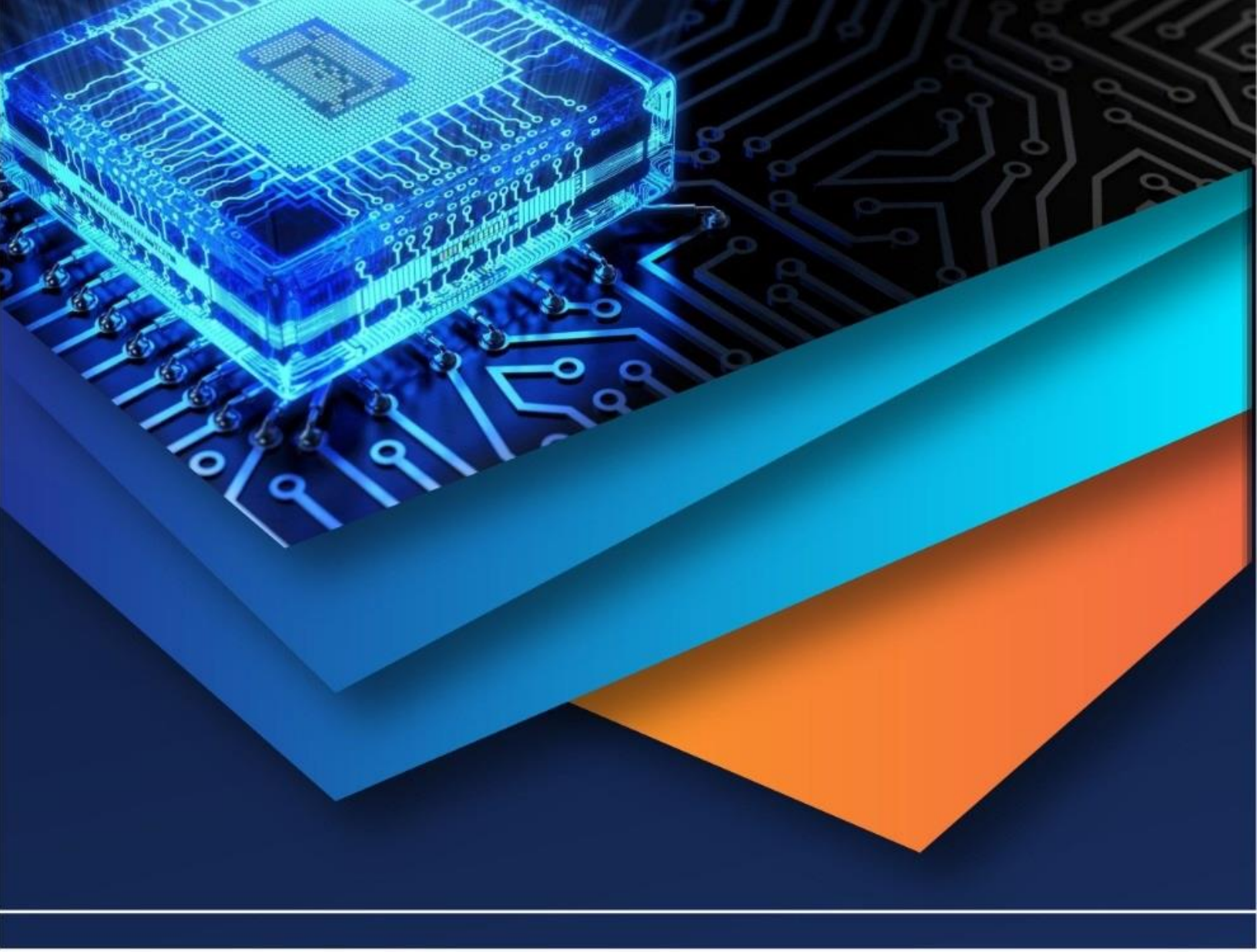
VIII. ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to all those who contributed to the completion of this research work. First and foremost, we extend our gratitude to Dr. Rakesh Kumar Arora, Professor at ADGIPS, for their valuable guidance and mentorship throughout the research process. This research work was made possible through the generous support of Dr. Akhilesh Das Gupta Institute of Technology and Professional Studies.Finally, we would like to acknowledge the entire research team, whose collaborative efforts and dedication have been invaluable to the success of this project.

REFERENCES

[1] V Sahayak, V Shete, A Pathan. "Sentiment analysis on twitter data," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, no. 1, pp. 178–183, 2015.
[2] F Ceci, AL Goncalves, R Weber "A model for sentiment analysis based on ontology and cases," IEEE Latin America Transactions, vol. 14, no. 11, pp. 4560–4566, 2016.
[3] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

[4] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. HerreraViedma, "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, vol. 311, pp. 18–38, 2015.
[5] E. Cambria, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.
[6] R. Upadhyay and A. Fujii, "Semantic knowledge extraction from research documents," in Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on. IEEE, 2016, pp. 439–445.
[7] S. S. A. Q. Mahlawi, "Structured data extraction from emails," 2017 International Conference on Networks Advances in Computational Technologies (NetACT), vol. 37, pp. 323–328, 2017.
[8] S. Rosenthal, N. Farra, and P. Nakov, "Semeval-2017 task 4: Sentiment analysis in twitter," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 502–518.
[9] N. P. M. Vadivukarassi and P. Aruna, "Sentimental analysis of tweets using naive bayes algorithm," World Applied Sciences Journal, vol. 35, no. 1, pp. 54–59, 2017.
[10] S. Ghosh, S. Pal "Sentiment analysis using averaged histogram," International Journal of Computer Applications, vol. 162, no. 12, 2017.
[11] N. P. M. Vadivukarassi "Sentiment analysis of twitter data: a survey of techniques," 2016.
[12] Chikersal P., Poria S., Cambria E., Gelbukh A., Siong C.E. Modelling public sentiment in twitter: using linguistic patterns to enhance supervised learning International Conference on Intelligent Text Processing and Computational Linguistics (2015), pp. 49-65.
[13] Kang K., Yoon C., Kim E.Y. Identifying depressive users in twitter using multimodal analysis 2016 International Conference on Big Data and Smart Computing (BigComp) (2016), pp. 231-238.
[14] Yadav P., Pandya D. SentiReview: Sentiment analysis based on text and emoticons 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (2017), pp. 467-472.
[15] Rakesh Kumar Arora, Manoj Kumar Gupta, Bhoopesh Singh Bhati Analysis of Various COVID-19 Prediction Techniques-IEIE Transactions on Smart Processing & Computing, 2021



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY