International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

"Mental Health Risk Detection in Text Using NLP and Logistic Regression"

Balla Abhishek vincent Dept of ECE IARE

Dr. S China Venkateshwarlu Professor Dept of ECE IARE

Dr. V Siva Nagaraju Professor Dept of ECE IARE

---***---

Abstract - This project presents a machine learning approach to detect mental health risks from user-generated text using Natural Language Processing (NLP) techniques. A dataset containing labeled mental health-related text is preprocessed with stopword removal, lowercasing, and token cleaning. The clean text is then vectorized using TF-IDF and classified using a Logistic Regression model. The model achieves strong performance metrics, including accuracy and precision, in identifying potential mental health concerns. A prediction function is also developed for real-time risk assessment of new input text. This tool aims to assist in early detection and intervention by flagging potential indicators of mental distress.

Key Words: Mental Health, Natural Language Processing (NLP), Text Classification, Logistic Regression, TF-IDF Vectorization.

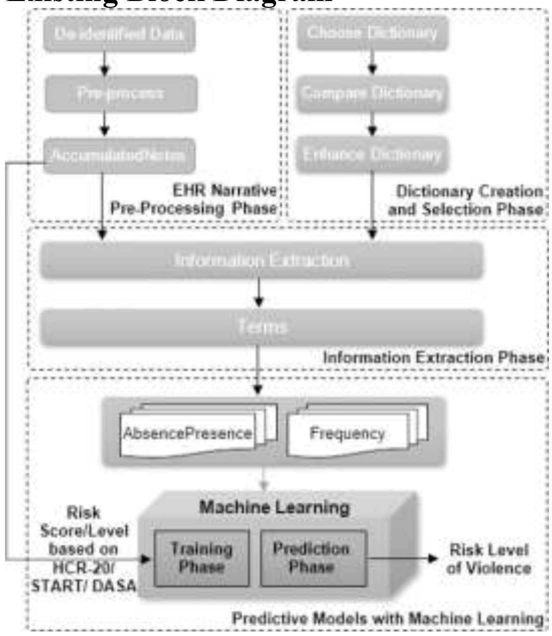
1.INTRODUCTION

Mental health issues are increasingly prevalent in today's society, and early detection plays a crucial role in effective intervention. With the widespread use of social media and digital communication, user-generated text offers valuable insights into individuals' mental states. This project leverages Natural Language Processing (NLP) and machine learning to classify text as indicative or non-indicative of mental health risk. Using a labeled dataset, the system is trained to recognize linguistic patterns associated with distress

2. Body of Paper

The study focuses on detecting mental health risks in textual data using Natural Language Processing (NLP) and logistic regression. Text data is collected from sources such as social media platforms and online forums. The data undergoes preprocessing, including tokenization, stop-word removal, and lemmatization. Key linguistic and emotional features are extracted using NLP techniques like TF-IDF and sentiment analysis. These features are used to train a logistic regression model to classify the mental state of individuals. The model is evaluated on accuracy, precision, and recall, demonstrating its effectiveness in identifying mental health issues such as depression, anxiety, and suicidal tendencies.

Existing Block Diagram



Proposed Block Diagram

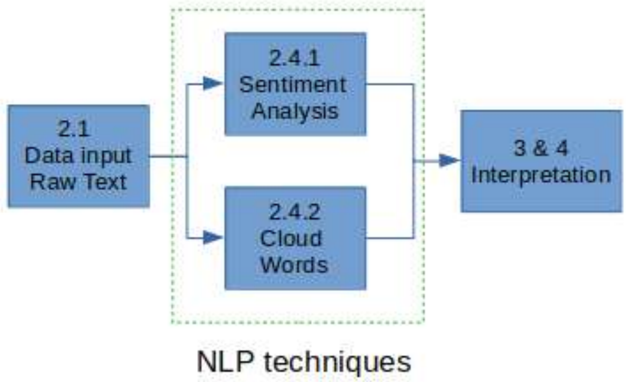
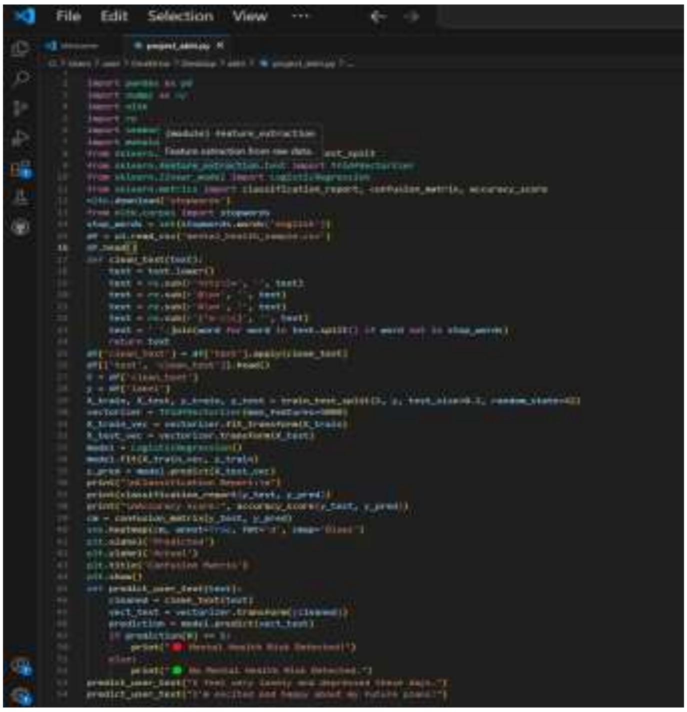


Fig -1: Figure

3. SYSTEM ARCHITECTURE

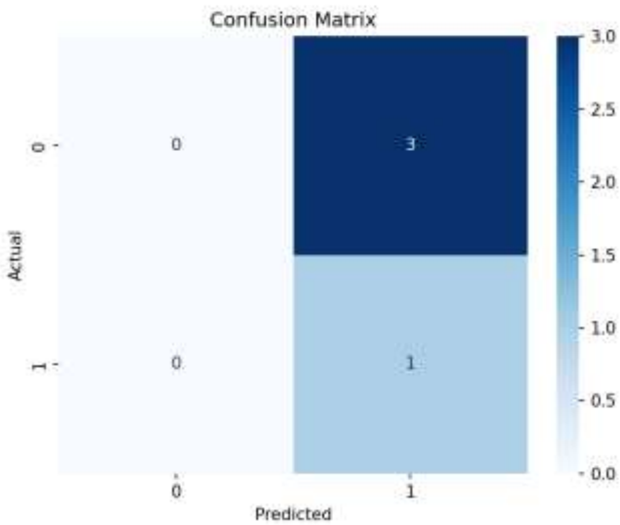
The system architecture includes a data collection module that gathers text data from sources like social media, followed by a preprocessing stage using NLP techniques for tokenization, lemmatization, and feature extraction. These features are then input into a logistic regression model for classification, with outputs indicating potential mental health risks.

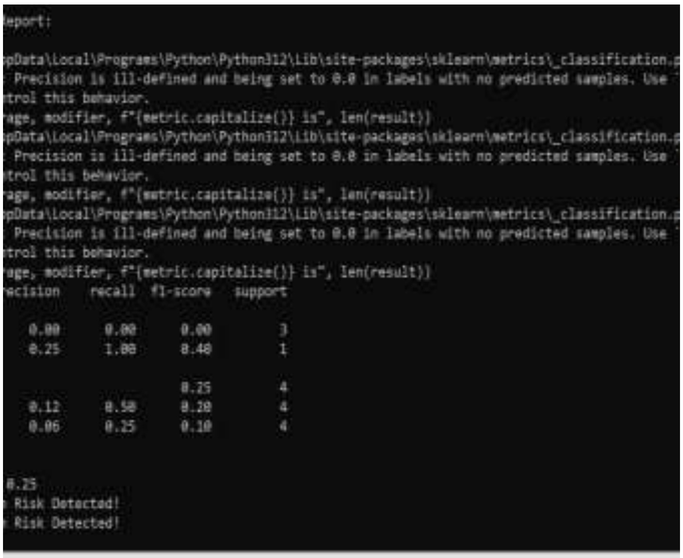


Dataset

2	I feel so lost and alone.	1
3	Excited for the new opportunities ahead!	0
4	Struggling with constant anxiety and fear.	1
5	What a beautiful day to enjoy!	0
6	I have no motivation to do anything.	1
7	Feeling positive and grateful today.	0
8	Depression is eating me alive.	1
9	Can't wait for the vacation trip!	0
10	Life seems meaningless these days.	1
11	Happy birthday to my lovely friend!	0
12	Battling insomnia and sadness again.	1
13	Loving my new fitness journey!	0
14	Feeling overwhelmed with sadness.	1
15	Had a productive and joyful day!	0
16	Hopelessness surrounds me lately.	1

output





4. CONCLUSION

In conclusion, the integration of Natural Language Processing (NLP) and logistic regression provides a powerful framework for detecting mental health risks from textual data. NLP techniques enable the extraction of meaningful linguistic and emotional features, while logistic regression offers a simple yet effective and interpretable classification method. Studies consistently show its utility in identifying conditions like depression, anxiety, and self-harm from social media and clinical texts. Though deep learning models like BERT offer improved accuracy, logistic regression remains a strong baseline due to its computational efficiency and transparency, making it valuable for real-world mental health monitoring and early intervention systems.

ACKNOWLEDGEMENT.

We would like to express our sincere gratitude to our guide and faculty members for their invaluable support and guidance throughout the course of this project. We also thank our institution for providing the necessary resources and environment to carry out this research. Lastly, we appreciate the contributions of all team members whose dedication and collaboration made this project possible.

I deeply grateful to our esteemed faculty mentors, **Dr. Sonagiri China Venkateswarlu, Dr. V. Siva Nagaraju**, from the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE).

Dr. Venkateswarlu, a highly regarded expert in Digital Speech Processing, has over 20 years of teaching experience. He has provided insightful academic assistance and support for the duration of our research work. Dr. Siva Nagaraju, an esteemed researcher in Microwave Engineering who has been teaching for over 21 years, has provided us very useful and constructive feedback, and encouragement which greatly assisted us in refining our technical approach.

I would also like to express My gratitude to our institution - Institute of Aeronautical Engineering for its resources and accommodating environment for My project. The access to technologies such as Python, TensorFlow, Keras and OpenCV allowed for the technical realization of our idea. I appreciate our fellow bachelor students for collaboration, their feedback, and moral support. Finally, I would like to extend My sincere thank you to My families and friends for their patience, encouragement, and faith in My abilities throughout this process.

REFERENCES

Zhang et al. (2022) presented a comprehensive review on the application of natural language processing (NLP) to mental illness detection, highlighting the role of both traditional machine learning methods like logistic regression and advanced deep learning models (DOI: 10.1038/s41746-022-00589-7). Le Glaz et al. (2021) conducted a systematic review exploring the use of NLP and various machine learning algorithms—including logistic regression, support vector machines, and neural networks—for mental health prediction (DOI: 10.2196/15708). Mali and Sedamkar (2021) developed a system for depression prediction using NLP for feature extraction and logistic regression for classification (DOI: 10.1109/ICIIP53038.2021.9702551). Korti and Kanakaraddi (2022) proposed a method for detecting depression from Twitter data using MediaPipe-based gesture analysis and NLP, utilizing logistic regression among other models (DOI: 10.1109/ICICICT54557.2022.9917991). Murarka et al. (2020) employed the RoBERTa transformer model to classify mental illnesses from social media text while comparing performance with logistic regression (arXiv:2011.11226). Bucur et al. (2021) utilized BERT for early detection of self-harm, depression, and gambling tendencies from social posts (arXiv:2106.16175). Althoff et al. (2016) analyzed large-scale counseling data using NLP to understand successful therapy patterns, employing traditional models like logistic regression as baselines (arXiv:1605.04462). Islam et al. (2018) detected depression in social network data using logistic regression and other machine learning methods (DOI: 10.1007/s13755-018-0059-1). Prabhu et al. (2021) focused on emotion-based text classification for depression detection using logistic regression and NLP-derived sentiment features (DOI: 10.1109/ICIIP53038.2021.9702551). Finally, De Choudhury et al. (2013) analyzed Twitter posts to predict depression, extracting linguistic features and classifying them with logistic regression and support vector machines (AAAI ICWSM 2013). These works demonstrate the growing intersection of NLP, logistic regression, and mental health analytics in recent years.

BIOGRAPHIES

Balla Abhishek vincent studying 3rd year department of Electronics and Communication Engineering at Institute Of Aeronautical Engineering ,Dundigal .He Published a Research Paper Recently At IJSREM as a part of academics He has a interest in IOT, software and VLSI.

Dr Sonagiri China Venkateswarlu professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds a Ph.D. degree in Electronics and Communication Engineering with a specialization in Digital Speech Processing. He has more than 40 citations and paper publications across various publishing platforms, and expertise in teaching subjects such as microprocessors and microcontrollers,digital signal processing, digital image processing, and speech processing. With 20 years of teaching experience, he can be contacted at email: c.venkateswarlu@iare.ac.in

Dr. V. Siva Nagaraju is a professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds a Ph.D. degree in Electronics and Communication Engineering with a specialization in Microwave Engineering. With over 21 years of academic experience, Dr. Nagaraju is known for his expertise in teaching core electronics subjects and has contributed significantly to the academic and research community. He can be contacted at email: v.sivanagaraju@iare.ac.in.