WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection and Estimation of Depression via Social Media by Deep Learning

Pokharkar Aishwarya Balasaheb[1], Auti Sakshi Sandeep[2], Naykodi Pradnya Chandrakant[3], Shinde Akash Dipak[4], Prof. Badhekar. P. M[5]*

[1,2,3,4]Department of Computer Engineering, Samarth Rural Educational Institute Polytechnic Belhe, India

[5] Department of Computer Engineering, Samarth Rural Educational Institute Polytechnic Belhe, India

A B S T R A C T

It is clear that students nowadays are always looking for ways to further their careers. However, in their pursuit of such chances, they frequently face obstacles such as restlessness, frustration, sadness, discontent, and despair, which have become inevitable features of their work environment. Depression has become a common problem in today's culture, impacting millions of students globally. As a result, it is critical to detect depression early on by utilizing deep learning and to provide counselling. During our recent encounters with students, we discovered that many of them are unsatisfied with their educational progress in their class.

The range of difficulties that students encounter, such as suicidal thoughts, sadness, loneliness, and the inferiority complex, makes it clear how seriously these problems should be taken. This initiative has been chosen in an effort to address and enhance students' mental health. It is vital to incorporate efficacious strategies for managing depression in order to ameliorate the present circumstances facing students. An efficient strategy for evaluating a student's level of depression is to employ statistical approaches such as Pearson correlation, K nearest neighbor, artificial neural networks, and decision trees. The efficacy of this strategy in predicting depression has been investigated.

Keywords: Depression Prediction, Decision Tree, Artificial Neural Network, K nearest Neighbours, Pearson Correlation.

1. INTRODUCTION

The relevance of a final project focused on estimating depression intensity via social media using a deep learning approach is multifaceted. As mental health awareness grows, the demand for innovative tools to detect and address issues like depression becomes increasingly critical. Social media platforms serve as a rich source of real-time data, allowing for the analysis of user-generated content to identify individuals in distress and facilitate timely interventions. By leveraging deep learning techniques, the project can uncover nuanced patterns in language, sentiment, and behavior that traditional methods may overlook, leading to more accurate assessments of mental health. Furthermore, automated systems can analyze vast amounts of data, enabling the monitoring of larger populations than manual approaches permit, thereby enhancing the scalability of mental health interventions. This analysis can also inform personalized support resources tailored to individual needs. However, it is essential to address the ethical considerations surrounding privacy and consent in such sensitive areas, ensuring that AI is used responsibly. Overall, this project not only contributes to academic research but also has practical implications for improving mental health support, bridging technology, psychology, and social science.

"Depression Intensity Estimation via Social Media: A Deep Learning Approach" focuses on leveraging the vast data generated on social media platforms to detect and estimate the severity of depression in individuals. With millions of users expressing thoughts, feelings, and experiences online, social media provides a unique opportunity to analyze mental health patterns using advanced computational methods. This approach utilizes deep learning algorithms, such as neural networks and natural language processing (NLP), to capture subtle cues, language patterns, sentiment changes, and behavior that might indicate depressive states. By examining textual data, images, and interactions, this technique aims to develop an automated and scalable system for depression estimation, offering valuable insights for mental health interventions, early warning systems, and public health monitoring.

1.1 Need

This final project focuses on estimating depression intensity through a deep learning approach using social media data. As mental health issues continue to rise, analyzing platforms like Twitter and Reddit offers critical insights into users' emotional states. We will collect and preprocess textual data while ensuring adherence to ethical standards and privacy regulations. By employing advanced deep learning models such as LSTM and BERT, we aim to classify the intensity of depression based on user-generated content. Our evaluation will utilize metrics like accuracy and F1 score to gauge the model's

effectiveness. Ultimately, this project aspires to enhance understanding of mental health dynamics in online environments and demonstrate how technology can play a role in mental health assessment and intervention.

1.2. Scope

The scope of this project encompasses the development of a deep learning model to assess depression intensity based on social media content. We will focus on collecting data from platforms like Twitter and Reddit, where users often express their thoughts and emotions. The project involves preprocessing this data to ensure it is suitable for analysis, followed by training and fine-tuning deep learning models to identify patterns indicative of varying levels of depression. Additionally, we will evaluate the model's performance using robust metrics, ensuring its reliability and effectiveness. The project aims not only to contribute to academic research on mental health and technology but also to provide insights that could inform future interventions and support systems for individuals experiencing depression. Ultimately, this work seeks to highlight the potential of social media analysis in understanding and addressing mental health issues.

2. LITERATURE SURVEY

1.Sentiment and Emotion Analysis of Social Media Posts

Several studies have focused on sentiment and emotion analysis to understand user expressions on social media platforms. Works by Cambria et al. (2017) explore emotion-rich sentiment analysis using deep learning, offering insights into distinguishing depressive language patterns from neutral or positive expressions. Researchers have highlighted that changes in linguistic tone, use of negative words, and increased sentiment variability may be strong indicators of depressive states.

2.Deep Learning Techniques for Depression Detection

Orabi et al. (2018) conducted studies using convolutional neural networks (CNN) and recurrent neural networks (RNN), such as LSTM, to analyze the text data from Twitter posts. These models were effective in identifying depression cues through context-specific word embeddings. Advanced attention mechanisms have further improved interpretability, helping researchers to focus on critical words or phrases indicating depression.

3.Natural Language Processing (NLP) for Mental Health Monitoring

Studies by Althoff et al. (2016) utilized NLP techniques to extract features like word count, sentence structure, and semantic meaning from Reddit posts. Their approach combined NLP with machine learning classifiers to detect early signs of depression, showing the potential of combining text analysis with predictive modeling for public health purposes.

4.Multimodal Analysis for Depression Detection

While text is a primary source for analysis, recent studies have expanded the scope to include other data modalities. Rizzo et al. (2020) investigated combining visual data (e.g., images shared by users) with textual posts. This multimodal approach employs deep neural networks to enhance accuracy, providing a richer context for understanding social media behavior and its relation to mental health states.

5.Challenges and Ethical Considerations

Research by Guntuku et al. (2019) has highlighted critical challenges in using social media data for mental health estimation, such as privacy concerns, data authenticity, and ethical issues. It emphasizes the need for ethical frameworks and informed consent when using personal data for research and AI applications.

6.Applications of Social Media-Based Depression Estimation

Various applications have emerged for using social media data for depression estimation, including early warning systems and automated mental health support. For example, Chancellor et al. (2019) discussed deploying AI-driven chatbots capable of engaging users in real-time, using depression detection insights to recommend professional support or immediate interventions.

3. PROBLEM STATEMENT

The increasing prevalence of depression poses a significant public health challenge, yet many individuals do not seek professional help or may go undiagnosed. Social media platforms serve as a unique lens into the emotional states of users, often revealing their struggles with mental health. However, the sheer volume of data generated on these platforms makes it difficult to manually analyze and identify patterns indicative of depression. This project seeks to address the gap in understanding and quantifying depression intensity by leveraging deep learning techniques to analyze social media content. The primary problem lies in developing an accurate and reliable model that can classify and quantify levels of depression based on user- generated text, ultimately enabling more timely and effective interventions for those in need.

4. METHODOLOGY

The methodology for estimating depression intensity via social media using a deep learning approach involves a multi-stage process. First, data is collected from social media platforms such as Twitter, Reddit, or Facebook, using appropriate web scraping tools, while adhering to ethical guidelines and privacy considerations. The collected data undergoes preprocessing, including tokenization, noise removal, handling of special characters, and normalization of emoticons and emojis. Key features, such as linguistic patterns, sentiment scores, and topics related to mental health, are extracted using tools like sentiment analysis and topic modeling. Text data is further transformed into embeddings using methods such as Word2Vec, GloVe, or transformer-based embeddings like BERT, ensuring contextual language representation.Deep learning models are then employed for analysis, starting with baseline classifiers such as Support Vector Machines (SVM) and then advancing to more complex architectures, including LSTM networks, transformers (e.g., BERT), and convolutional neural networks (CNNs). For scenarios where multimedia data (e.g., images) is analyzed alongside text, multimodal models are explored. During model training, data is divided into training, validation, and test sets, and hyperparameters are optimized using techniques like grid search. Regularization methods, such as dropout, are applied to prevent overfitting and enhance model generalization.

Evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, are used to assess model performance. Cross-validation ensures the robustness and reliability of the models. To enhance interpretability, attention mechanisms are incorporated, and tools like LIME or SHAP provide insights into model predictions. Finally, a prototype system is developed to analyze social media data in real-time, incorporating user feedback and adhering to ethical and privacy considerations for real-world deployment. This comprehensive methodology aims to create a scalable, effective solution for depression detection using deep learning and social media data analysis.

5. Conclusion

The present research article outlines the suggested approach for identifying Depression among students. The student's attributes are inputted via an interactive survey that features inquiries pertaining to Depression-inducing factors or indicators of Depression. The data of one hundred students has been efficiently collected and recorded in an Excel spreadsheet. The system is provided with a dataset pertaining to Depression status as its input. The dataset undergoes pre-processing procedures prior to its utilization in the framework. The dataset that has undergone pre-processing is subsequently utilized as an input source for the K-nearest neighbor clustering methodology to cluster the data. Upon the formation of the clusters, they are subsequently passed on to the Pearson Correlation component to facilitate the evaluation of the correlation among the Depression parameters and the dataset. The correlation module's output is utilized as an input for the generation of neurons in Artificial Neural Networks. The Artificial Neural Network (ANN) module is capable of achieving efficient calculation, which is presented in the format of a list of probabilities. The Decision Tree component is utilized to classify the probability list and achieve accurate Depression prediction based on the input attributes..

References

1. Predicting Anxiety, Depression, and Depression in Modern Life using Machine Learning Algorithms. Procedia Computer Science. 167. 1258-1267. 10.1016/j.procs.2020.03.442.

2.Zainudin, Z. & Hasan, S. & Shamsuddin, S.M. & Argawal, S.. (2021). Depression Detection using Machine Learning and Deep Learning. Journal of Physics: Conference Series. 1997. 012019. 10.1088/1742-6596/1997/1/012019.

3.Hossain, Elias & Alazeb, Abdulwahab & Almudawi, Naif & Alshehri, Mohammed & Gazi, M & Faruque, Golam & Rahman, Md. (2022). Forecasting Mental Depression Using Machine Learning Algorithms. Computers, Materials, and Continua. 72. 4945-4966. 10.32604/cmc.2022.027058.

4.Althoff, T., Clark, K., & Leskovec, J. (2016). "Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health." Transactions of the Association for Computational Linguistics, 4, 463-476.

The study demonstrates how NLP techniques can be used to analyze user conversations on platforms like Reddit for mental health monitoring.

5.Rizzo, G., Pereira, A. G., & Vieira, J. (2020). "Analyzing Depression-Related Images and Text on Social Media." Journal of Computational Social Science, 3, 57-74.

This paper introduces a multimodal approach, combining textual and visual cues, to better understand social media behaviors related to mental health.

6.Althoff, T., Clark, K., & Leskovec, J. (2016). "Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health." Transactions of the Association for Computational Linguistics, 4, 463-476.

7.Rizzo, G., Pereira, A. G., & Vieira, J. (2020). "Analyzing Depression-Related Images and Text on Social Media." Journal of Computational Social Science, 3, 57-74.

8.Chen, L., Lu, X., Yuan, J., & Luo, X. (2018). "A Survey of Depression Detection in Social Media Using Deep Learning Approaches." Journal of Artificial Intelligence Research, 59, 547-567.

9.Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R., Wang, H., & Ulhaq, A. (2018). "Depression Detection from Social Network Data Using Machine Learning Techniques." Health Information Science and Systems, 6, 8.

10.Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of NAACL-HLT.

11.Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space." arXiv preprint arXiv:1301.3781.

12.Lim, S. M., Lee, S., & Lee, J. H. (2020). "Detecting Depression Using a Machine Learning Approach and Social Media Posts." International Journal of Environmental Research and Public Health, 17(17), 6207.

13.Jiang, K., Li, Y., & Chang, M. (2018). "Detecting Social Media Depression Using Textual Data Analysis." Computational Intelligence and Neuroscience, 2018, Article ID 9521394.

14.Wang, W. Y., & Jurgens, D. (2018). "Detecting and Characterizing User Depression on Social Media: A Content-Based Approach." Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-28.

15.Huang, J., Luo, Y., & Lin, Z. (2017). "Sentiment Dynamics in Social Media and Its Implications for Mental Health." Computational Social Science, 3, 214-227.