

Mental Health Analysis Through Social Media Using Machine Learning and Deep Learning Approaches

CH . Komala Devi Mahalakshmi¹, P.V.N Pavan²,

¹Final year MCA Student, Department of Computer Applications, GIET Engineering College, Andhra Pradesh, India

² Assistance Professor, GIET Engineering College, Andhra Pradesh, India

Abstract

Over the past few years, many people have been suffering from psychological problems like depression, anxiety, and stress due to many factors that exist socially and personally. People tend to express their feelings and emotions on social media channels. Hence, social media is an excellent resource when it comes to getting insights about mental issues and their symptoms. This paper aims to study the use of artificial intelligence in social media to predict whether a person suffers from mental disorders or not using different machine learning and deep learning approaches. This system employs machine learning approaches, such as SVM (Support Vector Machine), Logistic Regression, and Random Forest classifiers, to carry out different tasks. Moreover, a complex ensemble classifier called XGBoost Classifier is employed to make predictions more efficient and accurate. When it comes to deep learning, approaches such as LSTM (Long Short-Term Memory) and Bidirectional LSTM are used to understand better the sequence of input data in text classification tasks. In order to increase the efficiency of models, text data needs to be processed through different steps such as tokenizing, stop-word removal, and stemming. Text features are extracted using TF-IDF vectorization technique in machine learning models.

1.INTRODUCTION

Introduction Mental health is a very important issue nowadays due to many reasons like increased stress, changes in people's lifestyle, and social pressure. People suffer from different types of mental diseases that range from depression and anxiety to stress, which affects millions of people across the world. The period when social media becomes very developed implies the necessity to write comments and messages in order to share the ideas and opinions. Social media pages contain plenty of data that might be used for analyzing people's mental condition. Traditionally, a mental health analysis process is done by conducting interviews with individuals. It is concluded from the answers of individuals what their mental health status is. Unfortunately, such traditional methods do not work effectively and sometimes give wrong results. That is why it is important to develop an automated tool. **Description of the Problem** This project suggests the development of a computer program for analyzing social media and classifying texts in terms of different mental states. Preprocessing stages include text cleaning,

tokenization, stopword removal, and stemming. **Introduction** The issue of mental well-being is currently of high significance since there are various factors affecting people's condition such as additional stress, changes in lifestyles, and societal pressure. The types of mental illnesses vary from anxiety, depression, to other kinds of mental diseases such as stress, which affects numerous individuals all around the globe. When the era of developed social media becomes prominent, it is important to use one's opportunities for writing posts and commenting to voice ideas and opinions. There is much information on social media sites that can be analyzed regarding individual's mental state. The conventional method for doing so implies interviewing people and concluding their condition based on the results. However, this method cannot be considered efficient since it provides false information from time to time. That is why it is vital to create an automated program. **Description of the Problem** This research focuses on developing a program for processing the information about mental well-being of individuals using social media. Cleaning, tokenization, stopwords removal, and stemming are among preprocessing procedures.

2. Methodology

2.1 Module of User Authentication:

This module takes care of the process of user registration, login, and logout functions in the system. It provides restricted access to the system based on the validity of the user. User credentials including usernames and passwords are kept confidential in the database. Verification of the user is done by authenticating them during login. It helps in keeping sessions active throughout the system. This helps in making the system more secure and private.

2.2 Module of Text Input and Preprocessing:

This module is required for collecting the input text from the users as well as preprocessing the text collected before feeding it to machine learning models for prediction. Some of the activities done in this process include removing the special characters, conversion of uppercase letters to lower case, removal of stopwords, and stemming of the words to reduce them to their basic forms

2.3 Feature Extraction Module (TF-IDF):

The Feature Extraction Module (TF-IDF) transforms the processed texts into numeric vectors using Term Frequency-Inverse Document Frequency. Words will be accorded weight depending on their frequency and importance in the data set. Numeric vectors produced by TF-IDF are used as the input for the machine learning model. This module helps in measuring the importance of words within the document. This increases the capacity of the model to classify the documents. It is important in creating effective text representations for classification purposes.

2.4 Machine Learning Model Training Module:

This module includes training of several machine learning models with the help of the features obtained above. Machine learning algorithms used include Logistic Regression, Support Vector Machine, Random Forest, Naive Bayes Classifier, and XGBoost. The models learn from the features provided to create predictions. They are then evaluated through measures such as accuracy, precision, and recall. The best performing model will be chosen according to the result of evaluation. This module is crucial for the construction of an efficient model for prediction.

2.5 Prediction and Classification Module:

This module processes input by making predictions regarding the user's mental health state by leveraging the pre-trained machine learning algorithm. Input text will be processed into TF-IDF features and sent as input to the pre-trained model. The algorithm then categorizes the given text in conditions such as depression, anxiety, stress, or even in normal cases. This module provides instantaneous results. This is the decision-making element of the entire system. This module engages the final users.

2.6 Result Visualization and Analysis Module:

This module is responsible for visualizing the output predictions in an understandable way. Graphical representations like bar charts can be provided to visualize the probabilistic outputs of the various models used. The system will also show description and treatment for each condition predicted. In addition to visualizing the output predictions in a graphical and easy-to-understand format, the module will also generate performance graphs for all the models used.

Results:

Fig-1: HOME PAGE



Fig-2 SIGNUP PAGE : OPTIONAL

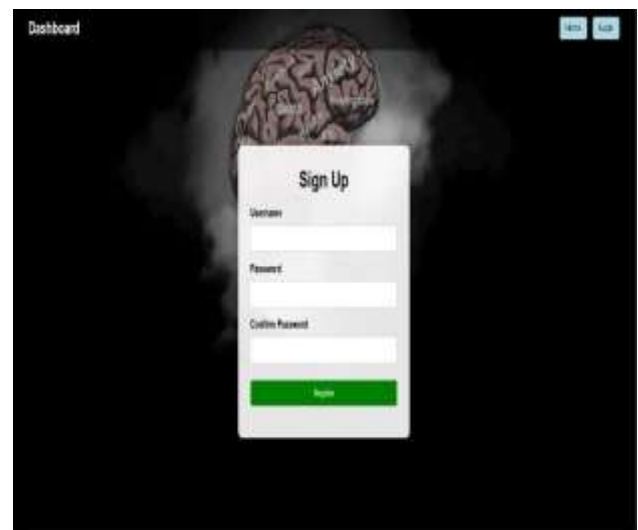


Fig-2.1 LOGIN PAGE : OPTIONAL

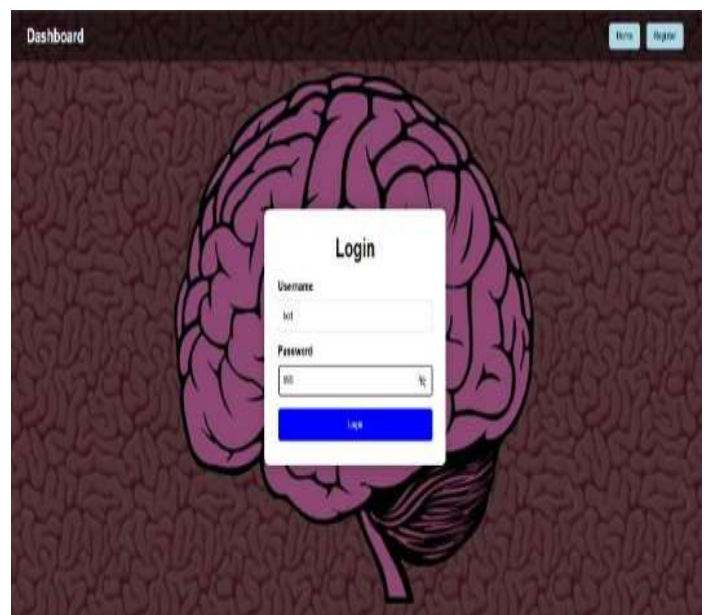


Fig-3: TEXT PREDICTION



Fig-4: PREDICTION RESULT

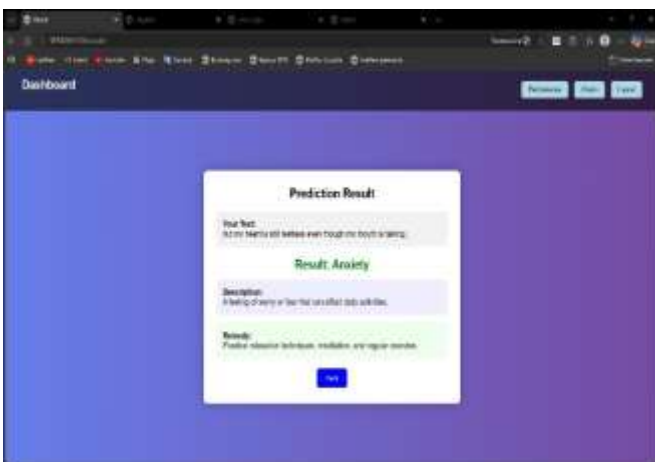


Fig-5: PREDICTION CHART



3. DISCUSSION

"Mental Health Analysis Through Social Media Using Machine Learning and Deep Learning Approaches" is a project that aims to analyze the mental condition of a

person based on social media posts and data. It is essential to emphasize the importance of social media as a channel where people express their thoughts and feelings.

Social media text data is analyzed through NLP operations, including preprocessing of posts and extracting features. Several Machine Learning algorithms, such as SVM, Random Forest, Naïve Bayes, and Logistic Regression, are compared with Deep Learning methods such as Long Short-Term Memory (LSTM). The experimental results indicate that automatic analysis helps to detect various symptoms of mental illness.

It should be stated that the deep learning approach outperforms other machine learning algorithms since it allows for recognizing the context and sequence of words in text. Meanwhile, the latter option is much more productive in terms of time and performance.

There are multiple benefits of utilizing automated analysis systems. These include fast identification of mental problems, the absence of human labor, scalability, and assistance for healthcare professionals. Challenges related to sarcasm, slang, and multilinguality may also occur.

4. CONCLUSIONS

Finally, the proposed system effectively shows how machine learning and deep learning can be used for recognizing mental health problems using social media data. The system uses different preprocessing methods to identify mental health cases, and then applies TF-IDF and classification algorithms like XGBoost and LSTM to achieve accurate predictions. The use of a user-friendly interface based on Django is another feature that adds value to the system. In addition to providing predictions, the system gives useful descriptions and solutions related to the problem. The visualization of results with charts is an important step in this process too. Thus, this system plays an important role in the early recognition of mental health problems and provides the basis for future research "Mental Health Analysis Through Social Media Using Machine Learning and Deep Learning Approaches" is a project that aims to analyze the mental condition of a person based on social media posts and data. It is essential to emphasize the importance of social media as a channel where people express their thoughts and feelings. Social media text data is analyzed through NLP operations, including preprocessing of posts and extracting features. Several Machine Learning algorithms, such as SVM, Random Forest, Naïve Bayes, and Logistic Regression, are compared with Deep Learning methods such as Long Short-Term Memory (LSTM). The experimental results indicate that automatic analysis helps to detect various symptoms of mental illness. It should be stated that the deep learning approach outperforms other machine learning algorithms since it allows for recognizing the context and sequence of words in text. Meanwhile, the latter option is much more productive in terms of time and performance. There are multiple benefits of utilizing automated analysis systems.

These include fast identification of mental problems, the absence of human labor, scalability, and assistance for healthcare professionals. Challenges related to sarcasm, slang, and multilinguality may also occur.

5. REFERENCES

- [1] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in Proc. 1st Instructional Conf. Machine Learning, 2003, pp. 133–142.
- [2] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," in Proc. ICLR, 2013.
- [3] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014, pp. 1746–1751.
- [4] A. Vaswani et al., "Attention Is All You Need," in Proc. NIPS, 2017, pp. 5998–6008.
- [5] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL, 2019, pp. 4171–4186.
- [6] F. Chollet, "Deep Learning with Python," Manning Publications, 2017.
- [7] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [8] C. Cortes and V. Vapnik, "Support Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
- [9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. KDD, 2016, pp. 785–794.
- [10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
- [11] I. Goodfellow et al., "Deep Learning," MIT Press, 2016.
- [12] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," MIT Press, 2012.
- [13] D. Jurafsky and J. H. Martin, "Speech and Language Processing," Pearson, 2019.
- [14] S. Bird, E. Klein, and E. Loper, "Natural Language Processing with Python," O'Reilly, 2009.
- [15] J. Pennington et al., "GloVe: Global Vectors for Word Representation," in Proc. EMNLP, 2014, pp. 1532–1543.
- [16] R. Socher et al., "Recursive Deep Models for Semantic Compositionality," in Proc. EMNLP, 2013.
- [17] H. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool, 2012.
- [18] M. Zhang and Y. LeCun, "Text Understanding from Scratch," arXiv, 2015.
- [19] J. Brownlee, "Machine Learning Mastery with Python," Machine Learning Mastery, 2017.

- [20] P. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” JMLR, vol. 12, pp. 2825–2830, 2011.
- [21] W. McKinney, “Data Structures for Statistical Computing in Python,” in Proc. SciPy, 2010.
- [22] H. Wickham, “ggplot2: Elegant Graphics for Data Analysis,” Springer, 2016.
- [23] S. Raschka, “Python Machine Learning,” Packt Publishing, 2015.
- [24] A. Géron, “Hands-On Machine Learning with Scikit-Learn & TensorFlow,” O’Reilly, 2019.
- [25] R. Feldman and J. Sanger, “The Text Mining Handbook,” Cambridge Univ. Press, 2007.
- [26] E. Cambria et al., “Affective Computing and Sentiment Analysis,” IEEE Intelligent Systems, 2013.
- [27] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in IR, 2008.
- [28] J. Howard and S. Ruder, “Universal Language Model Fine-tuning,” in Proc. ACL, 2018.
- [29] T. Young et al., “Recent Trends in Deep Learning for NLP,” IEEE Computational Intelligence Magazine, 2018.
- [30] M. Peters et al., “Deep Contextualized Word Representations,” in Proc. NAACL, 2018.