IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Text-Based Stress Detection And Classification Using Machine Learning

Harika Juluri[1] , Deepender Singh[2],T R Yograj Singh[3],G Sai Kiran Yadav[4]

[1]Department of Computer Science and Engineering,

[1]Methodist College of Engineering and Technology, Hyderabad, Telangana, India

Abstract:As mental health awareness is increasing, there is a growing need for tools to help detect and manage stress levels in individuals. It can be challenging for individuals to recognize and categorize their stress levels accurately, leading to insufficient or inappropriate coping strategies. The aim of this work is to develop a text-based stress detection and categorization model that utilizes natural language processing techniques to analyze text messages and detect stress levels. This wok involves several steps, including dataset preprocessing, feature extraction, training, stress detection, stress categorization and recommendation generation. The model will be trained using advanced machine learning techniques and evaluated using various performance metrics such as accuracy, precision, recall and F1 score. The results of this project will enable users to manage their stress effectively.

Key words **–** Machine learning, Random Forest, XGboost, performance metrics

1.INTRODUCTION

1.1 Understanding Stress

Stress is ubiquitous and serious condition that exerts profound effects on both mental and physical health. It arises when body responds to any change requiring adjustment or reaction, which can stem from external factors like environmental pressures or internal factors such as personal thoughts and emotions. The Pervasive ness of stress can lead to cascade of emotional, physical, and behavioral isssues, significantly disrupting and individual's daily life and overall well-being. The biological response to stress involves the activation of the hypothalamic-pituitary-adrenal (HPA) axis, leading to the release of cortisol and other stress hormones, which prepare the body for a fight-or-flight response. However, chronic activation of this response lead to detrimental health outcomes, including cardiovascular disease, weakened immune function and mental health disorders like anxiety and depression.

1.2 Symptoms and Impact

Stress manifests in various symptoms that can significantly impact a person's mental and physical health. Emotional symptoms include irritability, moodiness and feeling overwhelmed, while physical symptoms can range from low energy and headaches to rapid heartbeat and chest pain. Behavioral changes such as changes in appetite, procrastination and increased use of substances like alcohol or drugs are also common. These symptoms highlight the extensive reach fo stress throughout the body and underscore the need for effective detection and management strategies.

1.3 Addressing stress through technology

The development of tools and strategies to detect and address stress is an ongoing area of research and innovation, aiming to provide individuals with the resources they need to maintain mental and physical health. Text-based stress detection and classification using machine learning represents a promising approach in this field. By analyzing linguistic patterns and emotional cues in written text, ML Algorithms can identify indicators of stress, enabling timely interventions and support. This project

explores the potential of ML in stress detection, contributing to the broader efforts to improve mental health outcomes through technological advancements.

2. Literature Survey

2.1 Paper Title: "A Novel approach for text classification by combinig pre-trained BERT model with cnn classifier"

Authors: Chenxu Wang, Yulin Li, Ziying Wang.

Key findings:

- The proposed model which combines a pre-trained BERT model with a CNN classifier, achieves the highest performance on the AG News and Amazon product reviews datasets compared to baseline models.
- The model's superior performance is attributed to its ability to leverage the contextualized embeddings of BERT and the n-gram feature capturing of CNN.

2.2 Pape Title: "Research on domain text classification on method based on BERT"

Authors: Shaohui Xie, Yangsen Zhang, Zhengyu Hou, Zhenjiang Su, Renjie Wang, Ziyuan He.

Key findings:

- The proposed BERT_VCA Model outperformed the BERT_BASE modle by 1.12% in F1 value for domain text classification.

2.3 Paper Title: "Stress detection using natural language processing and machine learning over social interactions:

Authors: Tanya Nijhawan, Girija Attigeri, T Anantha Krishna

Key findings:

- Random forest classifier achieved the highest accuracy of 97.78% among the machine learning models test for sentiment analysis.
- The BERT Deep learning model achieved an accuracy of 94% for emotion classification.

2.4 Paper Title: "Bert-Enchanced Text Graph Neural Network for Classification"

Authors: Yiping Yang, Xiaohui Cui

Key findings:

- BEGNN which combines BERT-based and GNN-based feature extraction with a co-attention module, outperforms other baseline models.
- BEGNN shows more obvious imporvements on datasets with longer text, indicating it can better process longer texts.

2.5 Paper Title: "How to fine-tuen BERT for Text Classification?"

Authors: Chi Sun, Xipeng Qui, Yige Xu, Xuanjing Huang.

Key findings:

- The top layer of BERT is the most useful for the text classification tasks.
- Using an appropriate layer-wise decreasing learning rate helps BERT overcome the catastrophic forgetting problem.

3. Proposed Work

3.1 Introduction

Briefly introduce the problem of stress among the people and the impact of stress on individual's health. Highlight the importance of early detection of stress and intervention. State the research objectives and the significance of the proposed work.

In today's world, many of us rely on social media platforms such as Facebook, X (formerly Twitter), Snapchat, YouTube, TikTok, and Instagram to find and connect with each other. While each has its benefits, it's important to remember that social media can never be a replacement for real-world human connection. Human beings are social creatures. We need the companionship of others to thrive in life, and the strength of our connections has a huge impact on our mental health and happiness. Being socially connected to others can ease stress, anxiety, and depression, boost self-worth, provide comfort and joy, prevent loneliness, and even add years to your life. On the flip side, lacking strong social connections can pose a serious risk to your mental and emotional health.

3.2 Literature Review

Summarize existing research on stress detection, including methodologies, datasets and outcomes. Identify the key findings and gaps in the current literature. Discuss the application of machine learning in the context of veterinary epidemiology.

3.2 Research objectives

Clearly define the main goals of the proposed research. For Example: To develop a machine learning model for detection of stress among individuals using text input of their messages or posts. To create a user-friendly interface for individuals to post their feelings and use the model.

Detection of stress in texts from social media by current systems is quite a challenge. This is because models rely on the sensory data along with the text messages to predict stress, but in all the scenarios the user may not have access to such IoT devices or sensors to check stress making the models less accessible and user friendly.

3.3 Methodology

Explain the data collection process, including sources and types of data. Describe the preprocessing steps, data cleaning and feature engineering. Discuss the selection of machine learning algorithms and techniques(e.g., Decision Trees, Support Vector Machine, Neural Networks). Explain the model evaluation and validation process.

3.4 Data collection and Preprocessing

Detail how data was collected, anonymized and processed. Address issues related to data quality, missing values and data imbalance. Discuss the ethical consideration reagarding data usage and privacy.

Word2Vec is a powerful technique for natural language processing that captures semantic relationships between words. Developed by researchers at Google in 2013, Word2Vec transforms text into a numerical form that algorithms can easily process. This model creates word embeddings, which are dense vectors representing words in a continuous vector space, where words with similar meanings are positioned close to each other. The fundamental concept behind Word2Vec is to use the context of a word within a sentence to predict the word itself or vice versa.

BERT (Bidirectional Encoder Representations from Transformers) tokenizer is an advanced natural language processing tool designed to pre-process text data for the BERT model, a state-of-the-art transformer-based architecture developed by Google in 2018. The BERT tokenizer breaks down text into tokens, which are the basic units of meaning in the model. Unlike traditional tokenizers that often split text into words, the BERT tokenizer employs a WordPiece algorithm that breaks words into subword units. This allows BERT to handle out-of-vocabulary words effectively and understand context more precisely

by considering the subwords that make up a word. For instance, the word "unhappiness" might be tokenized into ["un", "##happiness"], allowing BERT to understand its components and contextual meaning more accurately.

3.5 Model Development and Training

Explain the development of machine learning models for the detection and classification of stress. Describe the algorithm selection, cross-validation strategies. Discuss the choice of evaluation metrics and model interpretation techniques.

The Random Forest Classifier is a robust and versatile machine learning algorithm renowned for its exceptional performance in classification tasks. It operates by creating an ensemble of decision trees, where each tree is trained on a random subset of the data and features, ensuring diversity and reducing overfitting. This ensemble approach aggregates the predictions of multiple decision trees, resulting in a highly accurate and stable model.trees, resulting in a highly accurate and stable model. Random Forest Classifier is particularly effective in handling complex datasets with both numerical and categorical features, making it suitable for a wide range of applications. Its ability to assess feature importance provides valuable insights into the underlying data patterns, aiding in feature selection and model interpretation. Additionally, Random Forest Classifier is scalable and computationally efficient, capable of handling large datasets with high dimensionality. Its robustness to noisy data and flexibility in handling various data types make it a popular choice among data scientists and machine learning practitioners for building reliable and accurate classification models.

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm known for its exceptional performance in a wide range of tasks, including regression, classification, and ranking. It belongs to the family of boosting algorithms, which combine weak learners (typically decision trees) to create a strong predictive model. XGBoost has gained popularity due to its scalability, flexibility, and ability to handle complex datasets. The key idea behind XGBoost is to iteratively build an ensemble of weak prediction models, called decision trees, to improve the overall predictive performance. The algorithm consists of three main components: a loss function, a weak learner, and a boosting framework. It is a powerful machine learning algorithm that combines the predictions of multiple weak learners to create a strong predictive model

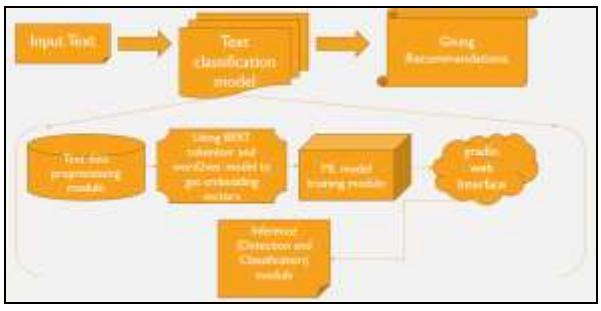
3.6 User Interface Development

If your research includes creating a user-friendly tool outline the development process. Mention the technologies and frameworks you plan to use for the interface

3.7 Results

Present the expected outcomes, such as the accuracy of the stress detection and classification model respectively. Model performance, the usability of the developed tool.

4 System Architecture



5. Conclusion

Text-based stress detection using machine learning holds immense promise for transforming the way we manage mental health and welfare. By harnessing the power of data, artificial intelligence and predictive modelling, we can work towards a future where stress detection, prevention an effective management become the norm in various contexts. This innovative approach offers the potential to improve stress-management outcome.

6. References

[1]Shaohui Xie, Yangsen Zhang, Zhengyu Hou, Zhenjiang Su, Renjie Wang, Ziyuan He – "Research on domain text classification method based on BERT"

[2]Chenxu Wang, Yulin Li, Ziying Wang –"A Novel Approach for Text Classification by Combining Pre trained BERT Model with CNN Classifier"

[3] Tanmay Gupta, Samarth Singh, Mr Arun Kumar – " Community Conversation Analyser"

[4] Ravishankara K, Dhanush, Vaishak, Srajan – "whatsapp chat Analyser".

[5] Weiguo Fan, Michael D Gordan – "The Power Of Social Media Analytics"

[6] M Ikonomakis, S Kotsaintis, V Tampakasm – "Text classification using machine learning techniques"

[7] Kern M L, Park G, Echstaedt J C, Schwartz H A, Sap M, Smith L K, Ungar L H – "Gaining insights from social media language: Methodologies and Challenges".

[8] Sergio Munoz, Carlos A Iglesias – "Information Processing and Management".

[9] Andrian Benton, Margaret Mitchell, Dirk Hovy – "Multi-task learning for mental health using social media texts".

[10] Joseph Lilleberg, Yun Zhu, Yanqing Zang – "support vector machine and wordevec for text classification with semantic features".

[11] M Nii, Yuya Tuchida, T Iwamoto, A Uchinuno, R Sakasita – "Nursing-care text evaluation using word vector representations realized by word2vec"

[12] Yonghui Zhang, Jingang Liu – "Microblogging Short Text Classification Based on Word2Vec"

[13] Lei Zhu Guijun Wang, X. Zou – "A Study of Chinese Document Representation and Classification with Word2vec"

[14] Md. Rafqul Islam, Ashir Ahmed,Abu Raihan M. Kamal,HuaWang,Anwaar Ulhaq - "Depression detection from social network data using machine learning techniques"

[15] Jeremy Howard, Sebastian Ruder – "Universal Language Model Fine-tuning for Text Classification".

[16] Chi sun, Xipeng Qiu, Yige Xu, Xuanjing Huang – "How to Fine-Tune BERT for Text Classification?".