Journal of Management & Entrepreneurship
ISSN 2229-5348
UGC Care Group I Journal
Vol-13 Issue-01 Mar 2024

DETECTION OF STRESS BASED ON SOCIAL MEDIA CONVERSATIONS

Mrs. Lanka DhanaLakshmi , Yarra Govardhana Naga Deepthi, Shaik Rabbani,
Vadavalli Hasri Mani Sandhya Harshitha, Vemuri Nivas
Assistant Professor, Department of CSE,
Seshadri Rao Gudlavalleru Engineering College,
Gudlavalleru, Krishna District, Andhra Pradesh.

ABSTRACT

Stress detection is a critical aspect of mental health monitoring, and the advent of machine learning provides an avenue to automate this process. This project aims to develop a stress detection model using machine learning techniques, utilizing a combination of structured questionnaire responses and unstructured Twitter data. The structured data consists of responses to a set of questions widely recognized as indicators of stress. These questions cover a spectrum of emotional and behavioral facets associated with stress, ranging from mood disturbances to impacts on daily life. The responses to these questions serve as labeled instances for training the machine learning model. The unstructured data, sourced from Twitter, comprises a collection of tweets, some of which may express stress-related sentiments. These tweets are treated as potential indicators of stress and are labeled accordingly. Preprocessing the data, extracting features, and training a machine learning model are all part of the process.Text data is transformed into numerical features using techniques such as TF-IDF. A suitable classification model, such as a Support Vector Machine, is trained on the structured questionnaire responses. This model is then applied to the unstructured Twitter data for stress prediction. The model's performance is assessed by the use of metrics such as F1 score, accuracy, precision, and recall. The successful development of this stress detection model could provide a valuable tool for automated mental health monitoring, offering insights into stress prevalence within a given population. The project not only contributes to the field of mental health technology but also highlights the potential of combining structured and unstructured data sources for enhanced predictive modeling. As stress continues to be a prevalent concern in modern society, the development of effective stress detection tools holds promise for improved well-being and early intervention. To enhance the project's scope, we introduce a section addressing remedies to overcome stress.

1. INTRODUCTION

Stress is an emotional and physical reaction that occurs when an individual perceives a threat, challenge, or demand that exceeds their ability to cope. It is a natural part of life and can manifest in various situations, both positive and negative. In the context of stress, machine learning (ML) can be extremely important in a number of ways: ML models can analyze various data sources (e.g., questionnaire responses, social media posts, speech patterns) to identify potential signs of stress. Early detection can lead to timely intervention and support.ML algorithms can analyze a person's individual characteristics, preferences, and responses to different treatments to tailor a treatment plan that is more likely to be effective. This can be used to monitor changes in mood and provide timelysupport. Over 70% of population experience stress, a condition linked to a compromised immune system, cancer, cardiovascular disease, depression, diabetes, and substance addiction. The profound impact of stress on both physiological health and psychological well-being underscores the critical need for robust and swift methods to detect human stress. Such technologies could facilitate continuous stress monitoring, empowering individuals to adapt their daily activities for stress reduction, while healthcare professionals could deliver more targeted treatment for stress-related conditions. Various mechanisms for detecting human stress have emerged in research.The use of physiological signals for stress detection.Typically, these investigations used conventional machine learning algorithms like K-nearest neighbors, decision trees,support vector machines, random forests, and linear discriminant analysis (LDA) to analyze physiological signals. Notably, Healey and Picard conducted pioneering research, using signals from multiple sensors for binary classification of stressed and non-stressed conditions. The manual creation of hand-crafted features, a prerequisite found in nearly all prior research, is a major drawback of traditional machine learning approaches. It was common practice to use physiological signal characteristics and statistics as features, which made it difficult to generate features for each sensor's signals.Furthermore, it is unclear if these features encompass the whole feature space required by machine learning algorithms because their representativeness and accuracy have not been demonstrated.

2. LITERATU REREVIEW

Bobade and Vani's study, "Stress Identification Using Multimodal Physiological Data and MachineLearning and Deep Learning,"[1] addresses the pervasive issue of stress by leveraging wearable sensors to capture multimodal physiological data. Stress, a common aspect of daily life, can lead to severe health[2] problems if not detected early. The study proposes the use of bio-signals, including thermal, electrical, impedance, acoustic, and optical, obtained from sensors [3]like three-axis acceleration, electrocardiogram, and others from the WESAD dataset[4] Results show promising accuracies [5]of up to 81.65% and 93.20% for three-class and binary classifications using machine learning and even higher, up to 84.32% and 95.21%, with deep learning. This underscores the potential of advanced computational methods, especially deep learning, in enhancing stress detection accuracy, with significant [6]implications for preventive healthcare.

Bannore et al.'s study on "Mental Stress Detection using Machine Learning Algorithm" addresses the challenging task of measuring and understanding mental stress[7] by introducing an appearance-based facial expression recognition system. Stress, a subjective phenomenon with diverse impacts on health, is explored[8] through the lens of facial expressions, serving as valuable indicators of emotional states. The study employs Convolutional Neural Network (CNN) and Local Binary Pattern (LBP) for appearance feature extraction, categorizing facial expressions[9] into four basic emotions (anger, fear, unhappy, and non-stressed). This non-invasive approach is experimentally validated on both Indian and Cohn-Kanade databases, contributing to the literature on stress detection by providing a nuanced and culturally [10]diverse perspective. The combination of machine learning algorithms and facial expression analysis demonstrates promising potential for objective and efficient stress detection.

In paper, "Detection of Stress in IT Employees using Machine Learning Technique," S. K. Kanaparthi et al. Tackle the crucial issue of identifying stress in IT professionals by employing machine learning and visual processing. The study represents an advancement over previous stress detection systems, incorporating real-time monitoring and individual counseling for a more comprehensive approach to employee well-being. By addressing the limitations of older systems, the research aligns with the evolving landscape of workplace mental health initiatives, emphasizing the need for adaptive and individualized approaches. The integration of a survey to collect data on employees' mental stress levels reflects a holistic methodology, considering both objective and subjective indicators of stress.

3. METHODOLOGY

Curating a diverse dataset comprising questionnaire responses from individuals with confirmed stress diagnoses and those without, involves careful consideration for coding category variables, managing missing values, and normalizing numerical aspects. Leveraging tools like NLTK for natural language processing tasks, we aim to explore various classification algorithms to identify the most suitable one for stress diagnosis. To ascertain the best performance in the context of stress diagnosis, a thorough comparison of a few chosen classification models will be carried out. We will measure important metrics like accuracy, precision, recall, F1-score, and AUC for classification performance evaluation, and we will thoroughly evaluate the model's efficacy using cross-validation techniques and a dedicated validation set.To enhance user accessibility, we will implement a user interface using Flask, providing an intuitive platform for individuals seeking mental health assessments. Users will be able to input their questionnaire responses, which will then undergo analysis by the integrated machine learning model. The system will furnish users with diagnostic feedback based on the model's evaluation of their responses, accompanied by personalized recommendations for overcoming stress.

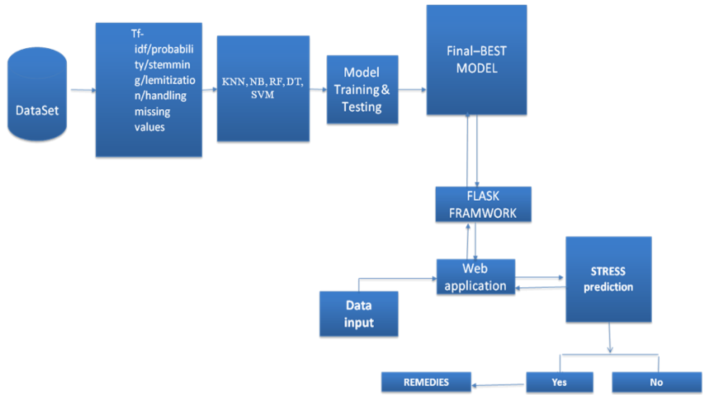


Fig3.1: Proposed system

MODULES

- Data Collection
- Data Preprocessing
- Model Training
- Model Testing
- Flask Framework
- Suggestion & Notification

A. Data collection

QUESTIONS

- " Minimal enthusiasm or enjoyment in one's activities ",
- " Feeling unhappy, melancholy, or forlorn ",
- " Problems getting to sleep, remaining asleep, or sleeping excessively ",
- " Having low energy or feeling exhausted ",
- " Insufficient appetite or overindulging in food ",

- " Having negative self-talk, feeling inadequate, or believing that you have let your family or yourself down “,
- " Difficulty focusing on tasks, like watching television or reading a newspaper ",
- " Speaking or moving so slowly that others could have observed ",
- " Feelings of injuring yourself or that it would be better if you died ",
- " If you have had any of the aforementioned challenges, how challenging have they made your life at home, at work, in school, or around other people?"

B. Data Pre-processing

In the context of stress detection, these text analysis and preprocessing techniques play a pivotal role in enhancing the accuracy and effectiveness of machine learning models. TF-IDF is particularly valuable for evaluating the relevance of words in stress-related documents, helping to identify key terms that may signify stress indicators. Probability assessment aids in predicting the likelihood of specific stress-related events or sentiments occurring in the analyzed text. Stemming and lemmatization are crucial for normalizing and simplifying stress-related language, allowing the model to recognize variations of words and capture their core meanings. This is especially important in stress detection, where subtle changes in language may convey different levels of distress. Additionally, effective handling of missing values is essential when dealing with responses to stress-related questionnaires or text data. Proper management of missing information ensures that the stress detection model is trained on complete and representative data, preventing potential biases in the analysis. In the realm of stress detection, integrating these text analysis techniques into the data preprocessing pipeline contributes to a more robust and reliable model. The enhanced understanding of stress-related language nuances, accurate representation of key terms, and careful management of missing information collectively empower the machine learning model to better identify and diagnose stress patterns in textual data.

C. Model Training

Training a machine learning model for stress detection involves several key steps, including data preparation, feature extraction, algorithm selection, model training, and evaluation. Let's explore each step in the context of stress detection, and provide an explanation for each of the mentioned algorithms: K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), and Support Vector Machine (SVM).

- K-Nearest Neighbors (KNN):

KNN classifies instances based on the majority class of their k-nearest neighbors. In the context of stress detection, KNN could be effective if similar stress patterns are close in the feature space.

- Random Forest (RF):

RF constructs several decision trees and aggregates the forecasts from them. RF is robust and can handle diverse datasets, making it suitable for stress detection where the relationship between features and stress may be complex.

- Decision Tree (DT):

A Decision Tree builds a structure recursively divided into sections according to features in the data.Decision Trees are interpretable and can be useful for understanding which features contribute most to stress detection.

- Naive Bayes (NB):

NB is predicated on the independence of features and is based on the Bayes theorem. It is particularly efficient for text data, making it relevant for stress detection involving questionnaire responses or other textual inputs.

- Support Vector Machine (SVM):

SVM finds a hyperplane that maximally separates different classes. SVM is effective in high-dimensional spaces, making it suitable for stress detection with diverse feature sets.

D. Model Testing

- **True Positive:** It is true that your prediction was positive.
- **True Negative:** It's true that you predicted the negative**.**
- **False Positive: (Type 1 Error):** Positive was your prediction, but it's not true.
- **False Negative: (Type 2 Error**): It's untrue that you predicted negatively**.**

Accuracy

Actually not a very good measure of performance, this is the most widely used statistic to evaluate a model. Inequality in courses leads to worse outcomes.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Precision

Percentage of positive cases relative to all positive cases that were projected.

$$\frac{TP}{TP + FP}$$

Recall/Sensitivity/True Positive Rate:

Percentage of successful cases relative to all successful cases in reality.

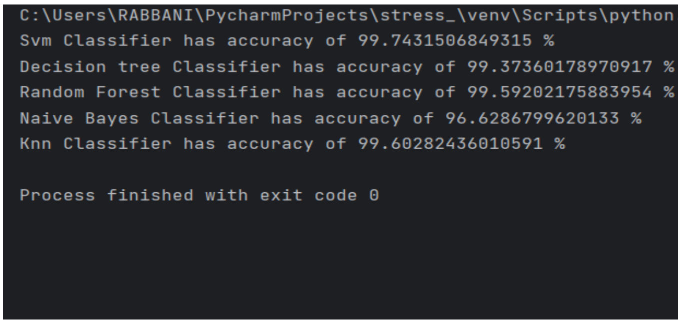
$$\frac{TP}{TP + FN}$$

F1 score

It is the precision and recall harmonic mean. The greater the F1 score, the better, as this takes into account both contributions.

$$\frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * precision * recall}{precision + recall}$$

The accuracy depends on two different genres like Quiz based and text based stress detector its accuracy with multiple algorithms are tested and displayed below:



```
C:\Users\RABBANI\PycharmProjects\stress_\venv\Scripts\python
Svm Classifier has accuracy of 99.7431506849315 %
Decision tree Classifier has accuracy of 99.37360178970917 %
Random Forest Classifier has accuracy of 99.59202175883954 %
Naive Bayes Classifier has accuracy of 96.6286799620133 %
Knn Classifier has accuracy of 99.60282436010591 %

Process finished with exit code 0
```

Fig 3.2: Accuracy percentages of different machine learning classifiers.

E. Flask framework Prediction

The front end of a website constitutes the user interface, encompassing visual elements such as text styles, colors, images, videos, graphs, tables, navigation menus, and interactive elements like buttons. In web development, HTML, CSS, and JavaScript are fundamental technologies employed to craft the front end of a site. Additionally, Flask, a Python web framework, is utilized for developing web applications, providing a seamless integration of Python with web technologies.The initial steps in creating a Flask web application involve importing the Flask class and instantiating it. The 'name' argument, representing the application's module or package, is passed to Flask, allowing it to locate resources like templates and static files. The route() decorator is then employed to specify the URLs that activate specific methods in the application. These methods define the content displayed in the user's browser, forming the foundation for dynamic and responsive web pages.

Flask operates on the Model-View-Controller (MVC) architecture, where the model manages data, the view handles the user interface, and the controller serves as an intermediary connecting the model and view. This architecture enhances code organization and maintainability in web development.

To leverage Flask for web application development, one must install it and define routes using the @app.route() decorator, which associates URL patterns with specific functionalities. This establishes the structure and navigation within the web application, allowing users to interact with different components seamlessly.

In essence, Flask provides a lightweight yet robust framework for building web applications rapidly, and its integration of Python simplifies the development process. By adopting Flask's MVC architecture and defining routes effectively, developers can create dynamic and engaging front-end experiences for users.

4.RESULTS

Fig4.1: User Interface for taking Stress Test

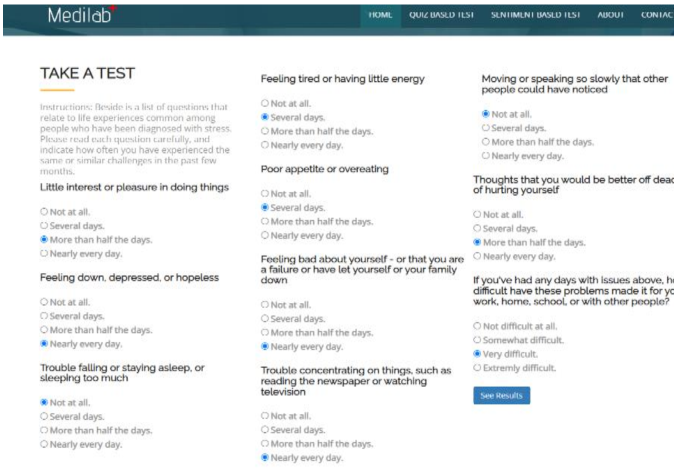


Fig 4.2: Demo Questions need to answer by user

Fig4.3: Based on the answers Stress can be classified

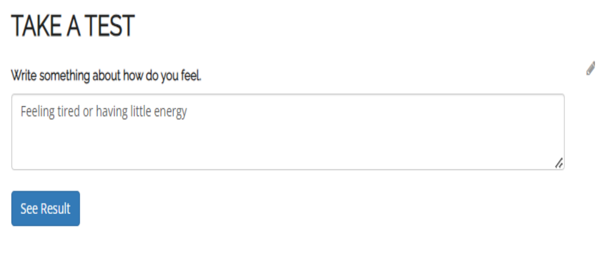


Fig4.4: Text based Stress Detection

Fig 4.5: Based on the Text Stress can be classified

5.CONCLUSION

The development of a stress detection model using machine learning techniques, integrating both structured questionnaire responses and unstructured Twitter data, presents a promising approach to automating the critical task of stress monitoring. By leveraging the power of advanced technologies, this project strives to contribute to the field of mental health by offering a comprehensive and proactive solution.The structured questionnaire responses, encompassing a broad spectrum of stress-related indicators, serve as a foundation for training the machine learning model. These responses provide a labeled dataset, enabling the model to learn and recognize patterns associated with stress across various emotional and behavioral dimensions.The incorporation of unstructured Twitter data introduces a dynamic and real-world element to the model. By treating tweets as potential indicators of stress, the project broadens its scope to capture the nuances of stress expression in diverse online contexts. This integration aims to enhance the model's adaptability and sensitivity to the evolving nature of stress-related communication.As we progress towards a more data-driven and technologically advanced era, the implications of this stress detection model extend beyond mere identification. The proactive nature of the model, coupled with its ability to process real-time social media data, opens avenues for timely intervention and mental health support. By combining structured and unstructured data sources, the model achieves a holistic understanding of stress, encompassing both individual responses and broader social expressions.

5.FUTURE SCOPE

The stress detection model using machine learning with questionnaire and Twitter data is vast and multidimensional. As the technology matures, we can anticipate the integration of more varied data sources such as voice inflection and facial expressions to refine the model's accuracy. The application of natural language processing could evolve to understand context and sentiment more deeply, enabling the model to distinguish between actual stress indicators and false positives with greater precision. There's also the possibility of personalizing stress management recommendations based on individual patterns detected over time. Furthermore, collaborations with healthcare providers could see the model becoming part of telemedicine platforms, offering users immediate feedback and professional support. In the grand scheme, the societal benefits include a reduction in healthcare costs due to preventative mental health care and a general increase in public awareness and understanding of stress and its impacts. This project could well be at the forefront of a shift towards more empathetic and responsive mental health technology.

7. REFERENCES

1. Brandão, Catarina. "P. Bazeley and K. Jackson, qualitative data analysis with Nvivo (2013). London: Sage." (2015): 492-494.

2. Jackson, Kristi, Pat Bazeley, and Patricia Bazeley. Qualitative data analysis with NVivo. Sage, 2019.

3. Kunz, Werner, and Sukanya Seshadri. "From virtual travelers to real friends: Relationship-building insights from an online travel community." Journal of business research 68, no. 9 (2015): 1822-1828.

4. Chung, Jin Young, and Dimitrios Buhalis. "Information needs in online social networks." Information Technology & Tourism 10.4 (2008): 267-281.

5. Espino, Mayel. "Method and system for brokering messages in a distributed system." U.S. Patent No. 8,849,892. 30 Sep. 2014.

6. Bhoi, Dhaval, and Amit Thakkar. "Sentiment analysis tools, process, methodologies: a survey." Int J Adv Sci Technol 29.4 (2021): 6280-6290.

7. Kumar, Yogesh, and Manish Mahajan. "Recent advancement of machine learning and deep learning in the field of healthcare system." Computational intelligence for machine learning and healthcare informatics 1 (2020): 77.

.8. Healy, Michael, et al. "A machine learning emotion detection platform to support affective well being." 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018.

9. Masood, Khalid, and Mohammed A. Alghamdi. "Modeling mental stress using a deep learning framework." IEEE access 7 (2019): 68446-68454.

10. Hyland, Michael E. "Do person variables exist in different ways?." American Psychologist 40.9 (1985): 1003.