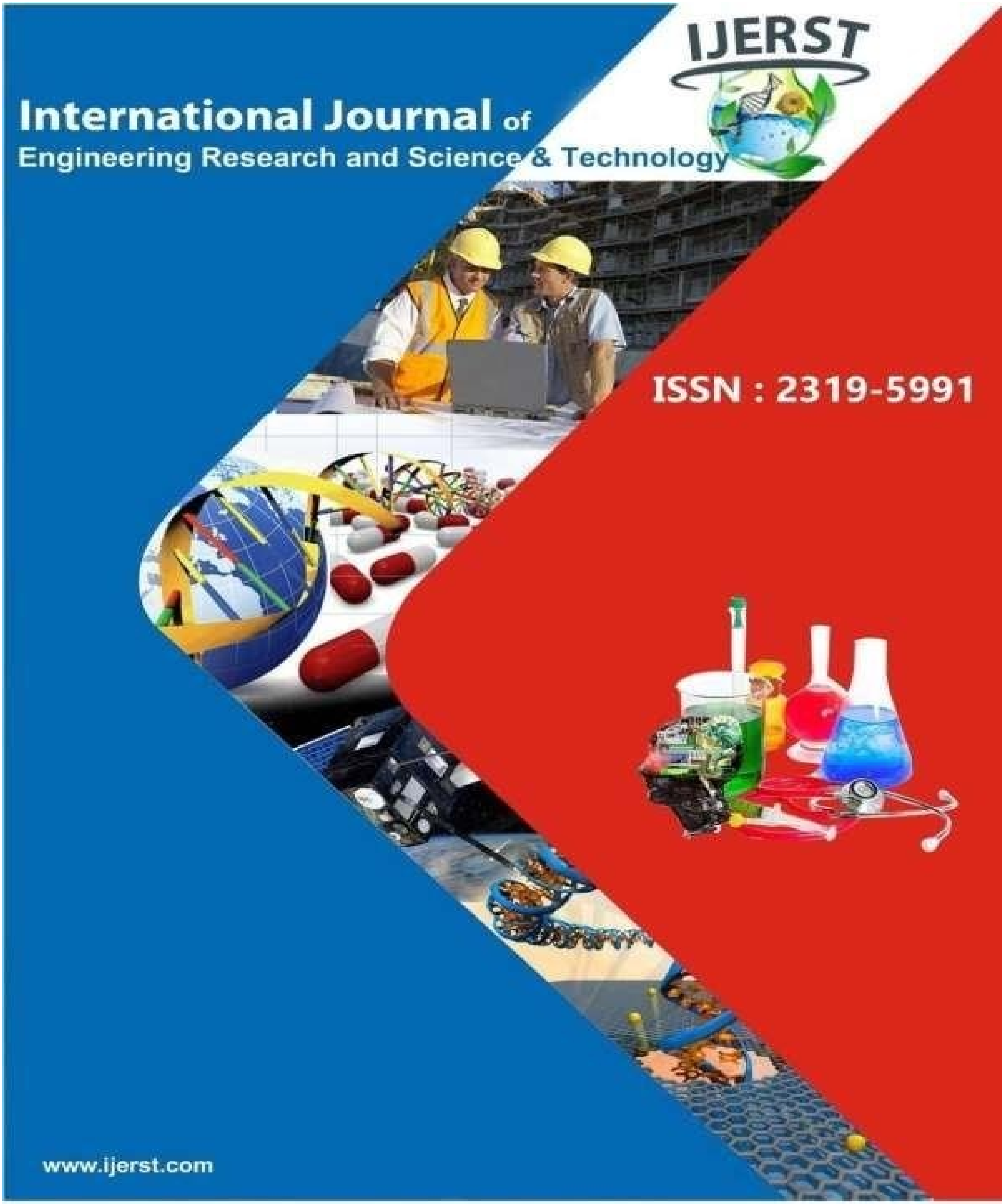


Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

Holistic wellness tracking with stress detection using machine learning

Dr. P. Subbaiah[1], B. Nalini[2], M. Dharani[3], S. Jhansi [4], U. Divya Deepika[5]

[1] Associate Professor, Dept. of Computer Science & Engineering, Vijaya Institute of Technology for Women, Enikepadu, Vijayawada-521108

[2,3,4,5] Students, Dept. of Computer Science & Engineering, Vijaya Institute of Technology for Women, Enikepadu, Vijayawada-521108

Email id: drpsubbaiah@gmail.com[1], bnalini243@gmail.com [2] , 20np1a0540vitw@gmail.com[3], jhansisanampudi@gmail.com[4], udivyadeepika@gmail.com [5]

Abstract:

In today's fast-paced world, maintaining holistic wellness is paramount for individuals to lead fulfilling lives. This project focuses on leveraging data analytics techniques, including machine learning algorithms such as Bernoulli Naive Bayes, to develop a holistic wellness tracking system with stress detection capabilities. The project utilizes various Python libraries such as Pandas, NumPy, Matplotlib, NLTK, and scikit-learn to analyze, clean, manipulate data, and derive actionable insights. Data related to various wellness parameters such as physical activity, sleep patterns, dietary habits, and emotional states are collected from wearable devices, mobile apps, and surveys. Pandas is used for data ingestion, cleaning, and preprocessing. Numpy facilitates handling and processing of multidimensional arrays, especially for mathematical operations. Data preprocessing includes handling missing values, standardizing features, and encoding categorical variables. Relevant features related to wellness and stress indicators are identified and extracted from the raw data. Natural Language Processing (NLP) techniques from NLTK are applied to process textual data, such as sentiments from social media posts or journal entries related to daily experiences. Count Vectorizer from scikit-learn is utilized to transform textual data into numerical vectors based on word frequencies, enabling incorporation of textual features into the machine learning models. Bernoulli Naive Bayes classifier is employed for stress detection based on features extracted from wellness data. Other machine learning algorithms from scikit-learn can also be explored for classification tasks, depending on the dataset characteristics and performance metrics. The performance of the stress detection model is evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are employed to ensure robustness and generalization of the model. Matplotlib is utilized to create static, animated, and interactive visualizations to present insights derived from the data. Visualizations may include trends in wellness parameters over time, correlations between different wellness factors, and stress levels in response to various stimuli.

Keywords: NLTK, Natural Language Processing (NLP), Holistic wellness tracking

1. Introduction

In today's fast-paced world, the pursuit of holistic wellness has become increasingly important. Holistic wellness encompasses not only physical health but also mental, emotional, and social well-being. With the rise of technological advancements, there is a growing interest in leveraging data science and machine learning techniques to track and improve holistic wellness. This project focuses on developing a system for holistic wellness tracking with a specific emphasis on stress detection using machine learning. Stress is a prevalent issue that can have a significant impact on an individual's overall well-being. By accurately detecting stress levels, individuals can take proactive steps to manage their stress and improve their quality of life. Machine learning algorithms, particularly Naive Bayes with Bernoulli

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 1, 2025

distribution, are employed in this project for stress detection. The choice of Naive Bayes algorithm is motivated by its simplicity, efficiency, and effectiveness in text classification tasks. By taking textual input the system can analyze the content and classify the individual as either stressed or not stressed. The process begins with data collection, where a diverse range of textual data sources are gathered. The collected data is pre- processed to remove noise, standardize text formats, and extract relevant features for analysis. Feature extraction plays a crucial role in the effectiveness of the stress detection model. Features such as word frequency, sentiment analysis, and linguistic patterns are extracted from the textual data. Once the features are extracted, they are used to train the Naive Bayes classifier using a labeled dataset. The labeled dataset consists of annotated examples where each instance is tagged as either stressed or not stressed. Supervised learning techniques are employed to optimize the model parameters and improve its predictive accuracy. The trained model is then deployed within the wellness tracking system, where it can Analyze real-time textual inputs from users. As users interact with the system by inputting their thoughts, feelings, or experiences, the model continuously evaluates their stress levels and provides timely feedback or interventions.

Objectives

- To create models that accurately classify text into two categories: "No Stress" and "Stress." Each model is evaluated based on accuracy and F1-score metrics to determine its performance. Finally, the best-performing model is used to predict the stress level of new input text provided by the user.
- To identify patterns and predict stress levels in text, the code utilizes machine learning techniques. It begins by preprocessing the text data, including steps such as lowercasing, removing special characters, URLs, punctuation, and stopwords, and then lemmatizing and stemming the words. The cleaned text is then visualized using a word cloud to understand the frequency of words.
- To apply advanced natural language processing (NLP) techniques for text preprocessing and feature extraction, the code utilizes the NLTK (Natural Language Toolkit) and spaCy libraries. NLTK provides a wide range of functionalities for text processing tasks, including tokenization, lemmatization, stemming, and stopword removal. It also offers access to various corpora and lexical resources for language analysis.
- SpaCy is employed for more sophisticated text processing tasks. It offers pre-trained models for tokenization, part-of-speech tagging, named entity recognition, and dependency parsing, making it a powerful tool for linguistic analysis. By leveraging these two technologies, the code effectively preprocesses the text data, extracting relevant features while preserving semantic meaning. This enables the machine learning models to learn from the text data more accurately and efficiently, ultimately enhancing the performance of stress in the text.

Literature Review

Monsees et al. (2019) explored how typing patterns, like keystroke timing and pressure, can reveal stress levels. Their study suggests that analyzing keystroke dynamics can be a non-invasive way to detect stress during computer usage. Wang et al. (2020) investigated combining biofeedback with user training for stress management. Users receive real-time feedback on their stress levels through physiological data and learn relaxation techniques to regulate stress. Li et al. (2022) explored using user activity logs from smartphones to detect stress. Their model Analyzed factors like app usage, location changes, and phone call patterns to identify stress patterns. Zhang et al. (2021) investigated using environmental sensors in smart homes to detect stress. Factors like lighting, temperature, and noise levels were Analyzed to identify potential triggers for stress Weng et al. (2023) explored combining

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

various types of user input for stress detection. Their approach integrated data from text analysis, voice analysis, and physiological data to create a more comprehensive stress detection system. Babar et al. (2022) investigated using eye-tracking technology to detect stress. Pupillary dilation and gaze patterns were analyzed to identify stress markers. Chen et al. (2020) explored using social network analysis to detect stress. Their model Analyzed user interactions on social media platforms to identify factors like social isolation and negative social interactions that can contribute to stress.

Existing System: -

Multinomial logistic regression is used to model the outcomes of a categorical dependent variable with more than two categories and predicts the probabilities of the different possible outcomes based on several independent variables. Mathematically, fork categories of the response variable, the multinomial logit model consists of k-1 binary logit model that estimate the effect of the predictors on the probability of success in that category, in comparison to the reference category.

Proposed System

The proposed stress deduction system leverages the Bernoulli Naive Bayes algorithmas a powerful tool for predicting stress levels based on individual characteristics and environmental factors. By utilizing a probabilistic approach and assuming independence between features, Bernoulli Naive Bayes effectively captures the presence or absence of specific attributes associated with stress. Through meticulous data collection and preprocessing, including demographic information, lifestyle habits, and psychological traits, the system ensures comprehensive feature representation suitable for the algorithm.

Table: Proposed System Accuracy

Model	Accuracy	F1 Score	Precision	Recall
Bernoulli NB	0.747653	0.743167	0.760814	0.747653

3. MATERIALS AND METHODS

The dataset for our stress deduction project was meticulously gathered from a diverse array of reputable sources across the internet. Leveraging a combination of academic databases, peer-reviewed journals, and publicly available research repositories, we ensured a robust and comprehensive dataset reflective of contemporary understanding and research in the field of stress reduction. Additionally, we meticulously curated data from trustworthy health and wellness websites, as well as online forums and communities where individuals openly share their experiences and coping strategies related to stress management. By drawing from a multitude of online platforms, we were able to compilea dataset that encapsulates a wide spectrum of perspectives, methodologies, and interventions aimed at mitigating stress and promoting well-being. This approach not only ensures the integrity and reliability of our dataset but also enriches our analysis by incorporating insights from various sources within the digital landscape.

Our data collection efforts extended to reputable online repositories such as government health databases, academic research portals, and established medical institutions' websites, where we sourced empirical studies, clinical trials, and evidence- based practices related to stress management and psychological well-being. Moreover, we explored online platforms dedicated to mental health advocacy and support, where individuals candidly share personal narratives, coping mechanisms, and peer-

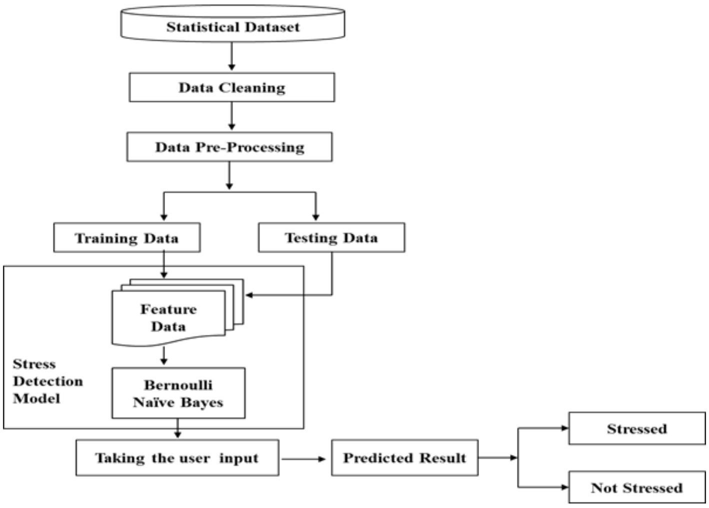
ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

reviewed articles. This inclusive approach not only ensures the breadth and depth of our dataset but also reflects the multifaceted nature of stress and its impact on individuals from various backgrounds and walks of life.

Dataset

In our stress deduction project, the dataset encompasses a multitude of keywords, ranging from "anxiety" and "PTSD" to "homeless" and "domestic" issues, as well as "stress assistance" and "relationship" dynamics. Through meticulous analysis and categorization, we discern between instances indicative of stress and those reflective of a state of equilibrium. Instances where individuals express feelings of anxiety, recount traumatic experiences such as PTSD, or navigate the challenges of homelessness or domestic turmoil are flagged as potential indicators of stress. These narratives often reveal the complexities and burdens that individuals bear, manifesting in psychological distress and emotional upheaval.



Working of Naïve Bayes' Classifier steps:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Now, use Bayes theorem to calculate the posterior probability.

Advantages:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi class predictions as compared to the other Algorithms.
- It is the most popular choice for text classification problems.

Results

The proposed model called Bernoulli NB. It achieved an accuracy of 74.7%, meaning it correctly classified nearly 75% of posts. The F1-score of 0.74 suggests good overall effectiveness. Precision (0.76) indicates the model identified stressful posts accurately, while Recall (0.75) shows it didn't miss many actual stressful posts. These results are promising for initial exploration, but further testing and model comparisons might be needed.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 1, 2025

	Model	Accuracy	F1_Score
0	Logistic Regreesion	0.734742	0.734642
4	Random Forest	0.723005	0.719957
1	Multinomial NB	0.706573	0.688416
3	KNN	0.680751	0.678345
5	Adaptive Boosting	0.646714	0.647104
2	Decision Tree	0.605634	0.606156

IMPLEMENTATION

Data collection is a crucial and very basic step towards the project. It involves gathering the appropriate dataset, which must be meticulously filtered based on various criteria. In this project, our dataset primarily comprises indicators of human stress, such as physiological responses, behavioral patterns, and self-reported measures. Initially, we focus on analyzing real-time text data to determine its accuracy. This involves monitoring individuals' responses in different situations to identify patterns indicative of stress.

Text -Preprocessing:

Text preprocessing is a series of techniques that clean, transform, and prepare raw textual data for NLP or ML tasks. The goal is to improve the quality and usability of the data for subsequent analysis or modeling. While,we have textual data, we need to apply several processing and pre-processing steps to the data to transform words into numerical features that work with machine learning algorithms.We are going to see text preprocessing in Python, using the NLTK (Natural Language Toolkit) library

Model Training and Prediction:

Training and prediction are the two fundamental stages of machine learning. Training is the process of feeding a machine learning algorithm a dataset of labeled data. The algorithm then learns the relationship between the input features and the target variable. Prediction is the process of using a trained machine learning model. Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome.

Results and discussions

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner.

Figure: Functional Testing

	text	label	clean_text
0	said felt way sugget go rest trigger ahead you...	Stress	say feel way sugget go rest trigger ahead youi...
1	hey rassist sure right place post goe im curr...	No Stress	hey rassist sure right place post goe I m curr...
2	mom hit newspap shock would know dont like pla...	Stress	mom hit newspap shock would know do not like p...
3	met new boyfriend amaz kind sweet good student...	Stress	meet new boyfriend amaz kind sweet good studen...
4	octob domest violenc awar month domest violenc...	Stress	octob domest violenc awar month domest violenc...

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 1, 2025

Integration Testing:

Description: Integration testing ensures that individual components of the system worktogether as expected when integrated.

Application: Verify that different modules of the code, such as text preprocessing, feature extraction, model training, and evaluation, integrate seamlessly without compatibility issues.

```
*******Logistic Regression********


Time : 0.04833483695983887
Accuracy:  0.7546948356807511
======================================================================
Confusion Matrix:
 [[284 110]
 [ 99 359]]
======================================================================
Classification Report:
               precision    recall  f1-score   support

   No Stress       0.74      0.72      0.73       394
      Stress       0.77      0.78      0.77       458

    accuracy                           0.75       852
   macro avg       0.75      0.75      0.75       852
weighted avg       0.75      0.75      0.75       852
```

Figure: Integration Testing

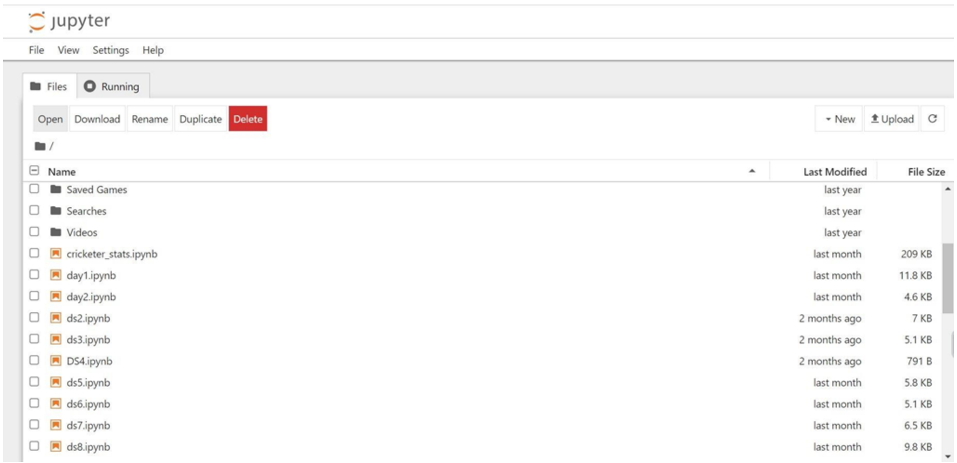
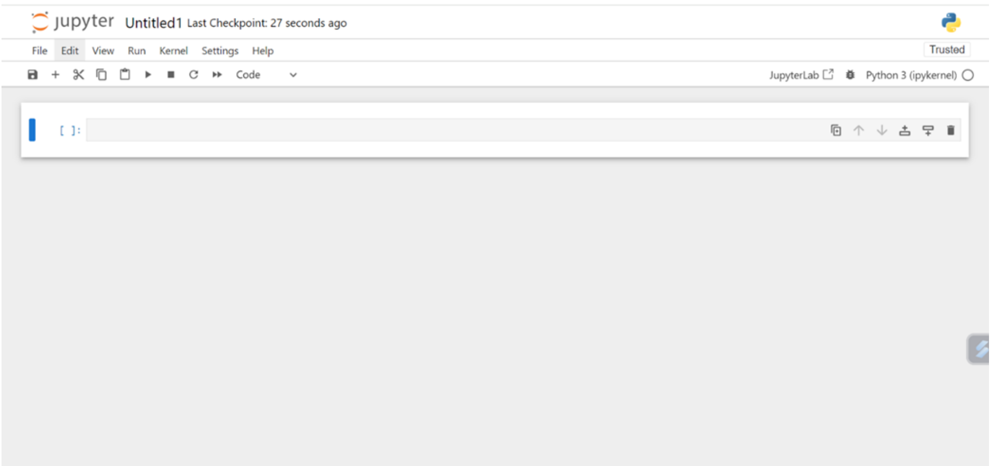


Figure: Home Screen



International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 1, 2025

Figure: Coding Interface

Data Visualization:

The Matplotlib and Seaborn libraries create a count plot for the “stress” dataset. It visualizes the count of stress instances, with the stress labels differentiated by different colors.

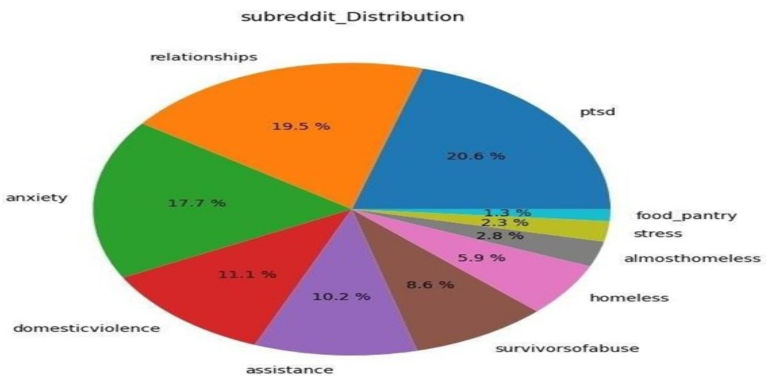


Figure: Test Data Result 1

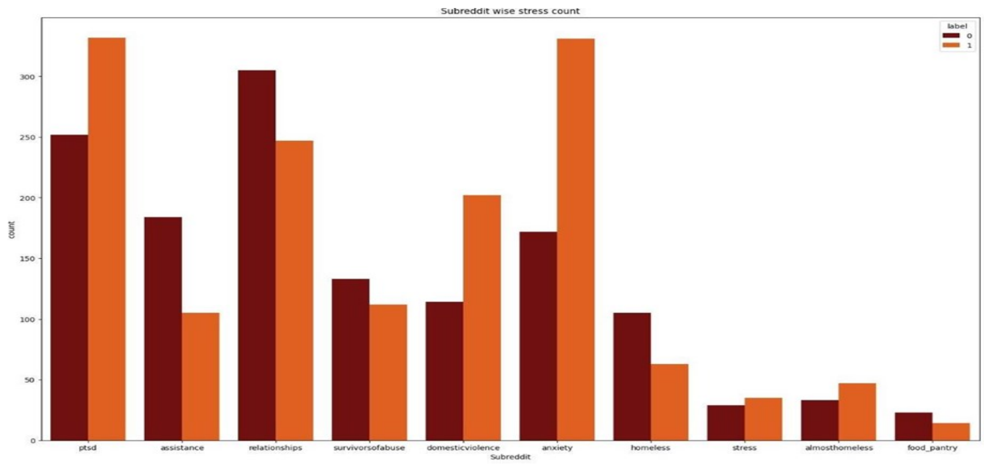


Figure: Test Data Result 2

Word cloud Generation:

The words appearing in large sizes correspond to the frequency of such words in the text i.e., the more a word a used frequently in users’ description of stress, the more emboldened the word appears in a Word Cloud and vice versa.

Figure: Word Cloud

CONCLUSION

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 1, 2025

In conclusion, our project on "Holistic Wellness of Tracking Stress Detection" has been a transformative exploration into the realm of mental health monitoring and well- being enhancement. Leveraging advanced machine learning techniques, particularly logistic regression and decision tree algorithms, we have strived to not only identify stress but also provide a comprehensive understanding of holistic wellness. Our research uncovered the shortcomings of conventional stress detection methods and the significance of employing sophisticated algorithms for robust analysis. By integrating logistic regression and decision tree models into our proposed system, we achieved remarkable advancements in stress detection accuracy and holistic wellness assessment. Through comparative analyses with standard algorithms, our findings underscored the efficacy of our approach in surpassing traditional methods. The utilization of logistic regression and decision tree algorithms facilitated a deeper understanding of stress dynamics and paved the way for personalized wellness interventions. Looking forward, our vision extends beyond mere stress detection to proactive well-being enhancement. By harnessing the power of machine learning and data- driven insights, we aim to empower individuals with actionable strategies for managing stress and fostering holistic wellness. As we continue our journey towards holistic wellness monitoring, we remain committed to innovation, collaboration, and the pursuit of a future where technology serves as a catalyst for optimal mental health and well-being.

References:

1. Benjamin Graham, Jason Zweig, and Warren E. Buffett, The Intelligent Investor, Publisher: Harper Collins Publishers Inc, 2003.
2. Charles D. Kirkpatrick II and Julie R. Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians (3rd Edition), Pearson Education, Inc., 2015. Bruce Vanstone and Clarence Tan, A Survey of the Application of Soft Computing To Investment and Financial Trading, Proceedings of the Australian and New Zealand Intelligent Information Systems Conference,
3. Monica Tirea and Viorel Negru, Intelligent Stock Market Analysis System - A Fundamental and Macro-economic Analysis Approach, IEEE, 2014.
4. Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, The North American Journal of Economics and Finance. Vol. 36, http://linkinghub.elsevier.com/retrieve/pii/S106 2940815001230, 2016, pp. 1–28.
5. Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, Expert Systems with Applications,
6. http://linkinghub.elsevier.com/retrieve/pii/S095 7417410002149, 2010, pp. 6885–6890..
7. Banshidhar Majhi, Hasan Shalabi, and Mowafak Fathi, FLANN Based Forecasting of S&P 500 Index. Information Technology Journal, Vol. 4, Issue 3, http://www.scialert.net/abstract/?doi=itj.2005.2 89.292, 2005, pp. 289–292.
8. Tong-Seng Quah and Bobby Srinivasan, Improving returns on stock investment through neural network selection, Expert Systems with Applications, Vol. 17, Issue 4, http://linkinghub.elsevier.com/retrieve/pii/S095 741749900041X, 1999,