Grenze International Journal of Engineering and Technology, June Issue

Stress Prediction using Machine Learning Algorithms

B.V.Sivaiah[1], Tangi Setti Gayatri [2], Thati Umadevi [3], Rajula Partha Siva Reddy[4] and Modiboina Sravya[5]
[1]Professor, Dept. of Data Science, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati.India
Email: Siva.bheem@hotmail.com
[2-5]UG Scholar, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, India
Email: tangisettygayatri@gmail.com, umadevithati9@gmail.com,parthasiva@gmail.com,sravyamodiboina1219@gmail.com

***Abstract*—Many people nowadays are dealing with psychological trauma. referred to as stress. Stress is defined as "a state of tension and pressure brought on by any thought or incident that makes you feel anxious, agitated, or frustrated". Many people have experienced stress as a result the current situation, particularly youth and working professionals. These days, a rise in stress can cause a number of issues, including heart attacks, depression, and suicide. Many academics have turned their attention to stress identification as mounting data indicates that stress related health illnesses associated with the demanding modern lifestyle are becoming more and more common. Consequently, this study's main objective is to predict stress levels, using user-retrieved data that is then confirmed using a subjective stress scale. In this case, a ML algorithm should be trained on the data to forecast the stress level. With the help of our stress prediction, you can quickly determine your current stress level and steer clear of any unfavorable circumstances. With the help of this approach, people may proactively handle future problems and avoid unfavorable consequences by having correct insights into their stress levels. The project's ultimate goal is to improve mental health by providing a tool for efficiently determining and controlling stress levels and encouraging a better way of living.**

***Index Terms*— Stress Prediction ,psychological, pressure, ,tension.**

I. INTRODUCTION

One of the primary symptoms of human existence is stress. Stress plays a significant part in people's lives. Numerous illnesses, including cancer, heart disease, lung issues, breathing difficulties, and other conditions, can be brought on by stress. The growth in global population has led to a rise in people's stress levels. Stress is a prevalent condition among all people these days. Major life-or-death issues are caused by stress in people all over the world. People that are under stress tend to behave differently from normal, which might cause issues with daily living. Inorder to effectively analyze their behavior within the system, we must continuously monitor their behavior using ML algorithms that are integrated into the system. This will provide us with real-time data. In India, stress affects 85% of the population. These days, many experience depression as a result of stress. The stress detection system will distinguish between a person exhibiting normal behavior and one who is under stress. Physiological sensors have been used to continuously monitor stressed individuals, helping clinicians forecast if a patient is normal or stressed out based on parametric data. In parallel, physiological sensors record a person's behavior and store it in a database. The data is then kept there for physicians to review at a later time. The physicians may readily estimate the patient's stress level based on such figures. Thus, these analyses

Grenze ID: 01.GIJET.10.2.664_1

determine if the subject is under stress or not.
Stress is characterized as an individual experiencing excessive strain and discomfort when working, sitting quietly and pondering, or feeling as though his mind is not with him. There are differences in the relationships with the individuals too. The person will not be content under stress. Whether the patient is sad or under stress, we need to speak with a doctor to receive advice. The doctor will assess the patient's stress level and recommend a neurologist and psychiatrist for additional testing and diagnosis. When someone is under stress, they become scared since they don't know if they will recover or not. When a patient experiences stress in its early stages, doctors may treat them with ease; but, if they already have a chronic illness, they will eventually develop new conditions. Stress can lead to the development of illnesses such as asthma, cancer, and heart-related conditions. When under stress, a person will sit by themselves and think constantly.
retrieval, and various NLP applications, benefiting researchers, businesses, and individuals relying on textual data interpretation.

II. Literature Survey

Ref. [1] In this research, a sentiment analysis technique to automatically identify human tension and relaxation using Twitter data is proposed. In order to identify stress and relaxation phrases on a scale from -5 to +5, the system uses a lexical approach and the TensiStrength framework for sentiment strength detection. Even with problems like Word Sense Disambiguation, the approach improves accuracy by employing pre-processing techniques like trigram, bigram, and unigram analysis for part-of-speech disambiguation. Results are improved when SVM is used with Ngram, yielding 65% precision and 67% recall. Finding overt and covert signs of tension and relaxation in tweets is the main goal.
Ref. [2] In order to increase the safety of women and girls, this project aims to develop a wearable device that combines physiological signal analysis and body position tracking. The device uses triple-axis accelerometer data to determine body position in addition to galvanic skin resistance analysis and body temperature. For real-time monitoring, a customized ML system recognizes activity from raw accelerometer data. Continuous monitoring is made possible via the wireless transfer of sensor data to a platform hosted on open-source clouds, and MATLAB is used for simultaneous analysis. The wearable gadget is intended to monitor the subject's characteristics on its own and react to any changes in signals that are being monitored, indicating possible danger. In certain situations, the gadget sounds an alert or notifies the user, guaranteeing quick response and encouraging proactive personal safety practices. This creative approach has the potential to improve general safety precautions for women and girls as well as empower people.
Ref. [3] This study presents a novel method for anticipating driver stress, which is a major cause of traffic fatalities and accidents. In contrast to conventional physiological signal measuring techniques, this approach makes use of an inertial motion unit sensor attached to the driver's glove, providing a less intrusive and more affordable option. Experiments were carried out to replicate various stress situations in urban, highway, and rural environments. In order to forecast stress levels, the motion sensor's recorded patterns of steering wheel movement were examined. The driver's palm was used to measure the galvanic skin reaction, which was used to communicate low-energy Bluetooth reference stress data. The outcomes showed how effective this indirect method was, with an average accuracy of 94.78%. The results point to the possibility of incorporating motion sensors for real-time stress level monitoring into wireless driver aid systems, which might enhance traffic safety.
Ref. [4] A cardiac surgeon has observed that the nonlinear structure of heart rate makes age prediction difficult. However, stress detection and fitness level prediction are made possible by utilizing heart rate data, particularly when age is known. Stress is a ubiquitous issue that affects many lives and requires early identification and treatment. This study suggests a way to forecast stress levels using heart rate as a measure in an effort to warn people about their unhealthy lifestyles before serious diseases develop. This prediction capacity is improved by the combination of ML and the internet of things (IoT). While IoT enables real-time communication to notify people of possible threats, ML algorithms evaluate the patient's status. By facilitating prompt intervention and lifestyle changes, this combination strategy provides a proactive solution that helps avoid stress-related adverse health consequences.
Ref. [5] This research tackles psychological stress by monitoring physiological markers during physical activity, such as body temperature, motion rate, and perspiration, and introduces a revolutionary stress detection system, iStress. With more than 150 examples serving as use a Mamdani-style fuzzy logic controller in conjunction with a neural network method. the model, the implementation uses a neural network technique with a Mamdani-type fuzzy logic controller. In order to monitor stress levels in real time and lower health risks, iStress gathers and sends data to the cloud. Surprisingly, in real-time circumstances, the system uses very little energy. The

suggested iStress system has a 97% accuracy rate, is reasonably priced, and has minimal complexity, making it a viable and effective stress detection method. This novel strategy has the potential to be widely adopted and improve overall wellbeing and proactive stress management.

III. METHODOLOGY

The suggested solution uses ML techniques to forecast stress in an effort to address the growing issue of stress. This creative method makes use of usersupplied data that is analyzed and trained using sophisticated ML models. The Perceived Stress Scale is used to confirm the prediction accuracy, providing a trustworthy way to gauge stress levels. The device is primarily intended for teens and working adults, and it acts as a preventative measure to precisely gauge stress levels. In doing so, it helps people deal with obstacles and steer clear of negative outcomes linked to elevated stress, such depression and heart-related problems. By putting this stress prediction model into practice, people may get insightful knowledge about their mental health, which helps them adopt better habits and take preventative action. In the end, the initiative hopes to support general mental wellbeing in today's demanding and fast-paced workplace and aid in the comprehensive management of stress.

A. Data Pre-processing

The processes of data preprocessing, feature selection, and feature extraction are critical to improving the efficacy and precision of stress level forecasts in the Stress Prediction System. Data Preprocessing: To guarantee the quality and consistency of the user-input data, the first stage is cleaning and preparing it. This procedure takes care of any outliers, inconsistent values, and missing values in the dataset. To ensure consistency in the input data, standardization and normalizing procedures can be used to bring all characteristics to a single scale. In order to convert categorical variables into a format that is appropriate for ML algorithms, they may also be encoded. By optimizing the dataset for additional analysis, data preparation increases the accuracy of stress forecasts.

System Architecture

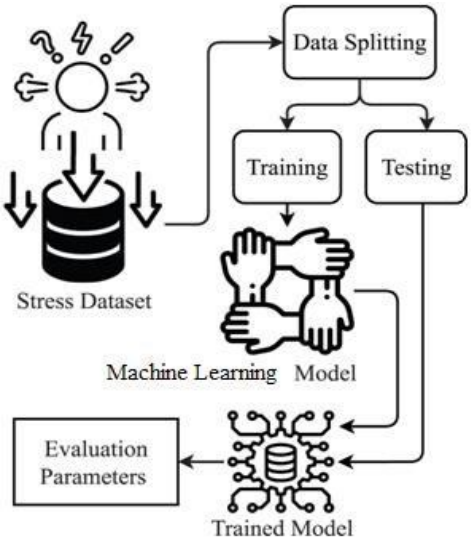

Fig 1 Proposed Architecture

B. Feature Selection and Engineering:

Feature selection is used to identify and retain the most relevant variables in order to optimise the efficacy of the stress prediction model. Techniques like Recursive Feature Elimination (RFE) or feature importance from tree-based algorithms (like Random Forest) can be used to select the most informative features. This stage reduces dimensionality, increases interpretability, and decreases overfitting risk. Relevant data significantly improves the anticipated accuracy of machine learning systems.

C. Feature Extraction

In feature extraction, important information is preserved while converting unprocessed data into a new collection of features, frequently with fewer dimensions. Latent patterns in the dataset can be discovered using methods

like Principal Component Analysis (PCA) or t-distributed Stochastic Neighbor Embedding (t-SNE). This enhances the overall performance of the model by producing a more effective representation of the data. By enhancing the predictive potential of the chosen characteristics, feature extraction helps to optimize the stress prediction system for precise and trustworthy outcomes.
Principal Component Analysis (PCA) or t-distributed Stochastic Neighbor Embedding (t-SNE). This enhances the overall performance of the model by producing a more effective representation of the data. By enhancing the predictive potential of the chosen characteristics, feature extraction helps to optimize the stress prediction system for precise and trustworthy outcomes.

D. Training and Testing

The ML algorithms in the Stress Prediction System go through a rigorous training procedure to identify patterns in the preprocessed information, giving the model the capacity to forecast stress levels with accuracy.
Preprocessed data is fed into a variety of ML methods, including LR, NB, DT, RF, and XGBoost, during the training phase. To create a prediction model, these algorithms examine the connections, correlations, and trends in the data. The algorithms modify internal parameters during training in order to maximize their capacity to generalize and produce precise predictions on unobserved data
The model is tested after it has been trained to assess how well it performs with fresh, untested data. For this, a different dataset—one different from the training dataset—is utilized. The model's capacity to generalize its learnt patterns and forecast stress accurately for situations it hasn't experienced during training is evaluated during the testing phase. Measures like recall, accuracy, precision, and F1-score are used to assess how well the model performs. Thorough testing guarantees the Stress Prediction System's validity and generalizability, enabling it to offer precise and significant insights into a person's stress levels. The model is subjected to ongoing assessment and improvement in order to maximize its performance in practical situations, hence augmenting its overall resilience and potency.

IV. ALGORITHMS

A. Logistic Regression

Among the supervised learning algorithms is the wellknown logistic regression algorithm. Based on a group of independent variables, it is used to predict a categorical dependent variable. Logistic regression predicts the outcome of a categorical variable, resulting in a discrete value between 0 and 1 instead of precise values. The predicted outcome can be binary, such as Yes or No, 0 or 1, true or false, and so on. The only difference between logistic regression and linear regression is the way they are applied. For classification issues, logistic regression is utilized, and for regression problems, linear regression.
Instead of employing a straight line, logistic regression uses a logistic function with a "S" shape to predict two maximum values.
The logistic function curve is like a line that helps us predict what might happen in different situations. For example, it can help us figure out if cells in our body are harmful or not, or guess if a mouse is too heavy or not. Logistic regression is a helpful way for computers to figure out the chances of something happening and put new information into different groups, using both smooth and separate sets of information. A statistical method for classifying observations and identifying the factors that influence the classification is called logistic regression. To reach these conclusions, it draws on many data sources and the logistic function.

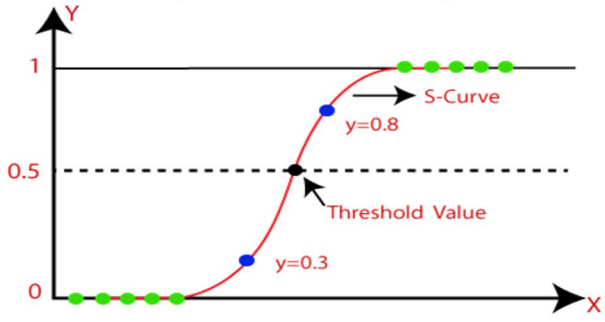

Fig 2 Logistic Regression

It can change a decimal number between 0 and 1 into a different kind of number. The logistic regression produces a curve that resembles the letter "S" because its result is limited to a range of 0 to 1. This curve is

sometimes called the Scurve because it looks like the letter "S". In logistic regression, we use a special number called a threshold to decide if something is more likely to be a 0 or a 1. If a number is bigger than the threshold, we think it is more likely to be a 1. But if the number is smaller than the threshold, we think it is more likely to be a 0.

The dependent variable must have a category component. There should not be any multicollinearity among the independent variables. The logistic regression equation is a way to predict things based on certain information. It is related to another equation called linear regression. To get the logistic regression equation, you have to do some special math steps. The following is one way to illustrate the straight line equation.

$$y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

The value of y LR needs to fall between 0 and 1. We shall divide the equation by (1-y) to take this into account.

$$\frac{y}{1-y}\text{; 0 for y= 0, and infinity for y=1}$$

It is necessary for the values to fall between negative and positive infinity. The equation changes as a result of taking its logarithm.

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

B. Naïve Bayes

The Naïve Bayes algorithm is a method for solving classification problems in supervised learning. It relies on the principles of the Bayes theorem. The main purpose of this is to categorize text, using a dataset that has many different dimensions.

The Naïve Bayes classifier is like a special tool that helps us teach computers to make predictions really quickly. It's simple to use and works really well in creating models that can guess things fast. As a probabilistic classifier, it predicts the likelihood of an object occurring. Some common uses of the Naïve Bayes algorithm include categorizing articles, analyzing sentiment, and filtering out spam.

The Bayes algorithm is made up of the terms "Naïve" and "Bayes", which refer to someone who is easily fooled and a statistical method, respectively.

Naïve: This concept is called the naive assumption because it assumes that one characteristic of something can be considered independently of other characteristics. For example, if we define a fruit based on its color, shape, and flavor, an apple would be described as a red, round, and sweet fruit. Each of these characteristics is seen as separate from the others when identifying an apple.

Bayes: It is called Bayes because it uses Bayes' Theorem.

Bayes Theorem:

Bayes' theorem is a way to figure out how likely something is based on what we already know. This probability is determined using conditional probability.

The following is the formula for Bayes' theorem:

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

C. Decision Tree

Although they may be applied to regression issues as well, DT are most frequently utilized to solve classification difficulties. This kind of classifier has a tree-like structure, where nodes stand for qualities, branches for decision rules, and leaf nodes for results. Although they may be applied to regression issues as well, DT are most frequently utilized to solve classification difficulties. This kind of classifier has a tree-like structure, where nodes stand for qualities, branches for decision rules, and leaf nodes for results. To execute a test or make decisions , the dataset's information is used. This tool visually displays all the possible solutions or decisions that can be made based on specific parameters. A Decision tree gets its name because it starts with a root node and expands with branches, resembling the structure of a tree. The CART algorithm is utilized for building trees in classification and regression tasks. A decision tree asks a question and then separates into smaller sections based on the answer.

When developing a ML model, the first thing to consider is choosing the best method for the specific dataset and problem. There are many ML algorithms to choose from, and one option is using a Decision tree. Decision trees

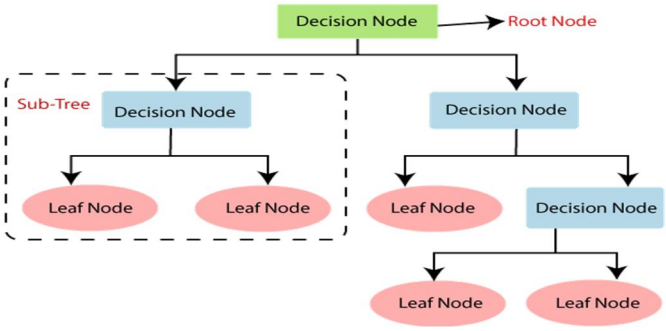

Fig 3 Decision Tree

are typically simple to comprehend since they simulate how people might think when faced with a decision. The decision tree's structure resembles a tree, making it easy to understand the reasoning behind it.

D. Radom forest

A well-liked ML approach for classification and regression issues is called RF. It is predicated on the idea of ensemble learning, which entails merging many classifiers to enhance the model's functionality and address challenging issues. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Rather than relying solely on the opinion of one DT, the random forest incorporates the views of several trees and bases its prediction on what the majority of the trees believe. Because there are more trees in the forest, accuracy is higher and overfitting is avoided.

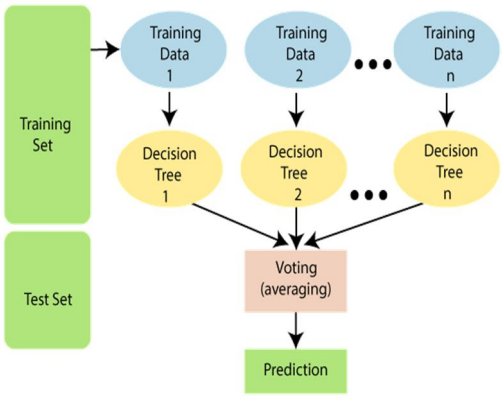

Fig 4 Random Forest

E. XGBoost

XGBoost is like a special tool that helps computers learn and make predictions. It is made to work well with lots of data and can make really accurate predictions by combining the guesses of many different models. Because it can handle big datasets and perform well in a variety of ML tasks, including regression and classification, XGBoost is extensively used and well-liked. XGBoost has a strong ability to handle missing values, making it suitable for real-world data without needing extensive preprocessing. Additionally, it has built-in parallel processing capabilities that allow for fast model training on large datasets. XGBoost has a wide range of applications such as predicting click-through rates, creating recommendation engines, and participating in Kaggle tournaments. Its high level of customization allows for performance optimization by adjusting various model parameters. XgBoost is a package developed by researchers at the University of Washington to enhance gradient boosting traning

V. Conclusion

In conclusion, there is great potential for resolving the growing mental health issues linked to stress through the creation of a stress prediction system that uses ML algorithms. The initiative offers a proactive and individualized approach to mental well-being since it focuses on precisely projecting stress levels using user-

provided data that is validated against recognized scales. This technology enables people to take timely preventative action and successfully manage the difficulties of modern life by giving them informed forecasts. The ability to detect stress levels at an early age not only helps prevent negative consequences such as heart-related problems and depression but also fosters a better way of living. With the increasing occurrence of stress-related health illnesses, the suggested approach is a useful instrument for raising mental health consciousness and developing a culture that values early intervention in stress management. All things considered, this research is a notable step toward utilizing technology to solve the widespread issues related to stress in modern life.

FUTURE WORK

The Stress Prediction System's potential integration with wearable technology and continuous monitoring applications represent its future promise. Increasing the diversity of datasets and adding real-time physiological data would improve the accuracy and application of the model. Furthermore, the system's predictive skills may be further enhanced by including more sophisticated ML algorithms and investigating partnerships with mental health specialists. This would promote proactive stress management and enhanced mental health on a larger scale.

REFERENCES

[1] Reshma Radheshamjee, and Supriya Kinariwala “Detection and Analysis of Stress using Machine Learning Techniques”International Journal of Engineering and Advanced Technology, pp. 22498958, 2019.

[2] Anand Jatti, Madhvi Kannan, Alisha R. M, Vijayalakshmi P, Shrestha Sinha, “Design and Development of an IOT based Wearable device for the safety and security of women and girl children,” IEEE International conference on recent trends in electronics information communication technology, 2016.

[3] Boon Giin Lee, Wan-Young Chung “Wearable Glove-Type Driver Stress Detection Using a Motion Sensor,”IEEE Transaction on Intelligent Transportation systems,” 2018.

[4] Purnendu Shekhar Pandey, “Machine Learning and IoT for Prediction and Detection of Stress,” 2017.

[5] Lavanya Rachakonda, Prabha Sundaravadivel, Saraju P. Mohanty, Elias Kougianos, Madhavi Ganapathiraju, “A smart sensor in the IoMT for stress level detection.”

[6] Tanmay Nema, Vedarth Kumtekar, Siddhivinayak Kulkarni, "Text Enhanced Ensemble for Human Stress Prediction", 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), pp.1-8, 2023

[7] Puneeth N. Thotad, Shanta Kallur, Laxmi Mundaragi, Swati H. Kadam, "Mental Health Tracker Using Machine Learning Approaches", 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), pp.1-5, 2023.

[8] Saloni Jage, Shubham Chaudhari, Manthan Jatte, Abhishek Mhatre, Vanita Mane, "Predicting Mental Health Illness using Machine Learning", 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), pp.1-5, 2023.

[9] Prathamesh Yadav, Shilpa Shinde, Rajashree Shedge, "Mental Health Disorder Detection Using Machine Learning and Deep Learning Tehniques", 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), pp.1-6, 2023.

[10] Suja Cherukullapurath Mana, G. Kalaiarasi, Yogitha R, L Suji Helen, R. Senthamil Selvi, "Application of Machine Learning in Healthcare: An Analysis", 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp.1611-1615, 2022.

[11] Sanduni Nilushika Gamage, P. P. G. Dinesh Asanka, "Machine Learning Approach to Predict Mental Distress of IT Workforce in Remote Working Environments", 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE), vol.5, pp.211-216, 2022.

[12] Swetha Reddy Anthay, Vangimalla Nagarjuna, Tathireddy Venkata Mahesh, Tamba Aravind Royal, Sagiraju Vivid Varma, "Detection of Stress in Humans Wearing Face Masks using Machine Learning and Image Processing", 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp.1104-1110, 2022.