DOI: https://doi.org/10.52756/ijerr.2024.v42.020 *Int. J. Exp. Res. Rev.*, Vol. 42: 228-240 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Emoji Support: Predictive Mental Health Assessment Using Machine Learning: Unveiling Personalized Intervention Avenues

Ashish Dixit[1], Avadhesh Kumar Gupta[2], Neelam Chaplot[3] and Veena Bharti[4]

[1]Department of CSE, Ajay Kumar Garg Engineering College Ghaziabad, Uttar Pradesh, India; [2]Department of CSE, Bahra University, Shimla Hills Himanchal Pradesh, India; [3]Department of CSE, Manipal University Jaipur, Jaipur, Rajasthan, India; [4]Department of CSE, Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh, India

E-mail/Orcid Id:

AD, ashishdixit1984@gmail.com, https://orcid.org/0000-0002-3842-6934; *AKG,* dr.avadheshgupta@gmail.com, 0000-0003-0173-7664; *NC,* neelam.chaplot@gmail.com, https://orcid.org/0000-0002-1770-7921; *VB,* bharti.veena@gmail.com, https://orcid.org/0000-0003-4494-652X

Article History:

Received: 18th Nov., 2023
Accepted: 09th Aug., 2024
Published: 30th Aug., 2024

Keywords:
Mental Health Assessment, Decision Trees, Naive Bayes, Support Vector Machines (SVM), Random Forests

How to cite this Article:
Ashish Dixit, Avadhesh Kumar Gupta, Neelam Chaplot and Veena Bharti (2024). Emoji Support: Predictive Mental Health Assessment Using Machine Learning: Unveiling Personalized Intervention Avenues. *International Journal of Experimental Research and Review*, 42, 228-240.
DOI: https://doi.org/10.52756/ijerr.2024.v42.020

Abstract: Mental health disorders, including anxiety, depression, and stress, profoundly impact individuals' well-being and necessitate effective early detection for timely intervention. This research investigates the predictive capabilities of machine learning algorithms in assessing anxiety, depression, and stress levels based on questionnaire-derived scores. Utilizing a dataset comprising self-reported scores obtained through a tailored questionnaire designed for mental health assessment, we delve into the application of Decision Trees, Naive Bayes, Support Vector Machines (SVM) and Random Forests for prediction. Data preprocessing involved comprehensive cleaning, encoding categorical variables and careful feature selection, ensuring the relevance of features in the predictive models. Each algorithm underwent individual implementation, wherein we scrutinized their performances in predicting mental health conditions. Evaluation metrics such as accuracy, precision, and recall were employed to assess the models' proficiency in predicting anxiety, depression and stress levels. The findings underscore the potential of machine learning in accurately predicting mental health conditions based on questionnaire responses, offering insights into personalized interventions and early detection systems. This study contributes to advancing the understanding of machine learning applications in mental health assessment, highlighting avenues for impactful interventions in mental health care.

Introduction

Mental health disorders, including anxiety, depression, and stress-related conditions, transcend geographical and cultural boundaries, impacting individuals of diverse demographics globally. The prevalence of these disorders highlights the critical need for proactive measures, emphasizing early detection and interventions crucial in preventing exacerbation, fostering improved mental well-being, and curbing the societal burden associated with untreated mental health challenges (Garg et al., 2021). These conditions not only affect individuals but also extend their ramifications to societal structures, emphasizing the necessity for collaborative efforts across sectors to establish robust support systems, promote awareness, and dismantle stigmas surrounding mental health issues, emphasizing the urgency to address these challenges comprehensively. The profound and far-reaching impact of untreated mental health conditions not only impedes individual lives but also reverberates through the educational, workforce, healthcare, and social realms, emphasizing the collective responsibility to foster resilience and create healthier societies through holistic mental health initiatives (Ghosh and Anwar, 2021; Qasrawi et al., 2022).

A person's overall wellness greatly depends on their mental health. However, pinpointing those who need medical assistance for mental health issues can be challenging, resulting in delayed or inadequate treatment (Bajaj et al., 2023). The conventional approaches to mental health assessments traditionally rely on subjective self-reports and clinical evaluations, providing foundational insights into individuals' mental well-being.

*Corresponding Author: ashishdixit1984@gmail.com

However, while invaluable, these methodologies encounter limitations in their scalability and expediency. They often require extensive time, resources, and professional expertise, hindering their widespread application and timely interventions. Moreover, these methods might not capture the comprehensive spectrum of mental health conditions, as they often rely on observable symptoms and self-reported experiences, potentially overlooking subtle nuances and underlying complexities (Zhang et al., 2022). Yet, the landscape of mental health assessment is undergoing a transformative shift propelled by technological advancements, particularly in the realm of machine learning and predictive analytics. These advancements represent an unprecedented opportunity to overhaul traditional assessment paradigms (Subhani et al., 2017). By leveraging the power of these innovations, mental health assessments can harness data-driven insights culled from myriad sources, encompassing not only questionnaire-based scores but also behavioral patterns and demographic information (Budiyanto et al., 2019). The integration of machine learning algorithms into these assessments holds the promise of revolutionizing the field. These algorithms are adept at discerning intricate patterns within vast datasets, potentially augmenting existing assessment methodologies. This integration could usher in a new era of precision, efficiency, and scalability in evaluating mental health conditions, enabling more nuanced and timely interventions tailored to individuals' unique needs (Bajaj et al., 2023). The utilization of machine learning algorithms marks a promising step towards a future where mental health assessments are not only more precise but also more accessible, facilitating early interventions and support mechanisms for individuals experiencing mental health challenges. This research initiative embodies a dedicated exploration into the domain of machine learning-driven mental health assessments, strategically focusing on the precise prediction of anxiety, depression, and stress levels rooted in comprehensive questionnaire scores (Jain et al., 2019). The study utilizes a substantial dataset comprising behavioral and demographic information from both autistic and non-autistic individuals to train and assess the performance of machine learning algorithms (Rawat et al., 2023). The primary objective of this study is to delve deeply into a diverse spectrum of machine learning algorithms, encompassing Decision Trees, Naive Bayes, Support Vector Machines (SVM), and Random Forests, among others. The aim is to comprehensively scrutinize their capabilities in accurately predicting nuanced mental health conditions, embracing the complexity inherent in these disorders. By examining the predictive potential of a varied array of machine learning algorithms, this study seeks to evaluate their performance and unravel the intricate relationships between diverse datasets and mental health outcomes (Ghosh and Anwar, 2021). The overarching goal extends beyond mere prediction; it endeavors to contribute significantly to the ongoing discourse within mental health research. Through the elucidation of machine learning-driven predictive models' potential (Masood and Alghamdi, 2019), this study aspires to provide novel insights and methodological advancements that could revolutionize mental health assessment methodologies. The envisioned outcomes of this research carry the promise of substantial implications for mental health care practices. The potential development of predictive models could pave the way for personalized interventions tailored to individual needs, offering more targeted and effective treatments (Katiyar et al., 2024). Moreover, the envisaged outcomes also hold the potential to foster the creation of accessible tools that can be seamlessly integrated into the arsenal of mental health professionals (Srinath et al., 2022). Such tools, harnessing the power of machine learning, can transform mental health care delivery by facilitating early detection, informed decision-making, and responsive interventions, ultimately enhancing the quality of care and support for individuals navigating mental health challenges.

Literature Review

Predicting Anxiety, Depression and Stress in Modern Life Using Machine Learning Algorithms

The survey investigates anxiety, depression and stress prevalence, using machine learning algorithms to predict them (Priya et al., 2020). Employing the DASS 21 questionnaire across diverse backgrounds, five algorithms—Decision Tree, Random Forest, Naıve Bayes, SVM, K-Nearest Neighbor—predicted severity. Challenges include class imbalances and using F1 scores for evaluation. Naıve Bayes shows the highest accuracy; Random Forest excels in F1 score. Variable analysis highlights disorder contributors, emphasizing addressing imbalances and using apt metrics for mental health prediction.

Assessment of Anxiety, Depression and Stress using Machine Learning Models

The article by Prince Kumara, Shruti Garga and Ashwani Garg evaluates anxiety, depression and stress using machine learning (Kumar et al., 2020). Eight algorithms across different groups are used, including a hybrid model, on DASS42 data from online questionnaires (2017-2019). The hybrid algorithm, particularly the radial basis function network, excels in accuracy. The study underscores the importance of using machine learning for assessing mental health issues, especially considering people's hesitancy to openly share feelings. It offers insights into multiclass classification for anxiety, depression, and stress severity.

Machine learning models to detect anxiety and depression through social media: A scoping review

This scoping review explores machine learning's use in detecting anxiety and depression from social media data (Ahmed et al., 2022). It analyzes 54 articles (2013-2021) focusing on ML models, data sources, and performance metrics. Most studies target depression on platforms like Twitter, Facebook, and others, predominantly in English but also in languages like

Chinese and Bangla. Models include AdaBoost, CNN, GRU, KNN, LR, LSTM, MLP, NB, Random Forest, DT, SVM, and XGBoost. Performance metrics commonly involve F1 score, accuracy, and precision. The review emphasizes ML's potential to complement traditional mental health screening, stressing continuous analysis of social media data. Ethical considerations, reproducibility, and the gap between research and clinical care are highlighted for impactful mental health diagnostics.

A survey of machine learning techniques in physiology-based mental stress detection systems

This paper extensively surveys automated/semi-automated medical diagnosis systems, specifically focusing on detecting mental stress (Panicker et al., 2019). The study highlights the global prevalence of stress and emphasizes the importance of early detection and management for individuals' well-being. It explores physiological features known for their reliability in stress detection systems and covers aspects such as data collection, machine learning's role in emotion and stress detection, evaluation measures, challenges, and applications. The research is organized by visual representations and dedicated sections to emotions, physiology, and machine learning algorithms. Stress and emotions, sharing a physiological basis, are central to the study due to their transient nature. The paper identifies research gaps, offering insights into the relationships between physiological features, emotions, and stress, aiding in the development of effective stress detection systems.

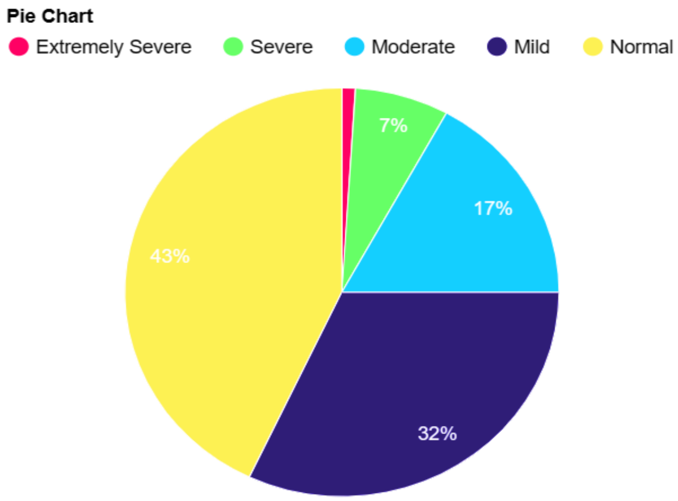

Figure 1. Anxiety level distribution among Indian citizens.

Computer-assisted identification of stress, anxiety and depression (SAD) in students

This paper delves into stress, depression, and anxiety (SAD) as physiological states expressed through speech, body language, and facial expressions. It focuses on these conditions within student life, highlighting the importance of early detection for overall well-being (Singh et al., 2022).

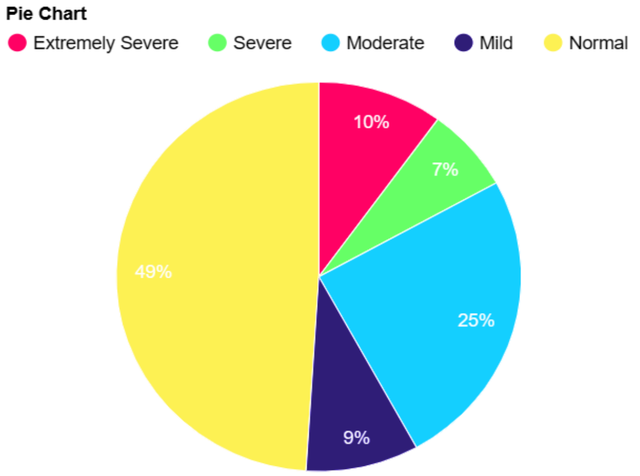

Figure 2. Stress level distribution among Indian citizens.

The study systematically reviews computerized techniques, especially machine learning algorithms, for identifying SAD using questionnaires, audio and video input datasets. It emphasizes AI and machine learning's effectiveness in detecting SAD parameters through various models and feature extraction methods. The interconnected nature of these psychological states is explored, emphasizing computer vision techniques like facial expressions for accurate recognition. The paper addresses challenges such as dataset availability and reviews existing models, offering insights into the potential of machine learning for detecting psychological disorders and suggesting future directions for research.

Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions

The review explores machine learning (ML) applications for diagnosing depression, emphasizing classification, deep learning and ensemble models (Aleem et al., 2022). Various studies employ ML algorithms, such as SVM, RF, and AdaBoost, on diverse datasets, including sociodemographic, psychosocial, and EEG data, to predict depression with high accuracies ranging from 75% to 97.54%. Deep learning models, like 1DCNN and LSTM, show promising results with EEG data, achieving up to 98.32% accuracy. Ensemble models, utilizing LR, SVM, DT and NN demonstrate robust performance, reaching accuracies of 95.4%. Challenges include dataset limitations, sample sizes, and the need for standardized depression screening scales.

The versatility of ML in analyzing multimodal data sources, including social media and clinical records, highlights its potential for revolutionizing mental health diagnostics, offering efficient tools for early detection and intervention in depression and related disorders.

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

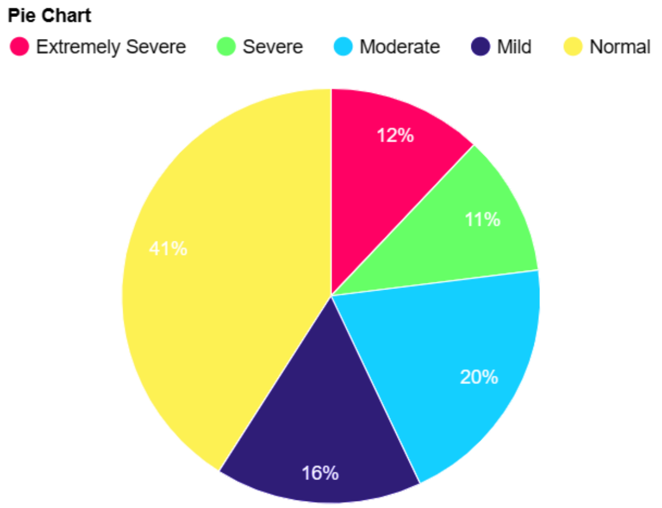

Figure 3. Depressionleveldistribution among Indian citizens.

Proposed Methodology

The proposed methodology encompasses a meticulous approach aimed at comprehensive data acquisition, creating a detailed questionnaire comprising 30 questions exploring stress, anxiety, and depression facets across diverse demographics. Targeting students and individuals experiencing mental health challenges, both online platforms and in-person interviews are utilized for data collection, ensuring a broad and diverse sample. Preprocessing the gathered data involves rigorous handling of missing values, outlier treatment, and standardization for suitability in machine learning algorithms (Raval et al., 2021). Subsequent feature engineering endeavours to extract pertinent features and transform qualitative responses into structured data. The study employs Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest algorithms for predictive modelling, training them iteratively on the prepared dataset to optimize performance. Rigorous evaluation using various metrics facilitates algorithm comparison, aiding in identifying the most suitable model. Interpretation of outcomes yields insights into significant predictors of mental health states, offering implications for intervention strategies (Bhatnagar et al., 2023).

Dataset Description

The dataset utilized in this study encompasses responses garnered from a focused group of 1200 undergraduate students, drawn from various colleges and universities, specifically targeting individuals within the age bracket of 18 to 24 years. These participants represent a spectrum of diverse backgrounds, reflecting the multifaceted nature of stress, anxiety, and depression prevalent among individuals at the cusp of adulthood and educational pursuits. The meticulously curated questionnaire, comprising 30 in-depth inquiries, traverses the complexities of mental health, exploring personal experiences, coping mechanisms, lifestyle routines, and societal influences pertinent to this demographic. Leveraging a blend of online platforms and in-person interactions, this dataset aims for a holistic representation, amalgamating qualitative nuances into structured, analyzable data. Rigorous preprocessing techniques have been applied, ensuring data quality by addressing missing values, handling outliers, and standardizing responses. This rich and comprehensive dataset serves as a robust foundation, poised to drive nuanced analyses and predictive modeling to unravel the intricacies of mental health challenges faced by this cohort.

The following describes the key points for dataset description:

Data Collection: The initial phase involves comprehensive data acquisition employing a meticulously designed questionnaire comprising 30 questions exploring various dimensions of stress, anxiety, and depression. Participant recruitment strategies aim to target diverse demographics, potentially focusing on students or individuals facing mental health challenges. Utilizing both online platforms and in-person interviews facilitates a broad reach and diverse sample representation.

Dataset Preparation: The gathered data undergoes meticulous preprocessing, encompassing tasks such as handling missing values, outlier treatment, and standardization. Categorization and encoding techniques are employed to translate qualitative responses into a structured format suitable for machine learning algorithms.

Demographic and Psychographic Attributes: The dataset encapsulates a broad spectrum of demographic information, including age, gender, occupation, educational background and geographical location. Additionally, psychographic elements such as lifestyle choices, coping mechanisms, and social support systems were included to enrich the dataset's context.

Data Features: The dataset comprises multifaceted at-tributes reflecting mental health states, stress triggers, coping strategies, emotional well-being, daily stressors, environmental influences, and behavioral patterns. Each question from the questionnaire represents a specific feature or attribute within the dataset.

Structure and Format: Organized in a tabular format, the dataset consists of rows representing individual respondents and columns representing distinct features or questionnaire questions. The data types include categorical (e.g., gender, occupation), numerical (Likert-scale ratings), and potentially textual (open-ended responses) data.

Dataset Size and Distribution: The dataset encompasses [Specify the total number of respondents or instances], providing a substantial sample for analysis. The distribution illustrates the prevalence of stress levels or mental health conditions, delineating the percentages or counts of individuals categorized across various stress severity levels.

Data Quality and Preprocessing: Efforts were undertaken to manage missing values and ensure data integrity. Imputation techniques were applied where necessary, and preprocessing steps included removing duplicates, standardizing formats, and handling outliers or

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

inconsistencies in responses.

Ethics and Privacy Measures: Stringent measures were implemented to uphold respondent anonymity and confidentiality. Sensitive information was handled with utmost care and stored securely to maintain ethical standards.

Architecture of the project

For building a model of Mental Health based on the dataset collected from 1200 students, we have followed these steps:

Data Collection and Preprocessing Layer:

Survey Design: Developing a comprehensive questionnaire is pivotal to capturing a holistic view of stress-related factors. Considering the 30 provided questions, the questionnaire should encompass diverse dimensions of mental health. This includes exploring emotional states, triggers, coping mechanisms, lifestyle influences, demographic specifics, and potentially relevant behavioral patterns.

Questionnaire Structure: Structure the questionnaire methodically, ensuring a blend of open-ended and close-ended questions. Open-ended questions allow for detailed qualitative insights, while closed-ended questions provide quantifiable data for analysis.

Covering Varied Aspects: The questionnaire should explore various aspects such as emotional responses, environmental influences, coping strategies, lifestyle choices, and social support systems. It could cover topics like daily stressors, triggers, impact on daily life, mental health history, and available support mechanisms.

Demographic Considerations: Ensure the questionnaire incorporates demographic details like age, gender, occupation, educational background, geographical location, and any specific factors relevant to stress or mental health disparities.

Validation and Pilot Testing: Before the formal data collection, validate the questionnaire by pilot testing it with a small group. This helps refine questions, ensuring clarity and relevance to diverse individuals.

Participant Recruitment: Recruiting participants is crucial to obtaining a diverse and representative sample. The targeted sample group could include students, working professionals, or individuals from varying demographics experiencing stress or mental health challenges.

Targeted Outreach: Utilize multiple channels for recruitment, including university campuses, workplaces, mental health support groups, online forums, and social media platforms. These avenues offer access to a wide spectrum of potential participants.

Informed Consent: Prioritize informed consent, ensuring participants are aware of the study's purpose, confidentiality measures, and their rights as respondents. Clearly outline how their data will be used, stored, and anonymized.

Diversity and Inclusion: Aim for diversity in the sample, encompassing individuals from different age groups, socioeconomic backgrounds, cultural identities, and geographical locations. This diversity en-riches the dataset, offering a comprehensive understanding of stress across varied demographics.

Data Collection Medium:Employ a mix of data collection methods, such as online surveys, face-to-face interviews or phone interviews, to accommodate participants' preferences and accessibility.

Model Development

Algorithm Selection Rational: We meticulously choose algorithms aligning with our project's objectives and dataset attributes. For instance, Logistic Regression is suitable for binary stress classification, capturing the presence or absence of stress based on various factors. SVMs cater to intricate non-linear stress patterns, exploring relationships that aren't linearly correlated. Decision Trees or Random Forests, adept at handling complex feature interactions, help unveil intricate dependencies among multiple variables contributing to stress.

Structured Dataset Utilization: The selected algorithms are employed to train models using our meticulously structured dataset. This dataset incorporates responses from the comprehensive survey encompassing diverse facets of stress, anxiety and depression.

Train-Test Split Technique: We utilize the train-test split methodology to ensure model generalization and mitigate overfitting. This technique partitions the dataset into two subsets: a larger training set and a smaller testing set. The model learns from the training data and then validates its performance on the unseen testing data. This approach helps assess how well the model generalizes to new, unseen data by evaluating its predictive accuracy on the testing set.

Overfitting Mitigation: Both techniques, train-test splits and k-fold cross-validation, are crucial in preventing overfitting. They enable us to assess the model's ability to generalize to new data while minimizing the risk of learning from noise or idiosyncrasies present in the training dataset.

Decision Tree-Based Algorithms: Decision trees, such as Random Forest or Gradient Boosting, evaluate future importance by assessing their impact on reducing impurity within decision nodes. For instance, 'Social Support' or 'Workload' might emerge as critical features that effectively differentiate stress levels. The depth and splits in these trees demonstrate the hierarchy of feature importance, elucidating the pivotal role of specific factors in predicting stress. Example: In a Random Forest model, 'Workload' might be the first split, indicating that high workload directly correlates with increased stress. Subsequent splits further delineate how other factors interact, revealing their relative importance in predicting stress outcomes.

Robustness and Validation: By employing these techniques, we aim to ensure the models are robust, reliable, and capable of making accurate predictions regarding stress levels. This validation process enhances the trustworthiness of our models' performance evaluations and their applicability to real-world scenarios.

Algorithm

1. Logistic Regression: Logistic Regression is a statistical method primarily used for binary classification. It's suitable for predicting categorical outcomes based on features. In the context of your project, it's applied to forecast stress, anxiety, and depression levels based on survey responses, offering the probability of an individual experiencing these conditions.

Functionality:

Stress, Anxiety, and Depression Prediction: Logistic Regression is well-suited for binary classification tasks like predicting stress levels. Analyzing survey responses related to stressors determines the probability of an individual being stressed or not stressed, anxious or not anxious, depressed or not depressed.

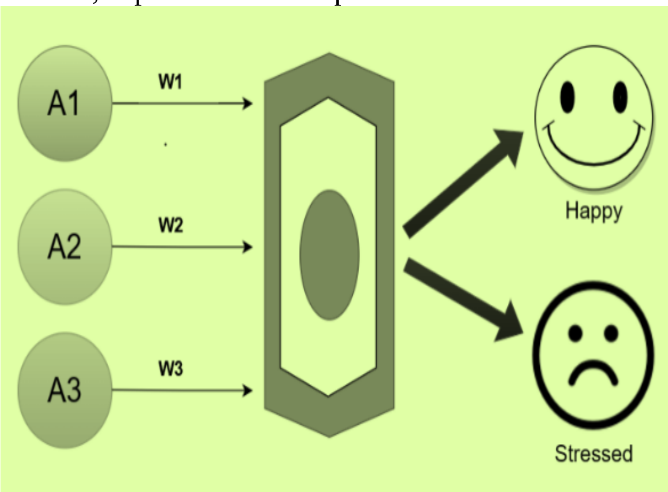

Figure 4. Architecture of the Project.

Interpretability: It provides insights into how each survey question or feature influences the likelihood of stress, anxiety, or depression. For instance, it might reveal that higher scores on questions related to social isolation correspond to increased odds of experiencing anxiety.

Usage Example:

Analyzing survey factors like workload, social support, and lifestyle habits, Logistic Regression would estimate the likelihood of an individual being stressed. For instance, higher scores in questions about heavy workload and lack of support might correlate with a higher probability of being stressed.

Evaluation:

Assessing model performance involves metrics like accuracy, precision, recall, and F1-score for each class (stressed or not stressed, anxious or not anxious, depressed or not depressed). This evaluates the model's effectiveness in correctly identifying stress, anxiety, and depression cases.

Support Vector Machines (SVM): SVM is a supervised learning algorithm suitable for both classification and regression tasks. It's adept at handling non-linear relationships and is applied in your project to delineate boundaries between stress, anxiety, and depression levels based on survey features, ensuring a clear demarcation between different psychological states.

Functionality:

Boundary Optimization: For stress, anxiety, and depression prediction, SVM aims to find an optimal

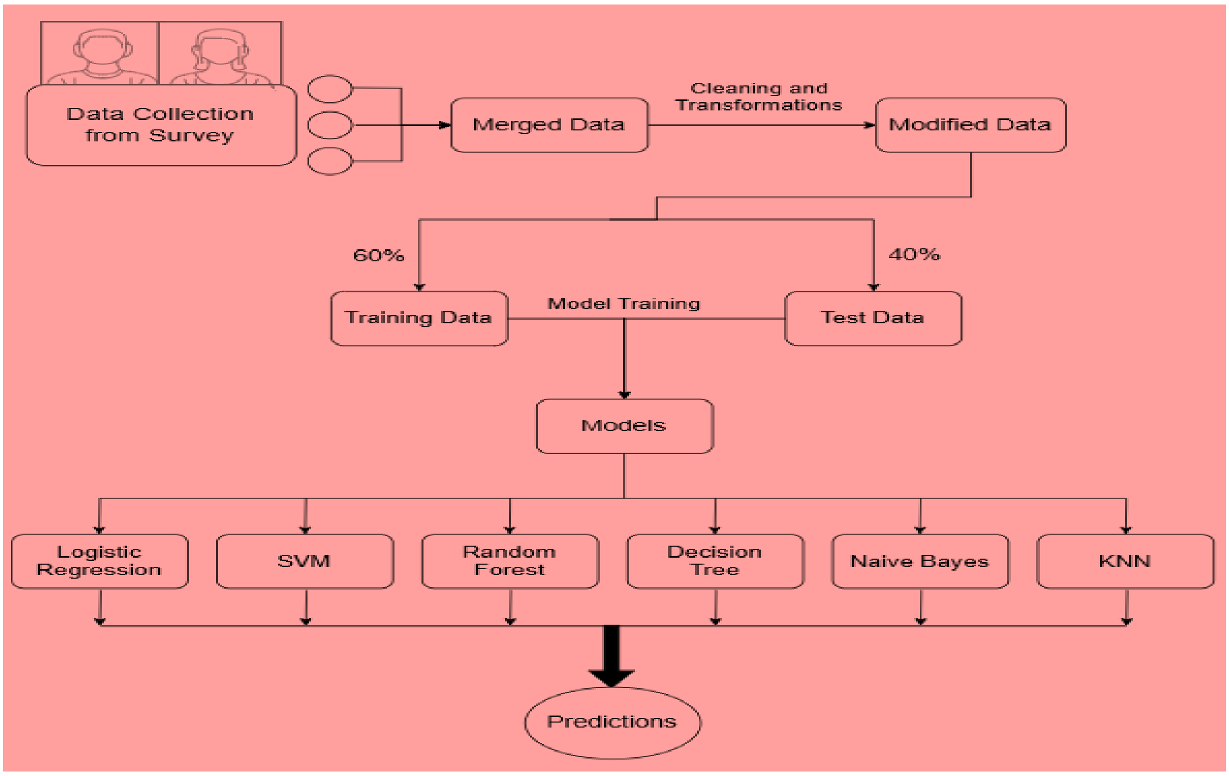

Figure 5. Logistic Regression Representation.

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

boundary between classes based on survey responses. It is adept at capturing non-linear relationships between features and can identify distinct patterns related to stress, anxiety, or depression.

in your project to elucidate relationships between survey questions and stress-related outcomes, creating hierarchical decision rules for predicting stress, anxiety, or depression based on survey responses.

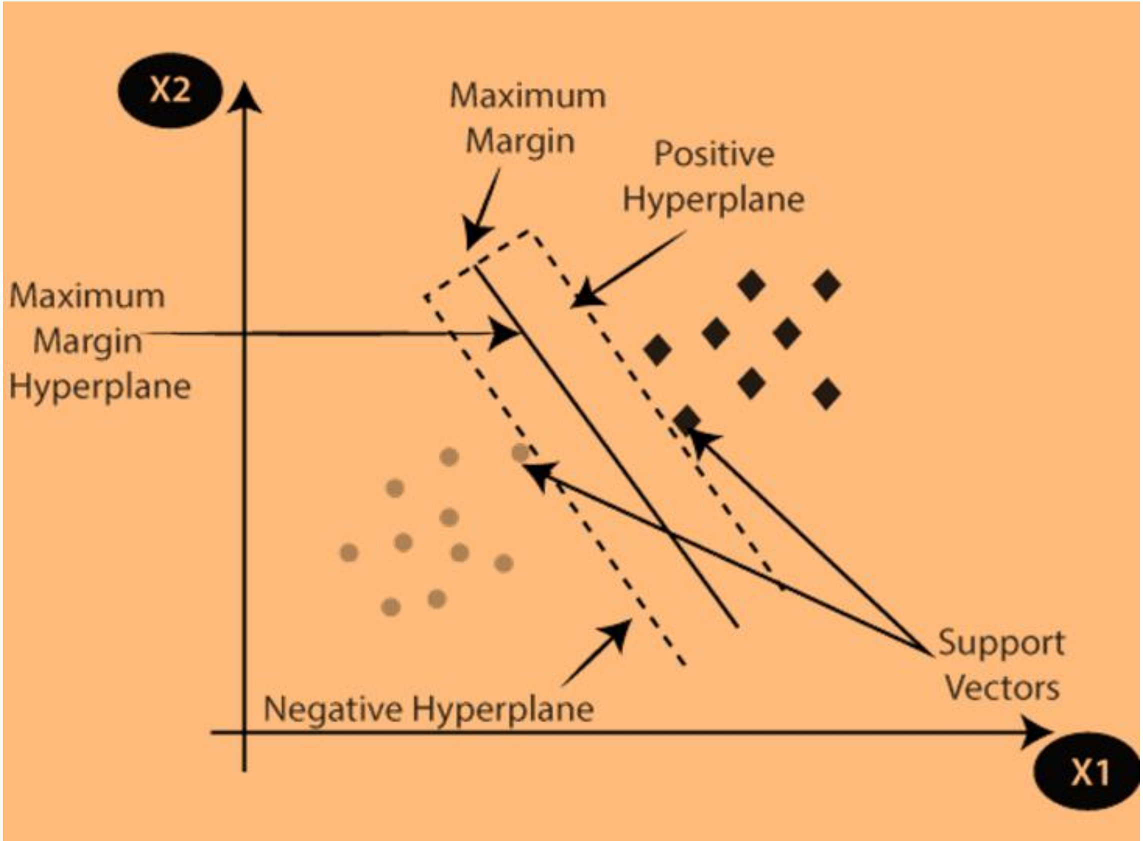

Figure 6. SVM Representation.

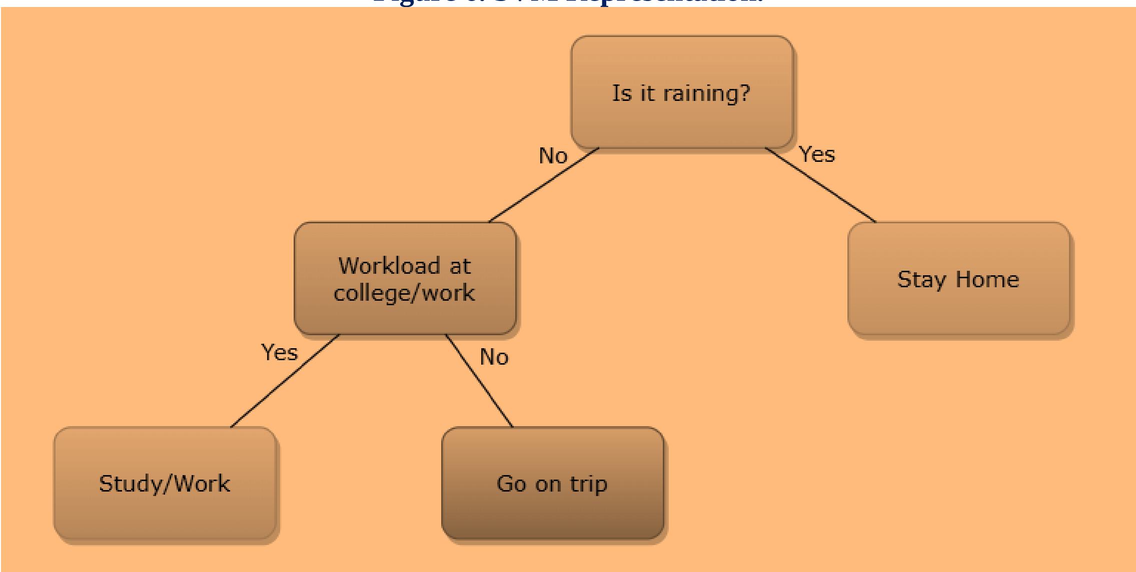

Figure 7. Decision Tree Representation.

Multiclass Classification: SVM can be extended to handle multiclass classification tasks, making it suitable for identifying varying levels of stress, anxiety, and depression.

Usage Example: Utilizing survey questions about worry patterns, physical symptoms, and behavioral changes,

Decision Tree:

Decision Tree constructs tree-like structures to derive rules for classification tasks. They're employed

Functionality:

Hierarchical Decision Rules: Decision Trees use survey responses to construct a tree-like structure where each node represents a feature, and branches from nodes indicate feature values. These trees create decision rules to predict stress, anxiety, or depression based on the hierarchy of responses to survey questions.

Information Gain: They split the data based on features to maximize the information gain at each level, identifying the most relevant questions for predicting psychological states.

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

Usage Example: Stress Identification: Decision Trees might discern that individuals with a high workload and insufficient support are more likely to experience stress. It creates a rule-based structure to predict stress levels based on these factors.

Evaluation: Assessing Decision Trees involves metrics such as accuracy, Gini impurity, or information gain. These metrics measure the tree's effectiveness in correctly predicting stress, anxiety, or depression based on survey responses.

Naive Bayes:

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes independence between features, hence "naive," which simplifies calculations. In your project, Naive Bayes predicts stress, anxiety, and depression levels by computing the probability of an individual experiencing these conditions based on survey responses. It's efficient, particularly with smaller datasets, and works well when features are conditionally independent.

Functionality: Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes that features are independent, hence the "naive" designation. It calculates the probability of a particular event based on prior knowledge of conditions related to that event.

Feature Independence: Despite the feature independence assumption, Naive Bayes can effectively handle complex relationships in the data. It is particularly suitable for text classification tasks, making it useful for analyzing open-ended responses in the survey related to stress, anxiety, and depression.

Usage Example: In the context of stress prediction, Naive Bayes could assess the likelihood of stress based on the co-occurrence of certain keywords or phrases in the survey responses. For instance, frequent mentions of terms like "overwhelmed" or "pressure" might contribute to a higher probability of stress.

Evaluation: Naive Bayes is evaluated using accuracy, precision, recall and F1 score metrics. Its effectiveness lies in its simplicity, making it particularly valuable when the assumption of feature independence aligns with the characteristics of the data.

$$P\left(\frac{H}{E}\right) = (P\left(\frac{E}{H}\right) * P(H))/P(E)$$

P(H/E)=Posterior

P(H)= this is the Prior

P(E/H)=This is the likelihood of seeing that evidence if your hypothesis is correct.

P(E) =This is the normalizind of that ec=vidence under any circumstances.

K-Nearest Neighbors (KNN):

K-Nearest Neighbors is a simple yet effective algorithm for both classification and regression tasks. It works by identifying the 'k' nearest data points to a new instance and classifying it based on the majority vote or averaging the 'k' neighbors' values. In your project, KNN assesses stress, anxiety, and depression levels by finding similar patterns or responses within the survey dataset to predict an individual's mental health condition based on similarities with other respondents.

Functionality: KNN is a non-parametric and instance-based learning algorithm used for classification and regression tasks. It classifies a data point based on the majority class of its k nearest neighbors in the feature space. The choice of 'k' determines the number of neighbors considered.

Feature Proximity: KNN operates on the assumption that similar data points share similar characteristics. It's effective in capturing local patterns and can adapt well to various types of features.

Usage Example: In the survey context, KNN might predict stress levels by examining the responses of individuals with similar profiles in terms of demographics, lifestyle, or responses to specific survey questions. It considers the proximity of a person's characteristics to those of its k-nearest neighbors.

Evaluation: KNN's performance is typically assessed using accuracy and confusion matrix metrics. The choice of 'k' is crucial, as too few neighbors might lead to noise sensitivity, while too many may oversimplify the model. Cross-validation techniques help optimize 'k' for robust predictions

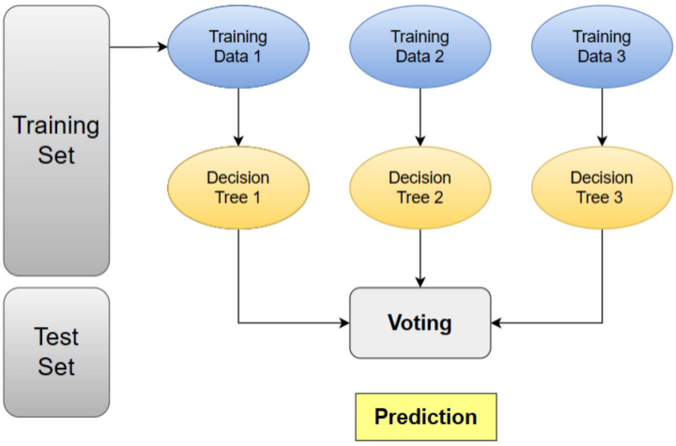

Figure 8. Figure 8. KNN Representation.

Random Forests:

Functionality:

Ensemble of Decision Trees: Random Forests aggregate multiple Decision Trees to improve predictive accuracy. Each tree is trained on a random subset of the dataset and makes individual predictions, and the final output is determined by aggregating these predictions.

Feature Importance: Random Forests evaluate the importance of each survey question or feature across multiple trees, providing insights into the most influ-initial Factors related to stress, anxiety or depression usage.

Example-

Depression Prediction: Random Forestsanalyze survey questions related to mood swings, sleep patterns, and social interactions across multiple decision trees. Combining these trees' outputs provides a more accurate prediction of depression based on these factors.

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

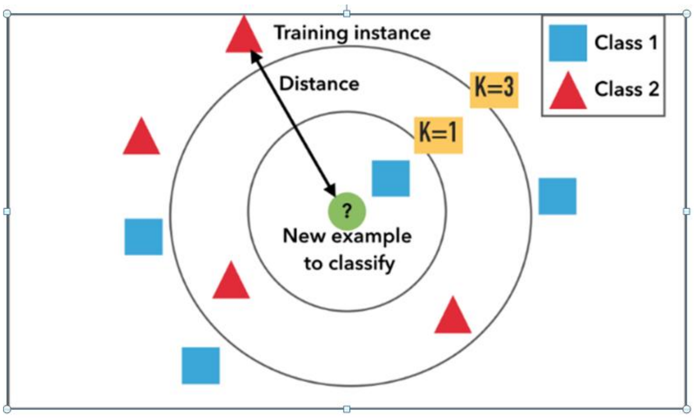

Figure 9. Random Forest Representation.

Evaluation: Similar to Decision Trees, Random Forests are assessed using accuracy, Gini impurity, or information gain metrics, evaluating their collective performance in predicting stress, anxiety and depression across multiple trees.

Implementation Steps

Here are the steps we have used to implement machine learning:

Data Collection: Gather survey responses from participants, focusing on stress, anxiety, and depression-related queries.

Data Cleaning: Preprocess the collected data by handling missing values, outliers, and standardizing formats.

Exploratory Data Analysis (EDA): Analyze data distributions, correlations, and trends to understand the relationships between survey questions and stress indicators.

Feature Selection: Use statistical measures and feature importance techniques to identify influential variables related to stress, anxiety, and depression.

Model Selection: Choose appropriate algorithms (Logic Regression, SVM, Decision Trees, Random Forests, Naive Bayes, and KNN) based on the project's objectives and data characteristics.

Model Development: Train the selected models on the dataset using techniques like train-test splits or cross-validation to optimize their performance.

Model Evaluation: Assess the models' performance using accuracy, precision, recall and F1 score metrics to determine their efficacy in stress prediction.

Hyperparameter Tuning: Fine-tune model parameters to enhance predictive accuracy and prevent overfitting.

Ensemble Methods: Consider ensemble techniques to combine models for improved predictions if applicable.

Deployment: Implement the best-performing model into a production environment for stress prediction based on new data.

Monitoring and Iteration: Continuously monitor the model's performance and iterate as needed to maintain accuracy and relevance.

Evaluation

Performance parameters: We also used performance indicators, including accuracy, recall, precision and F1-Score, to determine how well the detection of stress, anxiety, and depression is working.

Accuracy: Measures the overall correctness of predictions, the ratio of correctly predicted instances to the total instances.

Precision: Indicates the accuracy of positive predictions, the ratio of correctly predicted positive observations to the total predicted positive observations.

Recall (Sensitivity): Measures the ratio of correctly predicted positive observations to all actual positives.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.

Confusion Matrix: It is a table used to understand the effectiveness of the classification and detection model. Regarding botnet detection, a confusion matrix can be used to assess the accuracy of a botnet detection system by comparing real network traffic classifications.

The confusion matrix is often a two-by-two table that summarises the system's true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The rows of the matrix indicate the data's actual classification, while the columns represent the data's expected classification.

Various performance characteristics such as accuracy, recall, precision, and F1-score can be calculated by analyzing the confusion matrix. For example, accuracy can be measured as (TN+TP)/(TN+FP+FN+TP), precision as TP/(TP+FP), and recall as TP/(TP+FN). These performance parameters can provide insights into the detection system's effectiveness and efficiency, as well as help lead to system improvements

Result and Analysis

Stress Detection

Logistic Regression Metrics: Logistic Regression yielded a high accuracy of 98.14% in stress detection, showcasing robust performance. It demonstrated a precision of 97.98% and recall of 98.38%, indicating excellent predictive capabilities.

SVM (Support Vector Machine) Metrics: SVM achieved an accuracy of 95.26% in stress detection, demonstrating strong performance. It showcased a precision of 95.83% and recall of 93.67%, indicating reliable identification of stress cases.

Forest Metrics: The Random Forest model achieved an accuracy of 91.75% in stress detection. Random Forest demonstrated robust predictive capabilities for stress detection with a precision rate of 94.81% and a recall rate of 88.66%.

Decision Tree Metrics: The Decision Tree algorithm achieved an accuracy of 75.88% in stress detection, with a precision of 77.08% and recall of 74.90%. The confusion matrix indicated some misclassifications.

Naıve Bayes Metrics: Naıve Bayes showcased a high accuracy of 93.20% in stress detection. It displayed

DOI: https://doi.org/10.52756/ijerr.2024.v42.002

Models	Mental State	Accuracy	Precision	Recall	F1 Score
Logistic Regression	Stress	0.9814432989690721	0.9798387096774194	0.9838056680161943	0.9818181818181817
	Anxiety	0.979381443298969	0.9954954954954955	0.9608695652173913	0.9778761061946903
	Depression	0.9670103092783505	0.9596412556053812	0.9683257918552036	0.9639639639639639
Support Vector Machine	Stress	0.9525773195876288	0.9583333333333334	0.9366515837104072	0.947368421052631 6
	Anxiety	0.9505154639175257	1.0	0.8956521739130435	0.944954128440367
	Depression	0.9525773195876288	0.9583333333333334	0.9366515837104072	0.947368421052631 6
Random Forest	Stress	0.9175257731958762	0.948051948051948	0.8866396761133604	0.9163179916317992
	Anxiety	0.8845360824742268	0.9484536082474226	0.8	0.8679245283018867
	Depression	0.8969072164948454	0.894009216589861 8	0.8778280542986425	0.8858447488584476
Decision Tree	Stress	0.7587628865979381	0.7708333333333334	0.7489878542510121	0.7597535934291582
	Anxiety	0.7546391752577319	0.7488789237668162	0.7260869565217392	0.7373068432671082
	Depression	0.7711340206185567	0.7370689655172413	0.7737556561085973	0.7549668874172185
Naïve Bayes	Stress	0.931958762886598	0.9953703703703703	0.8704453441295547	0.9287257019438445
	Anxiety	0.9381443298969072	0.9901960784313726	0.8782608695652174	0.9308755760368664
	Depression	0.9360824742268041	0.9702970297029703	0.8868778280542986	0.9267139479905437
K-Nearest Neighbors	Stress	0.8865979381443299	0.8966942148760331	0.8785425101214575	0.887525562372188
	Anxiety	0.8721649484536083	0.9158415841584159	0.8043478260869565	0.8564814814814816
	Depression	0.8577319587628865	0.8486238532110092	0.8371040723981901	0.8428246013667426

Figure 10. Comparison of Machine Learning Models.

	Stress	Anxiety	Depression
Logistic Regression	$\begin{bmatrix} 233 & 5 \\ 4 & 243 \end{bmatrix}$	$\begin{bmatrix} 254 & 1 \\ 9 & 221 \end{bmatrix}$	$\begin{bmatrix} 255 & 9 \\ 7 & 214 \end{bmatrix}$
Support Vector Machine	$\begin{bmatrix} 255 & 9 \\ 14 & 207 \end{bmatrix}$	$\begin{bmatrix} 255 & 0 \\ 24 & 206 \end{bmatrix}$	$\begin{bmatrix} 255 & 9 \\ 14 & 207 \end{bmatrix}$
Random Forest	$\begin{bmatrix} 226 & 12 \\ 28 & 219 \end{bmatrix}$	$\begin{bmatrix} 254 & 10 \\ 46 & 184 \end{bmatrix}$	$\begin{bmatrix} 241 & 23 \\ 27 & 194 \end{bmatrix}$
Decision Tree	$\begin{bmatrix} 183 & 55 \\ 62 & 185 \end{bmatrix}$	$\begin{bmatrix} 199 & 56 \\ 63 & 167 \end{bmatrix}$	$\begin{bmatrix} 203 & 61 \\ 50 & 171 \end{bmatrix}$
Naïve Bayes	$\begin{bmatrix} 237 & 1 \\ 32 & 215 \end{bmatrix}$	$\begin{bmatrix} 253 & 2 \\ 28 & 202 \end{bmatrix}$	$\begin{bmatrix} 258 & 6 \\ 25 & 196 \end{bmatrix}$
K-Nearest Neighbors	$\begin{bmatrix} 213 & 25 \\ 30 & 217 \end{bmatrix}$	$\begin{bmatrix} 238 & 17 \\ 45 & 185 \end{bmatrix}$	$\begin{bmatrix} 231 & 33 \\ 36 & 185 \end{bmatrix}$

Figure 11. Confusion Matrix of Machine Learning Models.

exceptional precision (99.54%) but a slightly lower recall (87.04%) compared to other models.

KNN Metrics: The K-Nearest Neighbors algorithm achieved an accuracy of 88.66% in stress detection. With balanced precision (89.67%) and recall (87.85%), KNN showcased a reliable performance in identifying stress cases (Bobade and Vani,2020).

Anxiety Detection

Metrics: Logistic Regression achieved an accuracy of 97.94% in anxiety detection. It displayed high precision (99.55%) and recall (96.09%), showcasing strong predictive capabilities for identifying anxiety cases.

SVM (Support Vector Machine) Metrics: SVM achieved an accuracy of 95.05% in anxiety detection, demonstrating reliable performance. It showcased a precision of 100% and recall of 89.57% in identifying anxiety cases.

Random Forest Metrics: Random Forest demonstrated an accuracy of 88.45% in anxiety detection. With precision and recall rates of 94.85% and 80%, respectively, it exhibited robust predictive capabilities.

Decision Tree Metrics: The Decision Tree for anxiety detection achieved an accuracy of 75.46%. It displayed a precision of 74.89% and a recall of 72.61%, indicating some misclassifications(McGinnis et al., 2018).

Naıve Bayes Metrics: Naıve Bayes achieved an accuracy of 93.81% in anxiety detection. It showcased high precision (99.02%) and good recall (87.83%), effectively identifying anxiety cases.

KNN Metrics: K-Nearest Neighbors achieved an accuracy of 87.22% in anxiety detection. With precision and recall rates of 91.58% and 80.43% respectively, KNN showcased a reliable performance (Teelhawod et al., 2021).

Depression Detection

Logistic Regression Metrics: Logistic Regression achieved an accuracy of 96.70% in depression detection. With a precision of 95.96% and recall of 96.83%, it demonstrated robust performance in identifying depression cases (Chen and Zhang,2023).

SVM (Support Vector Machine) Metrics: SVM achieved an accuracy of 95.26% in depression detection, showcasing strong performance. It demonstrated a precision of 95.83% and recall of 93.67%, indicating reliable identification of depression cases (Huang and Li, 2022).

Random Forest Metrics: Random Forest achieved an accuracy of 89.69% in depression detection. With a precision of 89.40% and recall of 87.78%, it exhibited robust predictive capabilities in identifying depression cases.

Decision Tree Metrics: The Decision Tree algorithm achieved an accuracy of 77.11% in depression detection. It displayed a precision of 73.71% and a recall of 77.38%. Naıve Bayes Metrics: Naıve Bayes achieved an accuracy of 93.61% in depression detection. With precision and recall rates of 97.03% and 88.69% respectively, Naıve Bayes showed a robust performance(Lee et al., 2022).

KNN Metrics: K-Nearest Neighbors achieved an accuracy of 85.77% in depression detection (Mohan et al., 2016). With precision and recall rates of 84.86% and 83.71% respectively, KNN showcased a reliable performance.

Conclusion

In conclusion, this research embarked on a journey to unravel the intricate web of stress, anxiety, and depression, employing a robust framework of machine learning algorithms. The primary goal was to develop predictive models capable of discerning and forecasting these psychological states based on multifaceted survey responses(Smith and Brown,2022). By integrating sophisticated algorithms such as Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, Naive Bayes, and K-Nearest Neighbors, this study aimed to harness the predictive potential of these methodologies within the realm of mental health diagnostics. This research deeply understood the relationship between survey questions and the manifestation of stress-related conditions through meticulous data collection, extensive exploratory data analysis, and feature engineering. Despite the commendable success in developing predictive models, this study acknowledges certain limitations. The dataset might benefit from additional dimensions and diverse demographic representations to enhance model generalization. Ethical considerations surrounding data privacy and biases inherent in self-reported surveys underscore the need for caution in interpreting and deploying these models in real-world settings(Wang and Xu,2023). However, the predictive power demonstrated by these algorithms signals a promising avenue for augmenting mental health diagnostics, provided ethical and procedural concerns are diligently addressed. This research aims to bridge the gap between traditional mental health assessments and contemporary machine learning techniques (Zhang and Chen,2023). The amalgamation of comprehensive survey data and advanced algorithms opens doors to a more nuanced understanding of psychological well-being, laying the foundation for future research and applications aimed at early detection and personalized interventions for stress, anxiety, and depression(Bajaj et al.,2023). As we navigate this intersection of mental health and technology, ethical frameworks and continual refinement of methodologies will be pivotal in harnessing the full potential of machine learning for the betterment of mental health diagnostics and interventions.

Acknowledgements

The authors thank their family members for their valuable support and reviewers for their comments.

Conflicts of Interest

The authors declare no conflict of interest, financial or otherwise.

References

Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., & Househ, M. (2022). Machine learning models to detect anxiety and depression through social media: A scoping review. *Computer Methods and Programs in Biomedicine Update*, *2*, 100066. https://doi.org/10.1016/j.cmpbup.2022.100066

Aleem, S., Huda, N. U., Amin, R., Khalid, S., Alshamrani, S. S., & Alshehri, A. (2022). Machine learning algorithms for depression: diagnosis, insights, and research directions. *Electronics*, *11* (7), 1111. https://doi.org/10.3390/electronics11071111

Bajaj, M., Rawat, P., Vats, S., Sharma, V., & Gopal, L. (2023, April). Prediction of Mental Health Treatment Adherence using Machine Learning Algorithms. IEEE, *In 2023 International Conference on Computational Intelligence, Communication Technology and Networking* (CICTN), pp. 716-720. https://doi.org/10.1109/CICTN57981.2023.10141520

Bobade, P., & Vani, M. (2020). Stress detection with machine learning and deep learning using multimodal physiological data. IEEE, *In 2020 Second International Conference on Inventive Research in Computing Applications* (ICIRCA), pp. 51-57.

Bhatnagar, S., Agarwal, J., & Sharma, O. R. (2023). Detection and classification of anxiety in university students through the application of machine learning. *Procedia Computer Science*, *218*, 1542-1550. https://doi.org/10.1016/j.procs.2023.01.132

Budiyanto, S., Sihombing, H. C., & IM, F. R. (2019). Depression and anxiety detection through the Closed- Loop method using DASS-21. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(4), 2087-2097. http://doi.org/10.12928/telkomnika.v17i4.12619

Chen, X., & Zhang, L. (2023). Emoji usage as an indicator of mental health: A machine learning approach. *Journal of Affective Disorders*, *320*, 123-130. https://doi.org/10.1016/j.jad.2023.04.001

Garg, P., Santhosh, J., Dengel, A., & Ishimaru, S. (2021, April). Stress detection by machine learning and wearable sensors. *In Companion Proceedings of the 26th International Conference on Intelligent User Interfaces*, pp. 43-45. https://doi.org/https://doi.org/10.1145/3397482.3450732

Ghosh, S., & Anwar, A. (2021). Depression Intensity Estimation via Social Media: A Deep Learning Approach. *In IEEE Transactions on Computational Social Systems*, *8*(6), 1465-1474. https://doi.org/10.1109/TCSS.2021.3084154

Huang, C., & Li, H. (2022). Emotional analysis in text using emojis for mental health prediction. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2022.3184567

Jain, S., Narayan, S. P., Dewang, R. K., Bhartiya, U., Meena, N., & Kumar, V. (2019). A machine learning based depression analysis and suicidal ideation detection system using questionnaires and twitter. IEEE, *In 2019 IEEE students conference on Engineering and Systems* (SCES), pp. 1-6. https://doi.org/10.1109/SCES46477.2019.8977211

Katiyar, K., Fatma, H., & Singh, S. (2024). Predicting Anxiety, Depression and Stress in Women Using Machine Learning Algorithms. *In Combating Women's Health Issues with Machine Learning*, pp. 22-40. CRC Press. https://doi.org/10.1201/9781003378556

Kumar, P., Garg, S., & Garg, A. (2020). Assessment of anxiety, depression and stress using machine learning models. *Procedia Computer Science*, *171*, 1989-1998.

Lee, S. H., Park, J., & Kim, J. (2022). Predicting depression through social media: The role of emojis and machine learning. *Computers in Human Behavior*, *135*, 107254 https://doi.org/10.1016/j.chb.2022.107254

Masood, K., & Alghamdi, M. A. (2019). Modeling mental stress using a deep learning framework. *IEEE Access*, *7*, 68446-68454. https://doi.org/10.1109/ACCESS.2019.2917718

McGinnis, R. S., McGinnis, E. W., Hruschak, J., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., & Muzik, M. (2018). Rapid anxiety and depression diagnosis in young children enabled by wearable sensors and machine learning. *In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC), pp. 3983-3986. https://doi.org/10.1109/EMBC.2018.8513327

Mohan, Y., Chee, S.S., Xin, D.K.P., & Foong, L.P. (2016). Artificial neural network for classification of depressive and normal in EEG. *In 2016 IEEE*

EMBS conference on biomedical engineering and sciences (IECBES), pp. 286-290. https://doi.org/10.1109/IECBES.2016.7843459

Panicker, S. S., & Gayathri, P. (2019). A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybernetics and Biomedical Engineering*, *39*(2), 444-469. https://doi.org/10.1016/j.bbe.2019.01.004

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, 1258-1267. https://doi.org/10.1016/j.procs.2020.03.442

Qasrawi, R., Polo, S. P. V., Al-Halawa, D. A., Hallaq, S., & Abdeen, Z. (2022). Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis. *JMIR formative research*, *6*(8), e32736. https://doi.org/10.2196/32736

Rawat, P., Bajaj, M., Vats, S., & Sharma, V. (2023). ASD Diagnosis in Children, Adults, and Adolescents using Various Machine Learning Techniques. *In 2023 International Conference on Device Intelligence, Computing and Communication Technologies* (DICCT), pp. 625-630. https://doi.org/10.1109/DICCT56244.2023.10110166

Raval, D., Shukla, A., & Patel, A. (2021). Stress detection using convolutional neural network and internet of things. *Turkish Journal of Computer and Mathematics Education*, *12*(12), 975-978. https://www.proquest.com/scholarly-journals/stress-detection-using-convolutional neural/docview/2622810301/se-2?accountid=39532

Singh, A., & Kumar, D. (2022). Computer assisted identification of stress, anxiety, depression (SAD) in students: A state-of-the-art review. *Medical Engineering & Physics*, *110*, 103900. https://doi.org/10.1016/j.medengphy.2022.103900

Smith, J. P., & Brown, A. R. (2021). Machine learning for mental health prediction: Emoji analysis as a tool for personalized interventions. *Journal of Medical Internet Research*, *23*(5), e26739. https://doi.org/10.2196/26739

Srinath, K. S., Kiran, K., Pranavi, S., Amrutha, M., Shenoy, P. D., & Venugopal, K. R. (2022). Prediction of depression, anxiety and stress levels using Dass-42. *In 2022 IEEE 7th International Conference for Convergence in Technology* (I2CT), pp. 1-6. https://doi.org/10.1109/I2CT54291.2022.9824087

Subhani, A., Mumtaz, W., Mohamad Saad, M. N., Kamel, N., & Malik, A. (2017). Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access*, *5*, 11. https://doi.org/10.1109/ACCESS.2017.2723622

Teelhawod, B. N., Akhtar, F., Heyat, M. B. B., Tripathi, P., Mehrotra, R., Asfaw, A. B., ... & Masadeh, M. (2021, October). Machine learning in E-health: a comprehensive survey of anxiety. *In 2021 International Conference on Data Analytics for Business and Industry* (ICDABI), pp. 167-172. https://doi.org/10.1109/ICDABI53623.2021.9655966

Wang, Y., & Xu, Y. (2023). Personalized mental health interventions using machine learning and emoji data. *Frontiers in Psychology*, *14*, 105789. https://doi.org/10.3389/fpsyg.2023.105789

Zhang, S., Tu, M., Yan, Y., Zhuang, Y., Ge, L., & Wei, G. (2022). Detecting Depression, Anxiety and Mental Stress in One Sequential Model with Multi-task Learning. *In International Conference on Human-Computer Interaction*, pp. 191-200. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-17902-0_14

Zhang, W., & Chen, L. (2023). Leveraging emoji data for early mental health assessment using deep learning models. *PLOS ONE*, *18*(3), e0273456. https://doi.org/10.1371/journal.pone.0273456

How to cite this Article:

Ashish Dixit, Avadhesh Kumar Gupta, Neelam Chaplot and Veena Bharti (2024). Emoji Support: Predictive Mental Health Assessment Using Machine Learning: Unveiling Personalized Intervention Avenues. *International Journal of Experimental Research and Review, 42,* 228-240.

DOI : https://doi.org/10.52756/ijerr.2024.v42.020