

Predictive Analytics for Workplace Stress Assessment Using Ensemble Machine Learning

Dr. Sk. Moulali¹, Dr. M. Sree Ram Kiran Nag²

¹Professor, ²Associate Professor, ^{1,2}Department of Data Science, NRI Institute of Technology, Gundlapalem Village, Visadala Post, MEDikonduru Mandal, Guntur District, Andhra Pradesh -522438, India

Abstract— Employee stress has become a major concern in modern organizations due to its impact on employee productivity, mental well-being, job satisfaction, and organizational performance. Factors such as increasing workloads, tight deadlines, long working hours, and workplace pressure contribute significantly to employee stress. Traditional stress assessment methods, including manual surveys and human resource (HR) evaluations, are often subjective, time-consuming, and difficult to scale across large organizations. This paper presents a predictive analytics approach for workplace stress assessment using an Ensemble Voting Classifier. The proposed system analyzes more than 18 workplace-related and personal parameters to classify employees into two categories: *Low Stress* and *More Stress*. The ensemble model integrates six machine learning algorithms: Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Stochastic Gradient Descent (SGD) Classifier. By combining predictions from multiple classifiers, the approach aims to improve prediction reliability and classification performance. The system is implemented as a full-stack web application using the Django framework and MySQL database. Employee data is collected through a web interface, preprocessed using appropriate feature encoding techniques, and analyzed by the trained ensemble model. An administrative dashboard provides visualizations of organizational stress patterns through charts and analytical reports. The experimental evaluation compares the ensemble model with individual classifiers using classification performance metrics. The reported results indicate that the ensemble approach achieves approximately 88–93% prediction accuracy, demonstrating its potential for data-driven workplace stress assessment and employee wellness monitoring.

Keywords— Predictive Analytics, Workplace Stress Assessment, Employee Stress Prediction, Ensemble Machine Learning, Voting Classifier, Occupational Health, Employee Wellness, Data Preprocessing, Classification Algorithms, Human Resource Analytics, Support Vector Machine, Logistic Regression, Decision Tree, Naive Bayes, K-Nearest Neighbors, Django, MySQL.

I. INTRODUCTION

In today's competitive and fast-paced work environment, employee stress has become a serious organizational and social issue. Employees are frequently exposed to heavy workloads, strict deadlines, long working hours, performance pressure, and complex workplace interactions. These factors significantly affect mental well-being, productivity, decision-making ability, and overall job satisfaction. Continuous exposure to workplace stress may eventually lead to burnout, decreased efficiency, absenteeism, and high employee turnover, negatively impacting both employees and organizations.

Traditional methods of identifying employee stress mainly depend on manual surveys, interviews, counseling sessions, and human resource evaluations. Although these methods provide some insights, they are often subjective, time-consuming, and difficult to scale in large organizations. Employees may also hesitate to disclose their actual stress levels due to fear of judgment, professional consequences, or social stigma. As a result, many organizations fail to identify stress-related problems at an early stage.

Recent advancements in Machine Learning (ML) and data analytics have created new opportunities for automated stress prediction systems. Machine learning algorithms can analyze large volumes of employee-related data and identify hidden patterns associated with stress levels. By examining multiple workplace and personal factors simultaneously, ML models can provide accurate and objective predictions that support proactive decision-making in organizations. Ensemble learning techniques further enhance prediction performance by combining the strengths of multiple classifiers into a single robust model.

This project proposes a Machine Learning Based Employee Stress Prediction System that uses an Ensemble Voting Classifier to classify employees into two categories: *Low Stress* and *More Stress*. The system considers more than 18 workplace and personal parameters such as workload, working hours, performance ratings, monthly income, work satisfaction, and lifestyle-related factors. The ensemble model integrates six machine learning algorithms including Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest

Neighbors (KNN), Decision Tree, and SGD Classifier to achieve improved prediction accuracy and reliability.

The proposed system is implemented as a web-based application using the Django framework and MySQL database. Employees can securely submit their information through an interactive interface, while administrators can monitor organizational stress trends through analytical dashboards and visual reports. The system aims to provide organizations with a scalable, efficient, and cost-effective solution for early stress detection and employee wellness management.

I. REVIEW LITERATURE SURVEY

Employee stress prediction has gained significant attention in recent years due to its importance in maintaining employee well-being and organizational productivity. Researchers have explored different approaches ranging from physiological signal analysis to machine learning-based behavioral assessment systems. Several studies have demonstrated that machine learning techniques can effectively identify stress patterns by analyzing workplace, behavioral, and physiological factors.

One of the important studies in stress recognition was conducted by Neha Sharma and Tanveer Gedeon (2012), who presented a comprehensive survey on objective stress recognition methods using physiological, psychological, and behavioral measurements. Their work analyzed various stress indicators such as Electroencephalography (EEG), Galvanic Skin Response (GSR), Heart Rate Variability (HRV), speech analysis, and facial expressions. The study highlighted that combining multiple feature types significantly improves stress classification accuracy compared to using a single modality. However, the major drawback of this approach was the dependency on specialized biometric sensors and wearable devices, making large-scale organizational deployment expensive and difficult.

Another notable contribution was made by Martin Gjoreski et al. (2017), who investigated stress monitoring using wearable devices such as smartwatches and fitness trackers. Their research utilized machine learning algorithms including Support Vector Machine (SVM) and Random Forest classifiers to analyze physiological signals like heart rate, skin temperature, and motion data. The study demonstrated that ensemble learning methods achieve better performance than individual classifiers, with accuracy levels reaching approximately 87–91%. Although the

system achieved high accuracy, the continuous use of wearable devices introduced privacy concerns, sensor maintenance challenges, and user discomfort during prolonged usage.

Ane Alberdi et al. (2016) proposed an automatic stress recognition system for office environments using computer interaction patterns such as keyboard usage, mouse movement, and application switching behavior. Their work proved that workplace behavioral parameters and occupational data are strong indicators of employee stress levels. The study achieved considerable prediction accuracy using non-physiological features alone, confirming that stress can be effectively identified without expensive biometric hardware. However, continuous monitoring of employee computer activities raised ethical and privacy concerns related to workplace surveillance.

Recent studies have also explored machine learning approaches for employee stress and burnout prediction using structured organizational datasets. Researchers have implemented algorithms such as Logistic Regression, Decision Trees, Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines to classify employee stress levels based on workplace attributes including workload, working hours, job satisfaction, and performance ratings. Ensemble learning methods, especially Voting Classifiers, have shown improved prediction accuracy because they combine the strengths of multiple classifiers while reducing individual model limitations.

Despite the advancements in existing systems, several limitations remain unresolved. Many traditional systems rely heavily on manual HR surveys and periodic assessments, which are subjective and incapable of providing real-time monitoring. Sensor-based approaches require expensive hardware and often face scalability issues in large organizations. Additionally, some systems focus only on limited parameters and fail to analyze the complex relationships between multiple workplace and personal factors simultaneously.

To overcome these limitations, the proposed system introduces a Machine Learning Based Employee Stress Prediction System using an Ensemble Voting Classifier. The system analyzes more than 18 employee-related parameters and integrates six machine learning algorithms—Naive Bayes, SVM, Logistic Regression, KNN, Decision Tree, and SGD Classifier—to improve prediction accuracy and reliability. Unlike wearable-based approaches, the

proposed model relies on structured workplace and personal data, making it scalable, cost-effective, and suitable for practical organizational deployment.

II. RESEARCH METHODOLOGY

The proposed research methodology focuses on developing an automated Employee Stress Prediction System using Machine Learning techniques to identify stress levels among employees based on workplace and personal attributes. The methodology consists of multiple stages including data collection, data preprocessing, feature engineering, machine learning model development, ensemble integration, system implementation, and performance evaluation. The overall objective is to build a reliable and scalable system capable of predicting employee stress accurately and efficiently.

The first stage of the methodology involves data collection from structured employee datasets containing more than 18 workplace and personal parameters. These attributes include demographic information, department, education level, working hours, workload, performance ratings, monthly income, salary hike, work satisfaction, marital status, training time, and years at the company. The collected data represents multiple dimensions of employee behavior and workplace conditions that contribute to stress levels. The dataset is stored and managed using a MySQL relational database for secure and persistent data handling.

In the second stage, data preprocessing and feature engineering techniques are applied to prepare the dataset for machine learning analysis. Since machine learning algorithms require numerical input, categorical variables such as gender, department, region, and marital status are converted into numerical form using Label Encoding techniques. Missing values, invalid inputs, and inconsistent data entries are handled through validation and cleaning processes. Numerical features are normalized and organized into structured feature vectors to improve the efficiency and accuracy of the prediction models.

The third stage involves the development and training of multiple machine learning classifiers. Six different algorithms are used in the system: Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and SGD Classifier. Each algorithm independently learns patterns from the training dataset and predicts whether an employee belongs to the *Low Stress* or *More Stress* category. The dataset is divided into training and testing subsets using standard train-test split methods to evaluate the performance of each model on unseen data.

10.48047/jocaaa.2024.33.08.627

To improve prediction reliability and reduce individual model limitations, an Ensemble Voting Classifier is implemented in the fourth stage. The ensemble model combines predictions from all six classifiers using a hard voting mechanism. Each classifier generates its own binary prediction, and the final stress classification is determined based on the majority vote. This ensemble approach increases overall accuracy, minimizes overfitting, and improves robustness by leveraging the strengths of multiple algorithms instead of depending on a single model.

The fifth stage focuses on system implementation using the Django web framework. A secure and interactive web application is developed to allow employees to submit their stress-related information through online forms. The backend processes the submitted data, performs preprocessing, and passes the feature vector to the ensemble model for prediction. The prediction results are stored in the database and displayed to users through color-coded interfaces indicating stress levels. An administrative dashboard is also developed to provide visualization of organizational stress trends using charts, graphs, and summary reports.

Finally, the system is evaluated using various performance metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. Individual classifier performances are compared with the ensemble model to measure improvement in prediction capability. Experimental results demonstrate that the Voting Ensemble Classifier achieves higher accuracy, approximately 88–93%, compared to individual algorithms. The evaluation confirms that the proposed methodology provides an effective, scalable, and practical solution for early employee stress prediction and organizational wellness management.

III. EXISTING SYSTEM

The existing employee stress assessment systems mainly rely on traditional manual methods such as employee surveys, HR interviews, counseling sessions, and self-assessment questionnaires. In most organizations, Human Resource (HR) departments periodically collect employee feedback to identify stress-related issues and workplace dissatisfaction. Managers and HR professionals analyze employee behavior, absenteeism, work performance, and survey responses to determine whether employees are experiencing stress.

Although these traditional methods provide basic insights into employee well-being, they suffer from several limitations. One major drawback is subjectivity, as employees may hesitate to reveal their actual stress levels because of fear of negative judgment, professional consequences, or social stigma. As a

result, many employees underreport stress symptoms, leading to inaccurate assessments and delayed intervention.

Another limitation of existing systems is the lack of real-time monitoring. Most organizations conduct stress evaluations monthly or quarterly, which creates significant gaps in identifying sudden changes in employee mental health. Workplace stress can fluctuate rapidly due to project deadlines, workload increases, organizational changes, or personal factors, but traditional systems fail to capture these dynamic variations effectively.

Some early automated stress detection systems attempted to use simple rule-based approaches or limited workplace parameters such as absenteeism records and work performance metrics. However, these systems were unable to process complex relationships between multiple employee attributes simultaneously. They lacked the computational capability to analyze large-scale multi-dimensional data and therefore produced limited prediction accuracy.

Recent stress detection approaches have also incorporated wearable devices and physiological sensors such as heart rate monitors, EEG devices, and skin response sensors to measure stress levels. While these systems improved prediction accuracy, they introduced several practical challenges including high implementation cost, privacy concerns, device maintenance issues, and employee discomfort during continuous monitoring. Such systems are difficult to deploy in large organizations and may negatively affect employee trust and acceptance.

In addition, existing systems often lack scalability and automation. Conducting manual assessments for hundreds or thousands of employees requires significant time, effort, and financial resources from HR teams. These approaches are inefficient for large organizations that require continuous monitoring and rapid decision-making. Furthermore, most traditional systems do not provide advanced analytical dashboards or predictive insights that can help management proactively address employee stress trends.

Therefore, there is a need for an intelligent, automated, scalable, and data-driven stress prediction system capable of analyzing multiple employee-related parameters in real time. The limitations of existing systems motivate the development of the proposed Machine Learning Based Employee Stress Prediction System using an Ensemble Voting Classifier for accurate and efficient stress prediction.

IV. PROPOSED METHODOLOGY

The proposed methodology introduces an automated Machine Learning Based Employee Stress Prediction System designed to identify employee stress levels accurately using workplace and personal attributes. The system utilizes an Ensemble Voting Classifier that combines multiple machine learning algorithms to improve prediction accuracy, reliability, and robustness. Unlike traditional manual assessment methods, the proposed approach provides a data-driven, scalable, and real-time solution for employee stress monitoring.

The methodology begins with the collection of structured employee data containing more than 18 parameters related to workplace conditions, behavioral patterns, and personal information. These parameters include department, education, working hours, workload level, performance ratings, salary hike, work satisfaction, monthly income, training hours, marital status, and years at the company. The collected data is stored securely in a MySQL database and used as the input dataset for machine learning analysis.

After data collection, preprocessing techniques are applied to prepare the dataset for model training and prediction. Since machine learning algorithms require numerical input, categorical attributes such as gender, department, region, and marital status are converted into numerical values using Label Encoding. Missing values, inconsistent records, and invalid entries are handled during preprocessing to ensure data quality. The processed data is then transformed into structured feature vectors suitable for machine learning algorithms.

The core component of the proposed system is the Ensemble Voting Classifier, which integrates six different machine learning algorithms: Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and SGD Classifier. Each classifier independently analyzes the employee feature vector and predicts whether the employee belongs to the *Low Stress* or *More Stress* category. The final prediction is generated using a hard voting mechanism, where the majority vote among the six classifiers determines the output. This ensemble approach reduces the limitations of individual models and improves overall prediction performance.

The proposed system is implemented as a web-based application using the Django framework following the Model-View-Template (MVT) architecture. Employees can securely register, log in, and submit their stress-related information through an interactive web interface.

The backend processes the submitted data, performs preprocessing, and passes the feature vector to the ensemble model for prediction. The prediction results are displayed through a user-friendly interface with color-coded feedback indicating stress levels.

An administrative dashboard is also included in the system to support organizational stress analysis. The dashboard provides visual analytics such as pie charts, bar charts, line graphs, and summary reports showing overall stress distribution within the organization. Administrators can monitor stress trends, view prediction history, manage users, and analyze model performance metrics in real time.

To evaluate the effectiveness of the proposed methodology, the system is tested using standard performance metrics including Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. Experimental results demonstrate that the Voting Ensemble Classifier achieves superior performance with approximately 88–93% prediction accuracy, outperforming individual machine learning models. The proposed methodology therefore provides an efficient, scalable, and practical solution for early employee stress detection and proactive organizational wellness management.

V. BLOCK DIAGRAM

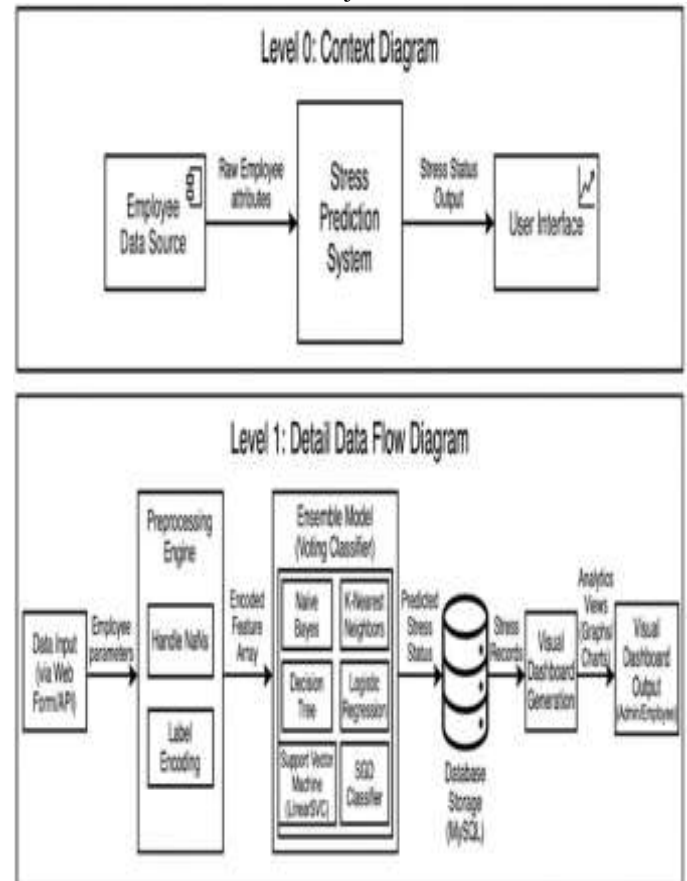


Fig. 6.2. Block Diagram

VI. RESULTS AND OUTCOMES

The proposed Machine Learning Based Employee Stress Prediction System was successfully implemented and tested using a structured employee dataset containing more than 18 workplace and personal parameters. The system effectively analyzed employee-related information and classified stress levels into two categories: *Low Stress* and *More Stress*. Experimental evaluation demonstrated that the integration of multiple machine learning algorithms through an Ensemble Voting Classifier significantly improved prediction performance and reliability.

During experimentation, six individual machine learning algorithms were trained and evaluated independently. These algorithms included Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, SGD Classifier, Logistic Regression, and Support Vector Machine (LinearSVC). Each model produced different levels of prediction accuracy depending on its learning capability and handling of employee-related features. The individual model accuracies ranged approximately from 78% to 89%.

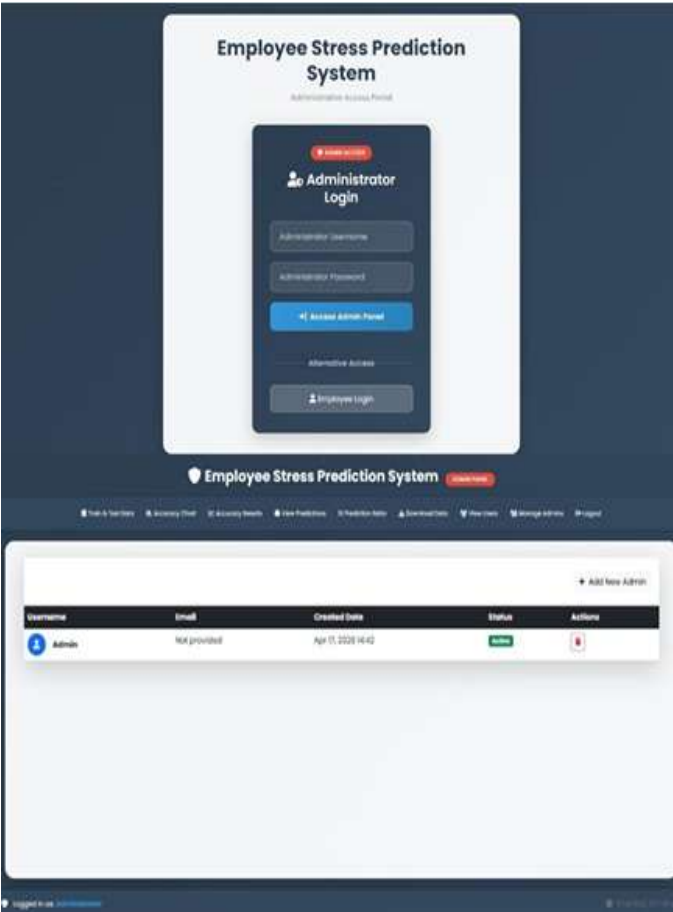


Fig: 7.1: Output 1

The Naive Bayes classifier achieved an accuracy of approximately 78–82%, serving as a fast baseline probabilistic model. K-Nearest Neighbors achieved around 80–85% accuracy by identifying similarities among employee profiles. Decision Tree and SGD Classifier models achieved accuracies between 82–87%, while Logistic Regression performed with approximately 84–88% accuracy. Among the individual classifiers, Support Vector Machine (LinearSVC) achieved the best performance with an accuracy of approximately 85–89%.

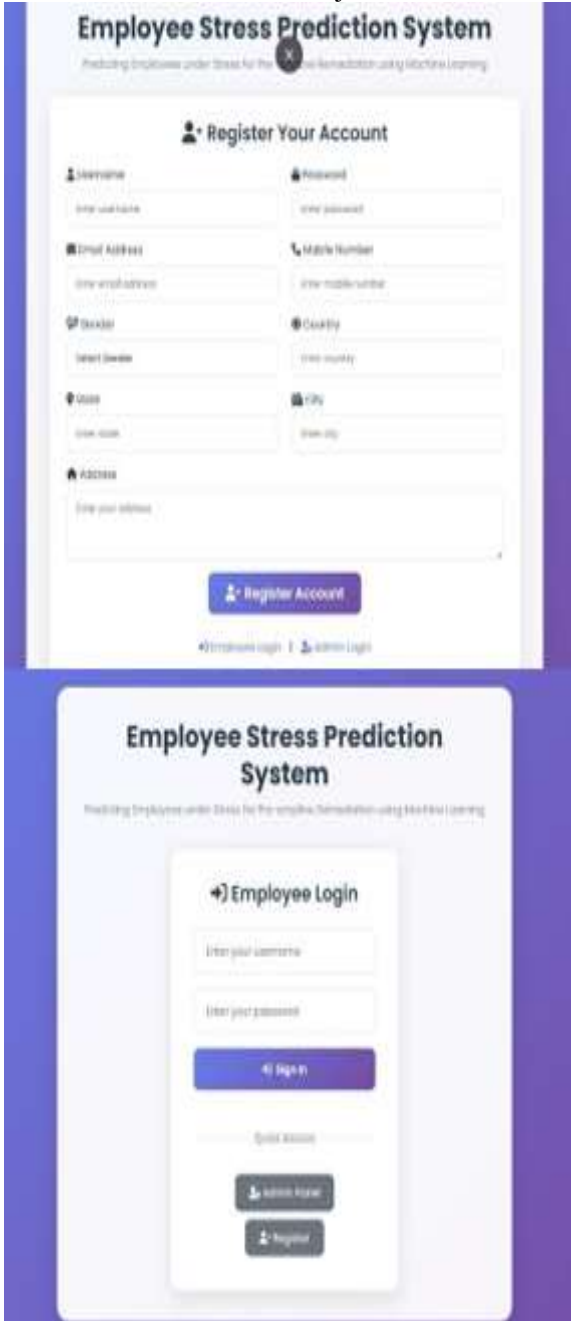


Fig: 7.2: Output 2

To improve overall performance, all six classifiers were combined using a Voting Ensemble Classifier. The ensemble model used a hard voting mechanism in which each classifier independently generated a prediction, and the final output was determined based on majority voting. The ensemble approach successfully reduced individual model bias and improved robustness. Experimental results showed that the Voting Ensemble Classifier achieved the highest prediction accuracy of approximately 88–93%, outperforming all individual algorithms.

10.48047/jocaaa.2024.33.08.627

interface. The administrator dashboard dynamically generated visual analytics such as stress distribution pie charts, bar graphs, and model performance charts, helping organizations monitor employee wellness trends effectively.

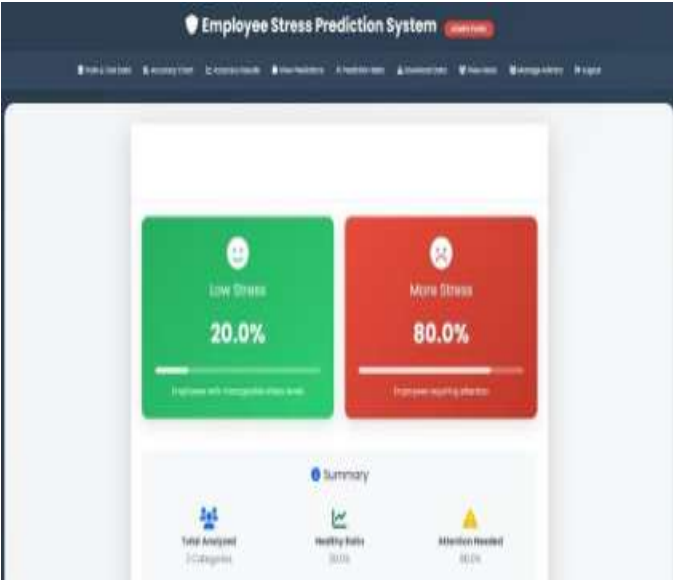


Fig: 7.3 : Output 3

The system was also evaluated using performance metrics such as Precision, Recall, F1-Score, and Confusion Matrix analysis. The ensemble model achieved high precision and recall values of approximately 0.90–0.91, indicating strong classification capability for both stress categories. The confusion matrix analysis confirmed that the model effectively minimized false predictions and improved classification reliability.



Fig: 7.5: Output 5

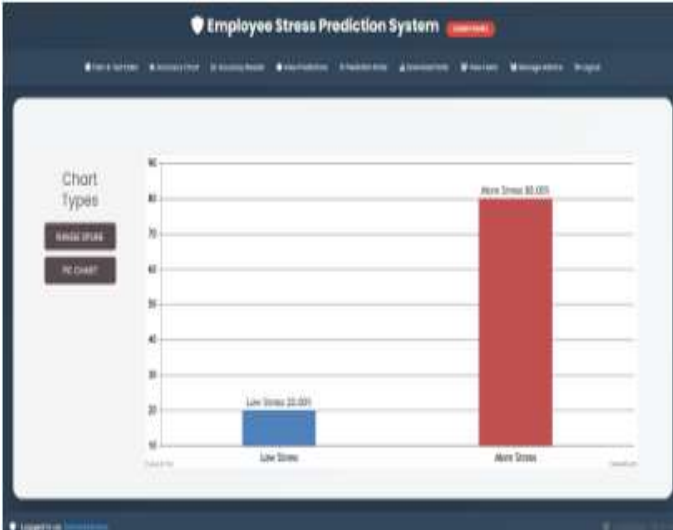


Fig: 7.4 : Output 4

In addition to prediction accuracy, the web-based application demonstrated efficient real-time functionality. Employees were able to register, submit stress-related data, and receive immediate prediction results through a user-friendly

The outcomes of the proposed system indicate that machine learning techniques can be effectively applied for early employee stress detection and organizational wellness management. The system provides several practical benefits including automated stress monitoring, reduced manual assessment effort, improved prediction reliability, scalable deployment, and proactive decision-making support for HR departments. By enabling early identification of stressed employees, the system can help organizations reduce burnout, improve employee satisfaction, and enhance overall productivity.

VII. CONCLUSION

The Machine Learning Based Employee Stress Prediction System successfully demonstrates the practical application of machine learning techniques in identifying employee stress levels within organizational environments. The proposed system effectively analyzes more than 18 workplace and personal parameters to classify employees into *Low Stress* and *More Stress* categories using an Ensemble Voting

Classifier. By combining six machine learning algorithms—Naive Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and SGD Classifier—the system achieves higher prediction accuracy and reliability compared to individual models.

The study highlights the limitations of traditional stress assessment methods such as manual surveys, HR interviews, and periodic evaluations, which are often subjective, time-consuming, and incapable of providing real-time monitoring. The proposed automated system overcomes these challenges by offering a scalable, objective, and data-driven solution for employee stress prediction. The integration of machine learning techniques enables efficient analysis of complex multi-dimensional employee data, supporting early detection of stress-related issues.

The implementation of the system using the Django framework and MySQL database provides a secure and user-friendly web application for both employees and administrators. Employees can perform self-assessment through an interactive interface, while administrators can monitor organizational stress trends using analytical dashboards and visual reports. The ensemble model achieved approximately 88–93% prediction accuracy, demonstrating the effectiveness of the proposed approach in real-world organizational scenarios.

The proposed system offers significant organizational benefits including reduced manual effort, improved employee wellness monitoring, early stress identification, and support for proactive HR interventions. By helping organizations identify at-risk employees before stress escalates into burnout or attrition, the system contributes to improved workforce productivity, employee satisfaction, and overall organizational performance.

In conclusion, the Employee Stress Prediction System provides an efficient, reliable, and scalable solution for occupational stress management using machine learning and ensemble learning techniques. The project demonstrates how intelligent data-driven systems can support modern organizations in maintaining healthier and more productive work environments.

REFERENCES

1. Neha Sharma & Tanveer Gedeon, “**Objective Measures, Sensors and Computational Techniques for Stress Recognition and Classification: A Survey**”, Published Year: 2012, [ScienceDirect Link](#)
2. Martin Gjoreski et al., “**Monitoring Stress with a Wrist Device Using Context**”, Published Year: 2017, [ScienceDirect Link](#)
3. Ane Alberdi et al., “**Towards an Automatic Early Stress Recognition System for Office Environments Based on Multimodal Measurements**”, Published Year: 2016, [ScienceDirect Link](#)
4. Amit Rastogi & Mayank Trivedi, “**Employee Stress and Burnout Prediction Using Machine Learning**”, Published Year: 2020, [ResearchGate Link](#)
5. Fabian Pedregosa et al., “**Scikit-learn: Machine Learning in Python**”, Published Year: 2011, [JMLR Link](#)
6. [Django Documentation](#), “**Django Documentation Version 4.2**”, Published Year: 2023. ([Django Project](#))
7. Leo Breiman, “**Bagging Predictors**”, Published Year: 1996, [Springer Link](#)
8. Corinna Cortes & Vladimir Vapnik, “**Support Vector Networks**”, Published Year: 1995, [Springer Link](#)
9. J. Ross Quinlan, “**Induction of Decision Trees**”, Published Year: 1986, [Springer Link](#)
10. Thomas M. Cover & Peter E. Hart, “**Nearest Neighbor Pattern Classification**”, Published Year: 1967, [IEEE Link](#)
11. David W. Hosmer et al., “**Applied Logistic Regression**”, Published Year: 2013, [Wiley Link](#)
12. Trevor Hastie, Robert Tibshirani & Jerome Friedman, “**The Elements of Statistical Learning**”, Published Year: 2009, [Springer Link](#)
13. Christopher Bishop, “**Pattern Recognition and Machine Learning**”, Published Year: 2006, [Springer Link](#)
14. Ian Goodfellow, Yoshua Bengio & Aaron Courville, “**Deep Learning**”, Published Year: 2016, [MIT Press Link](#)
15. Aurélien Géron, “**Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**”, Published Year: 2019, [O'Reilly Link](#)