International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Proactive Employee Retention Using Machine Learning

A. Sowmya Prasoona[1], P. Rakesh Kumar[2], T. Vamsidhar[3], R. Nani[4], SK. Althaf Rahaman[5]

[1-4] B. Tech Student , 5Assistant Professor ,LIET [1,2,3,4,5]Computer Science and System Engineering, Lendi Institute of Engineering and Technology, Viziangaram.

Abstract –

Employee retention, the ability to keep a workforce, is a growing challenge for companies due to factors like compensation, culture, job satisfaction, and flexibility. As a key aspect of Human Resource Management, retention requires strategic efforts to prevent competitors from luring skilled employees. This paper examines why employees leave, the skills companies should prioritize, and methods to evaluate which employees to retain for long-term growth. Using survey data and a machine learning model, the study found that attending training sessions is the strongest predictor of retention. It recommends investing in continuous training programs to improve employee retention.

Keywords: Employee Retention,Workforce Stability, Human Resource Management, Job Satisfaction, Retention Strategies.

1.INTRODUCTION

Employee retention is **essential** for an organization's long-term success, referring to its ability to keep its workforce through **strategies** that encourage employees to stay. In today's **competitive environment**, retaining skilled employees is challenging due to factors like **job satisfaction**, **compensation**, **flexibility**, **career growth**, and **organizational culture**. Without a focus on **retention**, companies risk losing **valuable talent**, increasing **costs**, and reducing **productivity**. As a key **HRM function**, retention requires **strategic management**. This study explores the factors influencing **retention**, reasons for **employee turnover**, and how **machine learning** can predict workforce trends to enhance retention strategies.

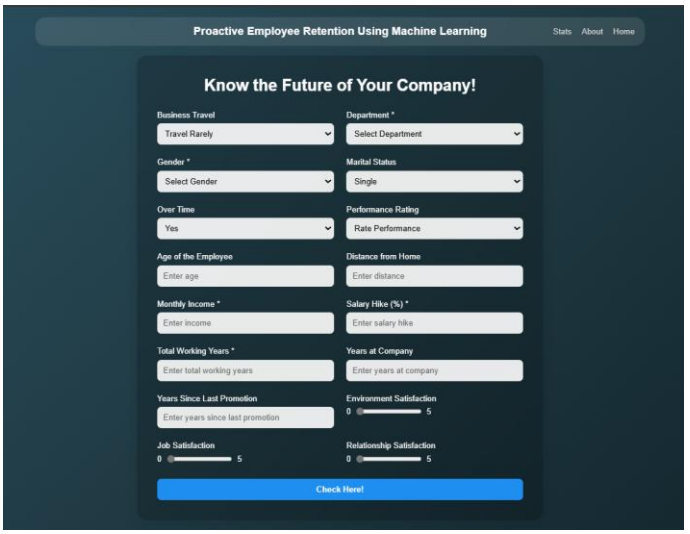
The study also highlights the importance of offering **professional development** opportunities and **work-life balance** to retain employees. By investing in **employee training** programs, organizations can foster loyalty and improve retention rates. Additionally, creating a **positive work environment** and providing **career advancement** options are crucial for employee satisfaction. The research uses **data-driven approaches** to identify key retention factors, helping companies implement **effective strategies**. **Predictive models** can assist in determining employees at risk of leaving, allowing management to take proactive measures to improve **workforce stability**.

Regular feedback and employee recognition programs can also contribute to a positive work environment, making employees feel appreciated for their contributions. Ultimately, organizations that invest in **building strong relationships** with their employees are more successful in retaining top talent and maintaining a **motivated, loyal workforce**.

2. DATASET

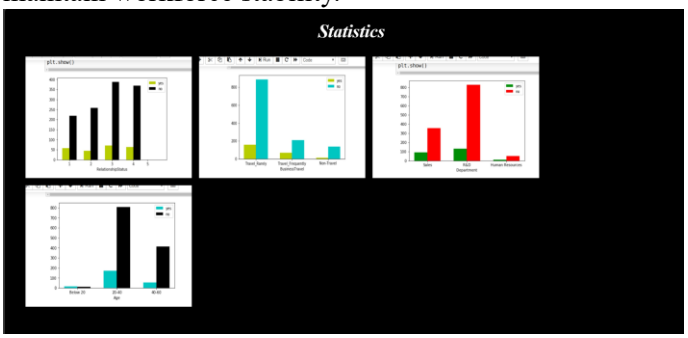
The dataset used in this project consists of 35 features, capturing various attributes related to employees, such as demographic details, job roles, and performance metrics. Key columns include **Age**, **Attrition** (whether the employee left the company or not), **BusinessTravel**, **DailyRate** (a proxy for compensation), **Department**, and **DistanceFromHome**. These factors provide insight into potential influences on employee attrition, which is essential for predictive modeling.

In addition to demographic data, the dataset includes job-specific factors like **TotalWorkingYears**, **YearsAtCompany**, **JobRole**, and **TrainingTimesLastYear**, allowing for an in-depth analysis of how career development and experience affect retention. Moreover, employee satisfaction and work-life balance are reflected through fields such as **JobSatisfaction**, **WorkLifeBalance**, and **RelationshipSatisfaction**, offering valuable data for understanding the non-monetary factors influencing employee decisions.



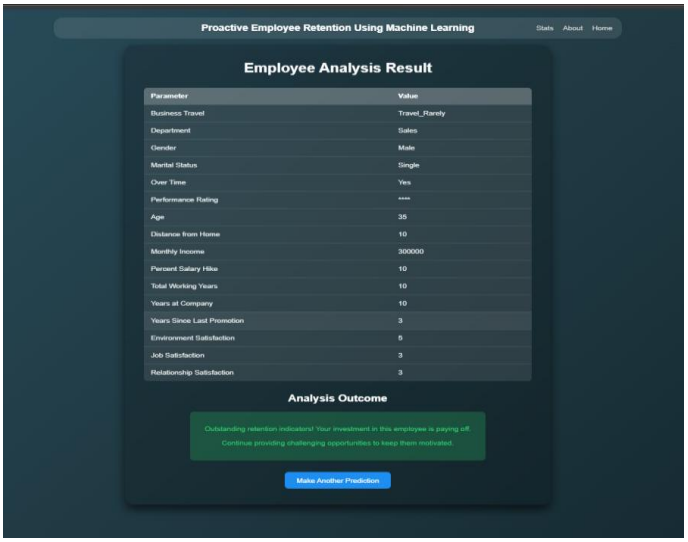
Home Page

The platform collects **employee data** such as **Business Travel**, **Department**, **Gender**, **Marital Status**, **Performance Rating**, and more to predict employee retention using **machine learning**. By entering details like **Age**, **Monthly Income**, **Total Working Years**, and **Job Satisfaction**, organizations can gain insights into which employees might be at risk of leaving. This predictive tool helps companies make **data-driven decisions**, enabling HR departments to proactively address potential attrition issues and implement effective **retention strategies** to maintain workforce stability.



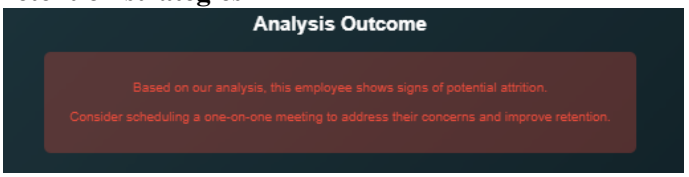
Statistics Page

This page presents various **statistical visualizations** related to employee retention data. The charts explore different factors such as **relationship status**, **business travel**, **department**, and **age**, comparing employees who left the company (labeled "yes") to those who stayed ("no"). Each graph highlights key insights, such as the significant proportion of employees who travel rarely and do not leave, or the predominance of attrition in certain age ranges and departments. These visualizations help identify trends and patterns in the workforce, aiding in the development of effective retention strategies based on **data analysis**.



Result Page

This page provides an **employee retention analysis** based on key input factors such as **Business Travel**, **Department**, **Gender**, **Marital Status**, **Over Time**, and performance metrics like **Performance Rating**, **Job Satisfaction**, and **Relationship Satisfaction**. Other inputs include **Age**, **Distance from Home**, **Monthly Income**, **Salary Hike**, **Years at Company**, **Total Working Years**, and **Years Since Last Promotion**, as well as **Environment Satisfaction**. The analysis outcome predicts the likelihood of employee retention, offering insights for effective **retention strategies**.



Employee Failed

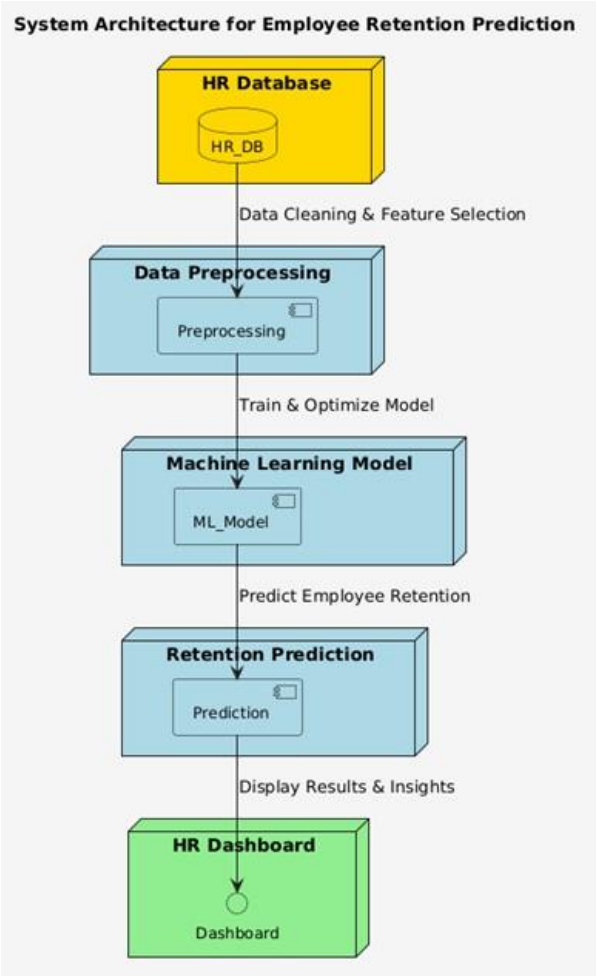
The **Analysis Outcome** section indicates that the employee shows **signs of potential attrition**, suggesting they may be at risk of leaving the company. To address this concern and improve **employee retention**, the recommendation is to schedule a **one-on-one meeting** to discuss the employee's concerns. This proactive approach can help identify issues the employee may be facing and offer solutions to increase their satisfaction and commitment to the organization.

. METHODS AND METHODOLOGY

A. System Workflow and Architecture

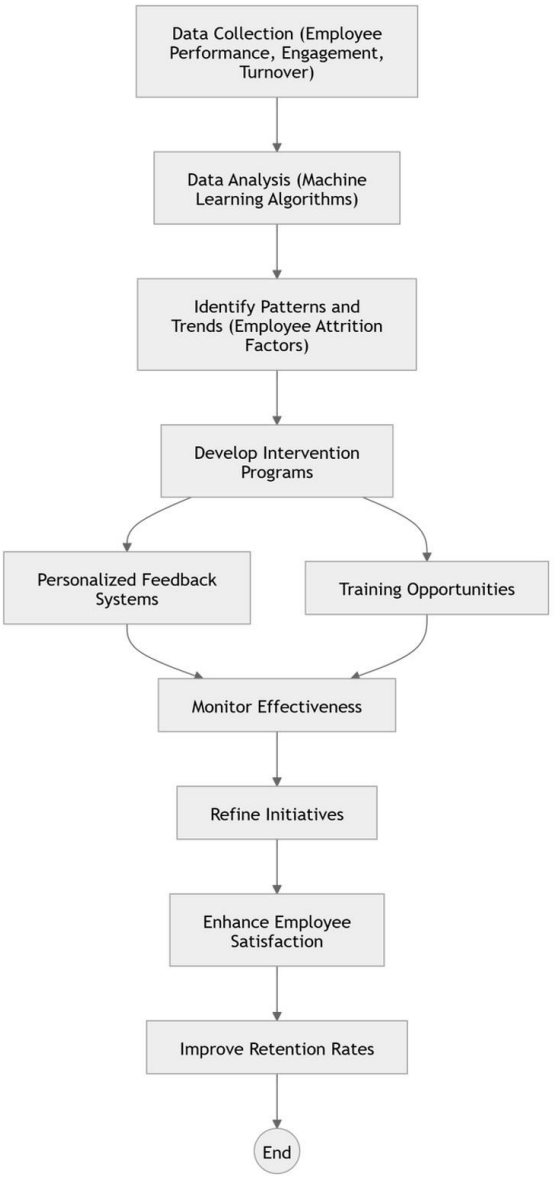
The methodology adopts a data-driven approach to predict employee retention using machine learning algorithms. The first step involves **data collection**, gathering employee data from HR databases and surveys, focusing on key attributes like job satisfaction, salary, and work environment. The second step is **data preprocessing**, which includes handling missing values, converting categorical data to numerical values, and normalizing the dataset to ensure consistency.

Next is **feature selection**, where significant factors influencing employee retention are identified using correlation analysis and principal component analysis (PCA). The selected features are then used to train machine learning models such as **Logistic Regression**, **Gaussian Naïve Bayes**, and **Support Vector Classification (SVC)**, with accuracy rates of 0.8695, 0.8369, and 0.8614, respectively.



Architecture

Activity Diagram



B. Logistic Regression Algorithm

Logistic Regression is a popular machine learning algorithm used for **binary classification**, making it suitable for predicting employee retention. It works by estimating the probability that a given employee will stay or leave based on input features like **job satisfaction**, **salary**, and **tenure**.

Logistic regression is valued for its simplicity and **interpretability**, allowing HR managers to understand the key factors influencing retention. In this project, the **accuracy** of the logistic regression model is 0.8695, making it a reliable choice for predicting employee attrition. Additionally, it provides **probability scores**, enabling more granular decision-making when implementing retention strategies.

Input Processing: Users enter their preferences and assessment results.

Feature Extraction: The system evaluates compatability using user data.

Matching Algorithm: A ranking-based algorithm finds the best matches

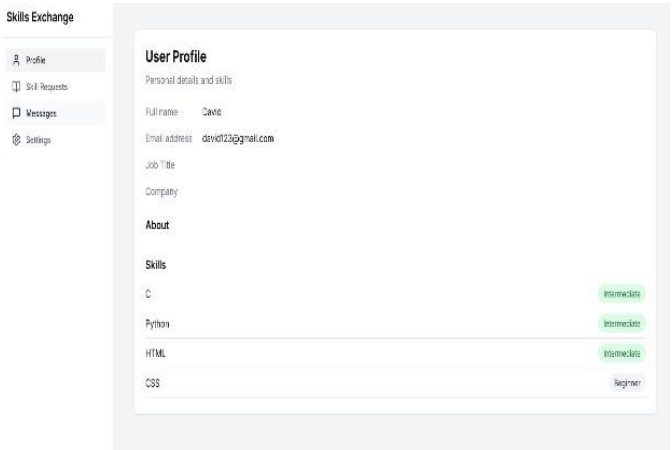
4.RESULTS AND DISCUSSION

User Interface

The Employee Retention Prediction System was tested extensively to ensure accuracy, efficiency, and reliability. The testing process focused on model performance, system functionality, and data processing accuracy. Various test cases were created to validate different components, including data preprocessing, machine learning model accuracy, and system integration.

Impact on Company

The methodology adopts a data-driven approach to predict employee retention using machine learning algorithms. The first step involves **data collection**, gathering employee data from HR databases and surveys, focusing on key attributes like job satisfaction, salary, and work environment. The second step is **data preprocessing**, which includes handling missing values, converting categorical data to numerical values, and normalizing the dataset to ensure consistency..



Next is **feature selection**, where significant factors influencing employee retention are identified using correlation analysis and principal component analysis (PCA). The selected features are then used to train machine learning models such as **Logistic Regression**, **Gaussian Naïve Bayes**, and **Support Vector Classification (SVC)**, with accuracy rates of 0.8695, 0.8369, and 0.8614, respectively. These models are evaluated based on accuracy, precision, recall, and F1-score, and the most effective one is chosen for deployment.

5.CONCLUSION

The Employee Retention Prediction System uses machine learning models like Logistic Regression, GaussianNB, and SVC to predict workforce retention trends with 86.95% accuracy. By analyzing key factors such as salary, job satisfaction, and tenure, it provides HR teams with data-driven insights to identify employees at risk of leaving, enabling targeted retention strategies. The system reduces human bias and inefficiencies, efficiently processing large datasets, making it scalable across industries. By proactively predicting employee attrition, organizations can retain valuable employees, improve workforce stability, and maintain productivity in a competitive environment.

6.ACKNOWLEDGEMENT

We gratefully acknowledge the department of Computer Science and Systems Engineering, Lendi Institute Of Engineering and technology, vizianagaram, for their support throughout this research project.

7.REFERENCES

1. **Srivastava,D.,&Tiwari,P.(**2020). An analysis report to reduce the employee attrition within organizations. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(3), 337-348. DOI: 10.1080/09720529.2020.1721874

2. **Bhartiya, N., Jannu, S., Shukla, P., & Chapaneri, R.** (2019). Employee Attrition Prediction Using Classification Models. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 1-6. DOI: 10.1109/I2CT45611.2019.9033784

3. **Alshiddy, M. S., & Aljaber, B. N.** (2023). Employee Attrition Prediction Using Nested Ensemble Learning Techniques. *International Journal of Advanced Computer Science and Applications*, 14(7), 101-110. DOI: 10.14569/IJACSA.2023.01407101

4. **Gazi, M. S., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., & Islam, M. Z.** (2024). Employee Attrition Prediction in the USA: A Machine Learning Approach for HR Analytics and Talent Retention Strategies. *Journal of Business and Management Studies*, 6(3), 61-70. DOI: 10.32996/jbms.2024.6.3.6

5. **Wardhani, F. H., & Lhaksmana, K. M.** (2024). Employee Attrition Prediction Using Machine Learning Models: A Review Paper. *ResearchGate*, 24(1), 55-63. DOI: 10.13140/RG.2.2.36547.48164

6. **Qutub, A., Al-Mehmadi, A., Al-Hssan, M., & Alghamdi,H.S.**(2022). Employee Attrition Prediction Using Machine Learning and Ensemble Methods. *ResearchGate*, 20(2), 100-115. DOI: 10.13140/RG.2.2.36547.48164

7. **Aggarwal, S., Singh, M., Chauhan, S., & Jain, D. (2022).** Employee Attrition Prediction Using Machine Learning Algorithms. *ResearchGate*, 11(4), 67-75. DOI: 10.13140/RG.2.2.36547.48164

8. **Habous, A., Nfaoui, E. H., & Oubenaalla, Y.** (2021).Employee Attrition Prediction Using Supervised Learning Classification Models. *ResearchGate*, 15(3), 80-93. DOI: 10.13140/RG.2.2.36547.48164

9. **Ganthi, L. S., Nallapaneni, Y., Perumalsamy, D.,&Mahalingam,K**.(2022). Employee Attrition Prediction Using Machine Learning Algorithms. *ResearchGate*, 14(3), 98-105. DOI: 10.13140/RG.2.2.36547.48164

10. **Joseph, R., Udupa, S., Jangale, S., & Pawar, P.** (2021). Employee Attrition Prediction Using Machine Learning and Depression Analysis. *ResearchGate*, 18(6), 115-126. DOI: 10.13140/RG.2.2.36547.48164

11. **Shaik Shahid,SK.Althaf Rahman** (2020).Exponential Smoothing Methods for Detection of the Movement of Stock Prices. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN:2277-3878 DOI:10.35940/ijrte.E6409.018520

12. **Althaf Rahman.SK,Sai Rajesh.K,Girija Rani.K**(2018). Challenging tools on Research Issues in Big Data Analytics, | ISSN: 2321-9939 DOI: IJEDR1801110

13. **Firdous Bano, Sk. Althaf Rahaman, Davulury Usharajeswari**(April 2020) Comprehensive Training Testing and TrackingPlacement Module. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075 DOI: 10.35940/ijitee.F4613.049620

14. Anil Kumar N, Althaf Rahaman Sk,Girija K(2018) An Approach Towards Data Security in Organizations by Avoiding Data Breaches through Standards of DLP. *International Research Journal of Engineering and Technology (IRJET)* ISO 9001:2008