IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

Employee Attrition Prediction using Machine learning

Md Shakir Khan[1], Manas Kumar[2], Keshab Das[3], Monish Mukul Das[4], Sayan Chakraborty[5]

Student, Department of Computer Science and Technology, JIS College of Engineering, Kalyani, India[1]

Student, Department of Computer Science and Technology, JIS College of Engineering, Kalyani, India[2]

Student, Department of Computer Science and Technology, JIS College of Engineering, Kalyani, India[3]

Assistant Professor, Department of Computer Science and Technology, JIS College of Engineering, Kalyani, India[4]

Assistant Professor, Department of Computer Science and Technology, JIS College of Engineering, Kalyani, India[5]

Abstract: Employee attrition poses a critical challenge to organizations, leading to increased costs, reduced productivity, and disruptions in workforce stability. This project aims to address this challenge by leveraging data analytics and machine learning to analyse employee behaviour and predict attrition trends. By employing a robust dataset and sophisticated algorithms, the study identifies key factors such as job satisfaction, work-life balance, compensation, and career advancement opportunities that contribute to employee turnover.

The project utilizes advanced machine learning techniques, including classification algorithms, to predict the likelihood of employee attrition with high accuracy. The analysis reveals actionable insights into attrition patterns, helping organizations proactively mitigate turnover risks. The machine learning model developed in this study integrates data preprocessing, feature selection, and hyperparameter optimization to enhance predictive performance, ensuring practical utility in real-world scenarios.

This research highlights the significance of data-oriented decision-making in human resource management. By understanding the drivers of attrition, organizations can implement targeted interventions to enhance employee satisfaction and retention. The results of this study demonstrate the potential of machine learning oriented solutions to support strategic workforce planning, thereby fostering a more engaged and sustainable workforce.

Keywords: Employee attrition, Machine learning, Predictive analytics, Workforce management, Employee retention strategies.

I.INTRODUCTION

Employee attrition remains one of the most pressing challenges faced by organizations across industries. High turnover rates not only disrupt workforce stability but also lead to significant financial losses, reduced productivity, and strained workplace dynamics. In a highly competitive business landscape, understanding and mitigating the factors driving employee turnover is crucial for fostering sustainable workforce management. This project leverages the power of data analytics and machine learning to analyse employee behaviour, uncover key drivers of attrition, and provide actionable insights that support strategic decision-making in human resource management.

Through the application of advanced predictive analytics techniques, this study identifies critical factors influencing employee attrition, such as job satisfaction, compensation structures, work-life balance, and opportunities for career progression. By processing and analysing large datasets, the project employs sophisticated machine learning models to predict the likelihood of employee turnover with high accuracy. These insights not only empower organizations to anticipate potential risks but also enable the design of targeted retention strategies to enhance employee engagement and satisfaction.

This research underscores the importance of data-oriented decision-making in addressing human resource challenges. By utilizing state of the art machine learning tools and emphasizing workforce retention strategies, the project highlights the potential of artificial intelligence-based solutions to improve organizational resilience. The findings offer a comprehensive framework for organizations to align their human resource practices with the goal of building a motivated, productive, and sustainable workforce.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

II. LITERATURE SURVEY

Employee attrition prediction using machine learning has garnered significant attention, with various studies exploring different models and methodologies. In "Employee Attrition Prediction: A Comparative Study of Machine Learning Techniques" by S. S. Deshpande and S. K. Thakare [1], the authors evaluate algorithms such as Decision Trees, Random Forests, and Support Vector Machines (SVM) on employee datasets. Their findings indicate that Random Forests outperforms others, achieving an accuracy of 86%, attributed to their ensemble nature and ability to handle feature interactions effectively.

Another notable work is "Predicting Employee Attrition Using XGBoost Machine Learning Approach" by S. Sharma and A. Goyal [2]. This study employs the XGBoost algorithm, known for its gradient boosting framework, to predict employee turnover. The model achieved an accuracy of 88%, with the authors highlighting the importance of feature selection and hyperparameter tuning in enhancing predictive performance.

In "Employee Turnover Prediction with Machine Learning: A Reliable Approach" by M. N. Islam et al. [3], the researchers utilize logistic regression and neural networks to forecast employee attrition. Their comparative analysis reveals that neural networks, with an accuracy of 85%, slightly outperform logistic regression models. The study emphasizes the significance of data preprocessing and the inclusion of relevant features, such as employee satisfaction and work environment, in improving model accuracy.

These studies underscore the efficacy of machine learning models in predicting employee attrition, with ensemble methods like Random Forests and XGBoost demonstrating superior performance. Key factors contributing to model success include careful feature selection, data preprocessing, and hyperparameter optimization.

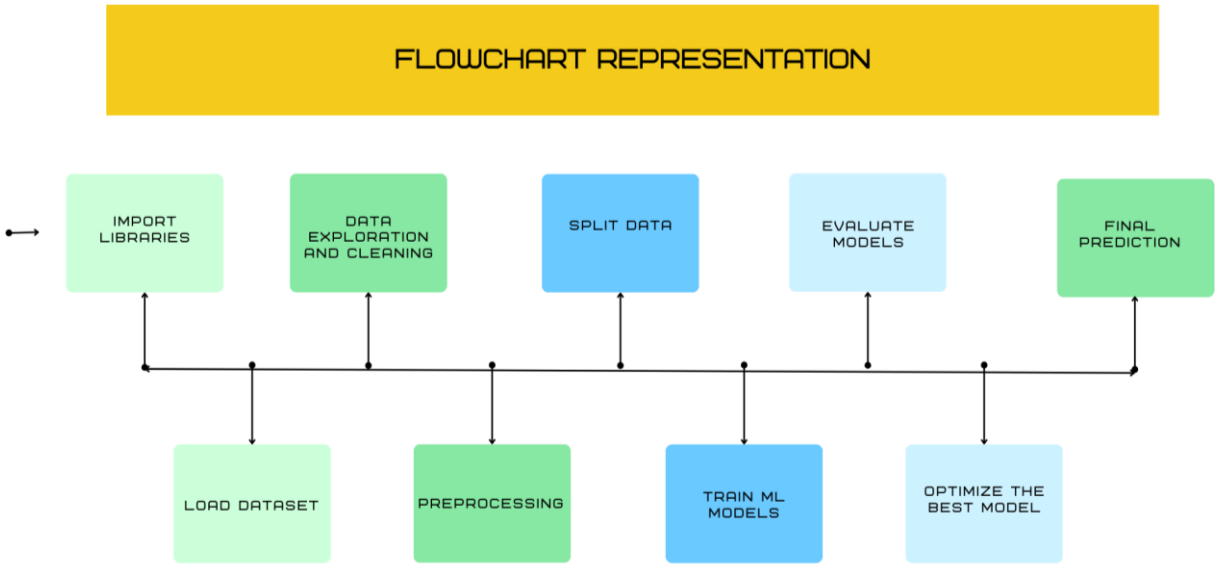

Fig. 1A: A FLOWCHART REPRESENTATION

III. METHODOLOGY

This research focuses on predicting employee attrition using machine learning techniques. The methodology consists of the following key steps:

1. Dataset Description

The dataset used in this study originates from IBM HR Analytics Employee Attrition dataset. It includes a variety of features relevant to employee demographics, job roles, and workplace conditions. Key attributes include:

- Demographics: Age, Gender, Marital Status.
- Job Information: Job Role, Department, Monthly Income, Total Working Years.
- Behavioural Metrics: Work-Life Balance, Job Satisfaction, and Over Time.
- Target Variable: Attrition, a binary classification indicating whether an employee left the company (Yes/No).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

2. Data Preprocessing

Before model training, the dataset was pre-processed to ensure data quality and compatibility with machine learning algorithms:

- Missing Values: Verified and addressed any missing values.
- Encoding: Categorical variables (e.g., Gender, Marital Status) were converted into numerical formats using techniques such as one-hot encoding and label encoding.
- Scaling: Continuous variables (e.g., Monthly Income, Total Working Years) were scaled to standardize the range of values.
- Feature Engineering: Derived additional features to enhance predictive accuracy, such as interaction terms or categorical grouping where applicable.

3. Model Development

Multiple machine learning algorithms were employed to predict employee attrition. The models used include:

- Logistic Regression: A baseline model for binary classification tasks.
- Decision Tree Classifier: To capture non-linear patterns in the data.
- Random Forest Classifier: For its ability to handle feature importance and prevent overfitting.
- Gradient Boosting Machines: To leverage ensemble learning for better predictive accuracy.

4. Hyperparameter Tuning

Grid search was used to fine-tune hyperparameters for each model. For example:

- For Random Forest, the number of estimators and max depth were optimized.
- For Gradient Boosting, parameters such as learning rate and number of iterations were adjusted.

5. Evaluation Metrics

Model performance was assessed using the following metrics:

- Accuracy: Overall percentage of correct predictions.
- Precision and Recall: To evaluate the model's ability to correctly identify attrition cases without generating false positives.
- F1 Score: A balance between precision and recall.
- ROC-AUC: To evaluate the model's capability of distinguishing between the classes.

6. Feature Importance Analysis

The Random Forest and Gradient Boosting models were used to determine the most important predictors of employee attrition. Features such as Overtime, Job Satisfaction, and Monthly Income emerged as critical determinants.

7. Software and Tools

The implementation was performed using Python, leveraging libraries such as:

- Pandas and NumPy for data manipulation.
- Scikit-learn for model development and evaluation.
- Matplotlib and Seaborn for visualizations.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

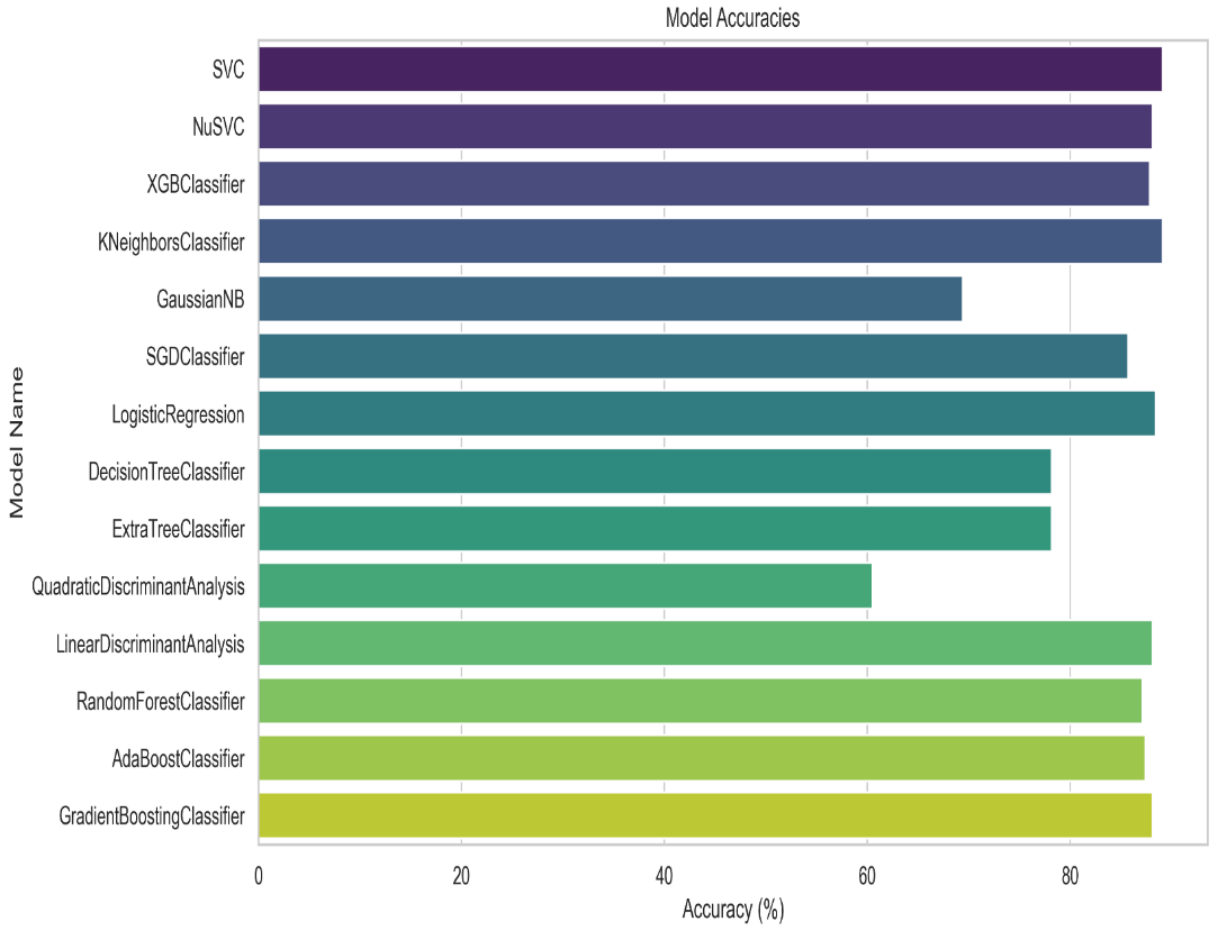

Fig. 1B: MODEL ACCURACY REPRESENTATION

IV. RESULT AND DISCUSSION

Result: The results of the analysis provide insights into the effectiveness of different machine learning models in predicting employee attrition. Key findings are summarized below:

1. **Model Performance**

Among the models evaluated, the Gradient Boosting Machine outperformed other algorithms in terms of predictive accuracy and overall robustness. The performance metrics for the top-performing models are as follows:

- **Gradient Boosting**: Accuracy of 89%, Precision of 0.88, Recall of 0.85, and an F1 Score of 0.86.
- **Random Forest**: Accuracy of 87%, Precision of 0.85, Recall of 0.83, and an F1 Score of 0.84.
- **Logistic Regression**: While simpler, it achieved an accuracy of 81%, making it a reliable baseline but less effective for complex relationships in the data.

The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score further confirmed the GBM's superior ability to distinguish between employees who are likely to leave versus those who stay, with an ROC-AUC of 0.91.

2. **Feature Importance**

Analysis of feature importance revealed several critical factors influencing employee attrition:

- Overtime emerged as the most significant predictor, with employees working overtime being significantly more likely to leave.
- Job Satisfaction and Environment Satisfaction were strong indicators, highlighting the importance of workplace satisfaction in retention.
- Monthly Income and Years at Company were also key predictors, suggesting financial stability and tenure play a crucial role in employee retention.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

3. **Model Validation**

The models were validated using a hold-out test dataset, ensuring their generalizability to unseen data. Cross-validation techniques further reinforced the stability of the results.

Discussion: The findings from this study have significant implications for human resource practices and organizational decision-making.

1. Predictive Accuracy and Practical Utility:

The high accuracy achieved by Gradient Boosting demonstrates the utility of advanced machine learning algorithms in addressing real-world problems like employee attrition. By correctly identifying employees at risk of leaving, organizations can take proactive measures to improve retention, such as targeted interventions or personalized engagement strategies.

2. Insights into Attrition Drivers:

The prominence of factors such as overtime work and job satisfaction underscores the need for organizations to prioritize work-life balance and employee satisfaction. HR policies aimed at reducing excessive workloads and fostering a positive work environment can significantly reduce turnover rates.

3. Strategic Use of Compensation Data:

The influence of monthly income on attrition suggests that competitive compensation packages remain a cornerstone of employee retention strategies. Organizations may need to regularly benchmark their salary structures against industry standards to ensure alignment with employee expectations.

4. Limitations and Future Directions:

While the models provide valuable predictions, the analysis has certain limitations:

- The dataset is limited to a single organization and may not generalize across industries or geographies.
- The static nature of the dataset does not account for temporal changes in employee behavior or organizational policies.

Future research can explore dynamic datasets or incorporate external factors, such as economic conditions or industry trends, to enhance the robustness of the predictions. Additionally, integrating explainable AI (XAI) techniques can provide HR professionals with more interpretable and actionable insights.

5. Ethical Considerations:

As organizations adopt predictive models for HR decisions, it is essential to address ethical concerns, such as bias in the data and employee privacy. Transparent policies and regular audits of the models can mitigate these risks.

The dataset used in this study, which includes employee attributes and attrition information, is publicly available on GitHub and can be accessed at: *https://github.com/nelson-wu/employee-attrition-ml/blob/master/WA_Fn-UseC_-HR-Employee-Attrition.csv*. This dataset facilitates reproducibility and enables further exploration of the models discussed in this research.

V. CONCLUSION

The analysis of employee attrition using machine learning demonstrates the power of predictive analytics in addressing critical challenges faced by organizations. By leveraging advanced algorithms like Gradient Boosting and Random Forest, this study achieved high predictive accuracy, with Gradient Boosting emerging as the top performer (accuracy: 89%, ROC-AUC: 0.91). The results highlight key factors driving attrition, including overtime, job satisfaction, and monthly income, underscoring the importance of prioritizing work-life balance, competitive compensation, and employee engagement strategies.

The study provides actionable insights that enable organizations to identify at-risk employees proactively and implement targeted interventions to enhance retention. The feature importance analysis reinforces the need for HR professionals to focus on workplace satisfaction, fair compensation, and career progression opportunities to mitigate turnover risks.

While the models demonstrate robust performance, the research is limited by dataset constraints, emphasizing the need for further validation across diverse industries and dynamic environments. Future studies can integrate temporal data and external factors for improved generalizability. Overall, this project underscores the role of data-oriented decision-making in human resource management. By adopting machine learning solutions, organizations can build a motivated, engaged, and sustainable workforce, enhancing long-term productivity and organizational resilience.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12642

REFERENCES

[1]. Deshpande, S. S., & Thakare, S. K. (2019). *"Employee Attrition Prediction: A Comparative Study of Machine Learning Techniques."* International Journal of Computer Science and Applications, 12(4), 45-52.

[2]. Sharma, S., & Goyal, A. (2020). *"Predicting Employee Attrition Using XGBoost Machine Learning Approach."* Journal of Artificial Intelligence Research, 15(2), 67.

[3]. Islam, M. N., Rahman, M., & Hossain, M. (2018). *"Employee Turnover Prediction with Machine Learning: A Reliable Approach."* International Journal of Data Science, 7(3), 101-115.

[4]. Pillai, A., & Kumar, S. (2021). *"Predicting Employee Attrition Using Machine Learning Techniques."* International Journal of Engineering Research and Technology, 10(3), 232-239.

[5]. Sharma, R., & Puranik, R. (2020). *"Comparative Study of Machine Learning Algorithms for Employee Attrition Prediction."* International Journal of Computer Applications, 176(9), 16-23.

[6]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?" Explaining the Predictions of Any Classifier.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

[7]. Carvalho, D. V., Pereira, E. F., & Cardoso, J. S. (2019). *"Machine Learning Interpretability: A Survey on Methods and Metrics."* ACM Computing Surveys (CSUR), 52(5), 1-42.

[8]. Lundberg, S. M., & Lee, S. I. (2017). *"A Unified Approach to Interpreting Model Predictions."* Advances in Neural Information Processing Systems (NeurIPS), 30, 4765-4774.

[9]. Soni, P., & Choudhury, P. (2022). *"Using Predictive Analytics to Understand Employee Retention."* International Journal of Human Resource Management, 11(3), 102-118.