

Forecasting Employee Attrition in HRM Based on Advanced Supervised Machine Learning Models

Dr. Parth Gautam
Associate Professor
Department of Computer Sciences and Applications
Mandsaur University, Mandsaur
parth.gautam@meu.edu.in

Abstract—Human resources are a company's most valuable asset, so losing employees is a big deal for the business. Human resources managers and department heads have a number of challenges when trying to predict when employees may start to leave their positions. Consequently, this research provides a machine learning-based methodology for HR planning-related employee turnover prediction utilizing the IBM HR dataset. Data cleaning, label encoding, min-max scaling, and class balancing with SMOTE were all part of the preprocessing stages to ensure reliability and reduce bias. Including 35 demographic, job-related, and satisfaction-related factors, the dataset comprised 1,470 employee records. Decision Tree, AdaBoost, and Support Vector Machine were some of the more conventional models tested with two more modern ensemble models, XGBoost and Random Forest (RF). As shown in the testing findings, XGBoost achieved a 96.95% accuracy rate and RF a 93.90% accuracy rate, both of which were better than the benchmark models. With its outstanding generalizability and predictive capacity, XGBoost stood out as the most effective of the bunch. The results show that the suggested models can handle complicated, high-dimensional HR data and provide dependable predictions of employee turnover, demonstrating their resilience and scalability. By allowing firms to detect people at risk early and put effective retention initiatives in place, this adds to proactive HR decision-making.

Keywords—Employee Turnover, Human Resource Planning, Machine Learning, Predictive analytics, Attrition Prediction.

I. INTRODUCTION

People management is an essential component of organizational processes, as human resources and the organization itself are inseparable. Employees in well-managed organizations are often regarded as the primary source of productivity growth, highlighting the strategic importance of human resource management (HRM) [1][2]. Within HRM, human resource planning (HRP) is critical, not only as a responsibility of HR managers but also as a collective task involving all organizational leaders to align personnel needs with organizational progress.

Employee loss has become one of the biggest problems that businesses all over the world have to deal with. In today's globally competitive business world, change not only slows things down, but it also costs a lot in terms of money and work output [3][4]. Predicting turnover is hard because many things affect each other. Affective commitment, work satisfaction, embeddedness in one's work, and burnout are all part of this category. It is well-known that all of these indicate a strong desire to depart [5][6]. Traditional ways of measuring turnover are limited by different types of data, noise, and the failure to keep up with how the workforce is changing.

The rise of HR analytics and big data has provided organizations with new tools for strategic workforce planning. HR analytics involves systematic analysis of workforce data, enabling organizations to identify patterns in performance, engagement, and organizational dynamics to support evidence-based decision-making [7][8]. Integrating Artificial Intelligence (AI) into this domain has further advanced predictive capabilities, offering proactive insights into turnover by analyzing behavioural, performance, compensation, and financial data [9][10]. AI-driven predictions enable HR and finance teams to implement

targeted interventions such as optimizing compensation structures, enhancing engagement, and addressing dissatisfaction before attrition occurs.

Machine learning (ML), a subset of AI, is particularly promising for turnover prediction as it can process heterogeneous, large-scale datasets to uncover hidden patterns and make accurate predictions [11]. Despite limited adoption in practice, ML-based predictive analytics is rapidly gaining attention for its ability to automate data-driven insights [12]. Moreover, recent advances in deep learning further enhance HR applications, ranging from recruitment optimization to employee well-being monitoring, including frameworks such as visual heart rate estimation for stress analysis [13][14]. This research compares and contrasts supervised ML models that are used for strategic human resource planning to forecast employee turnover. Organizations can benefit from robust prediction models that help reduce the risks of employee turnover, increase workforce stability, and boost overall organizational performance. This study tries to find these models by methodically comparing several algorithms.

A. Motivation and Contribution

A significant issue for firms is the growing complexity of staff turnover, which causes a drop in productivity, increases the expenses of recruiting and training new employees, and can even lead to a loss of valuable information. Traditional approaches, such as turnover intention surveys, often fail to capture the multifaceted and dynamic factors influencing attrition, including job satisfaction, commitment, and organizational dynamics, especially in the presence of noisy and heterogeneous data. This limitation highlights the need for more robust, data-driven methods to accurately predict and manage employee turnover. Motivated by this gap, the study leverages machine learning and HR analytics to systematically analyze diverse employee-related factors and

develop predictive models that can identify at-risk employees. By integrating advanced algorithms, the study aims to support proactive HR decision-making, enhance workforce stability, and provide organizations with effective tools to design targeted retention strategies. Several important advances in the area of human resource planning-based employee turnover prediction have resulted from this study:

- Utilized the IBM HR dataset from Kaggle as a benchmark source for analyzing employee turnover.
- Implemented systematic steps, including data cleaning, label encoding for categorical features, min-max scaling for normalization, and SMOTE for addressing class imbalance.
- Applied advanced ensemble learning models, namely XGBoost and Random Forest, to capture complex relationships influencing employee attrition.
- Analyzed the efficacy of the model by calculating its area under the curve (AUC), recall, accuracy, precision, and F1-score.

B. Novelty and Justification

The novelty of this study lies in integrating advanced ensemble learning techniques, specifically XGBoost and Random Forest, with systematic pre-processing strategies to address challenges commonly overlooked in employee turnover prediction, such as class imbalance and high-dimensional categorical data. Unlike conventional approaches that often rely on simpler models or neglect data balancing, this research leverages SMOTE for resampling, along with label encoding and normalization, to ensure fair representation and reliable predictions. The demand for proactive retention measures, which can be achieved through accurate risk identification of employee attrition, is the driving force behind this study in human resource planning. Academic research and real-world HR management practices are both enhanced by the study's more thorough and practical framework for predicting employee turnover, which is achieved by integrating robust algorithms with balanced and well-processed data.

C. Structure of the Paper

The outline of the paper is as follows The Literature Review, which covers previous research and current methods for predicting employee turnover and analyzing HR data, is presented in Section II. Data pre-processing is covered in Section III, which discusses the methodology. The Results are discussed in Section IV. Section V wraps up the research and shows where the field is heading in terms of future work, including topics like fairness in HR analytics and explainable AI.

II. LITERATURE REVIEW

An extensive review and analysis of significant research on employee turnover prediction within HR planning has been conducted to guide and strengthen the foundation of this study.

Fang and Zhang's (2025) performance was compared between the training and validation sets. Finally, cross-validation methods were used to confirm the model's generalizability. Training on GAN-balanced data with only five features chosen by RFE yielded the greatest results for the RF model, which had a precision of 0.98, recall of 0.92, F1-score of 0.95, and Cohen's Kappa of 0.94. With the same methodology applied to the IBM HR dataset, they were able

to confirm that their strategy was resilient and applicable to other datasets [15].

M et al. (2025) present a fresh approach to workforce management by suggesting an ML-based personnel classification system to tackle the pressing problem of staff turnover. Employing state-of-the-art ML techniques, the system uses a dataset from XYZ Corporation to forecast employee turnover with remarkable precision and interpretability. The top-performing algorithm was the XGBoost model, which achieved a precision of 89.7%, an accuracy of 92.3%, and an F1-score of 90.6% [16].

Pachica, Fajardo and Medina (2024) show that compared to other optimizers, such as Nadam, Admix, and Adam, the Adam optimizer using the tanh-based warm-up method performs the best, with a convergence time of 50 epochs and an accuracy of 95%. Organizations looking for economical and accurate models to forecast employee turnover may find this unique optimization technique useful, according to these findings [17].

Recilla et al. (2024) The FMSAX operator demonstrated superior performance compared to the other GA crossover operators, according to the study. Compared to the initial GA, the optimized result produced by the FMSAX operator is 73.33% better, or 69.24% better. After optimizing with the AX operator (43.33% rate of success), they used the CAX operator (56.66% success rate) to get the optimal number of variables [18].

Sharma and Sharma (2023). The dataset used in the study consists of 1471 rows and 35 columns and contains information about employees' demographics, jobs, and performance. In terms of accuracy, the logistic regression model came in at 85% and the random forest classifier model at 87% [19].

Yağmur, Sarikaya and Najaflou (2023) Researchers also tested the RFE and Boruta feature selection methods to see how well they cut down on the number of features in the data set. In order to make predictions, they employed DL, ensemble techniques, and regular ML. According to the findings, oversampling yields better results than under-sampling over an extended period of time. With an F1 Score of 90%, XGBOOST outperformed the competition [20].

Despite the growing body of research on employee turnover prediction, several gaps remain. Most existing studies focus on improving model performance through advanced algorithms, optimization techniques, or feature selection, but they often rely on limited or domain-specific datasets, which restricts the generalizability of their findings. Many approaches also emphasize accuracy and related metrics without adequately addressing data imbalance issues or exploring model interpretability, which is crucial for practical HR decision-making. Furthermore, while some studies highlight key predictors of attrition, there is limited integration of demographic, job-related, and behavioral factors in a unified framework that can be applied across diverse organizational settings. These gaps highlight the need for robust, scalable, and interpretable ML models that can effectively handle imbalanced HR datasets, capture complex relationships among variables, and provide actionable insights for proactive human resource planning.

The Table I provide a summary of recent studies on employee turnover prediction within HR planning, outlining

the proposed models, datasets utilized, key findings, and the challenges or recommendations identified in each work.

TABLE I. RECENTRECENT STUDIES ON EMPLOYEE TURNOVER WITHIN HR PLANNING

Author	Proposed Work	Dataset	Key Findings	Challenges & Future Work
Fang and Zhang (2025)	Applied GAN for data balancing and RFE for feature selection with RF and XGBoost models	Internal HR dataset + IBM HR dataset	RF with GAN-balanced data and 5 selected features achieved superior performance; XGBoost also showed comparable results	Need to test on larger and more diverse datasets to enhance generalizability
M et al. (2025)	Proposed ML-based employee classification system for attrition prediction	XYZ Corporation dataset	XGBoost achieved best results with high accuracy, precision, and F1-score	Interpretability of models and adaptability to multi-domain datasets require further exploration
Pachica, Fajardo, & Medina (2024)	Developed Adam optimizer with tanh-based warm-up strategy for turnover prediction	HR dataset	Achieved highest accuracy (95%) with faster convergence (50 epochs) compared to other optimizers	Requires validation across different datasets and neural network architectures
Recilla et al. (2024)	Proposed FMSAX operator in GA for turnover prediction	HR dataset	FMSAX operator outperformed other GA crossover operators; identified key turnover predictors (salary, workload, family, career transition)	Broader validation on larger datasets needed; more focus on real-world interpretability
Sharma & Sharma (2023)	Compared Random Forest and Logistic Regression classifiers	Dataset of 1471 rows and 35 features	RF achieved higher accuracy (87%) than LR (85%), with better precision, recall, and F1-score	Limited dataset size; future studies should explore scalability and ensemble techniques
Yağmur, Sarıkaya & Najafloo (2023)	Applied data balancing methods (oversampling, undersampling) and feature selection (RFE, Boruta) across ML, ensemble, and DL models	HR dataset	Oversampling outperformed undersampling; XGBoost achieved best performance (90.90% F1 score)	Future work should integrate hybrid balancing methods and test DL scalability

III. RESEARCH METHODOLOGY

The proposed methodology for predicting employee turnover within HR planning begins with the IBM HR dataset sourced from Kaggle. The data undergoes pre-processing, which includes cleaning, label encoding, and balancing the classes. After pre-processing, the dataset is split into training and testing subsets. Applying XGBoost and RF, two prediction models, to the training data follows. In order to guarantee accurate and strong prediction results, the models are assessed using performance metrics like recall, F1-score, accuracy, precision, and AUC. Finally, the predicted results are generated, enabling informed HR decision-making strategies. The whole process steps are illustrated in Figure 1.

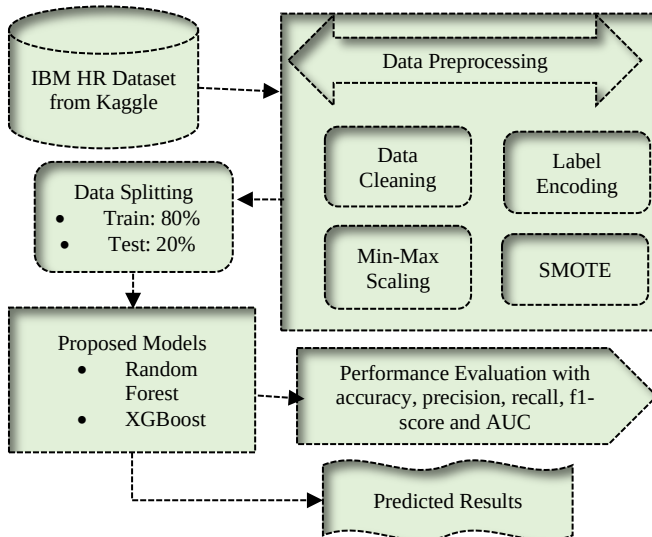


Fig. 1. Proposed Flowchart for Employee Turnover within HR Planning

The following section presents a detailed explanation of each step in the proposed flowchart for predicting employee turnover within HR planning using ML.

A. Data Gathering and Analysis

This study aims to identify the sources of attrition by simulating real-world HR data using a synthetic dataset developed by IBM data scientists. The dataset is derived from Kaggle and has a grand total of 1470 records. Age, gender, marital status, degree of education, and occupation are some of the demographic factors. Some of the employment features include the following: department, job level, years at company, monthly salary, job travel frequency, job title, and distance from home. Evaluating performance, being involved in one's work, being satisfied with one's work environment, having a good work-life balance, and being satisfied with one's relationships are all components of the performance and satisfaction measurements. To further understand the dataset, an exploratory data analysis (EDA) was conducted. The results, shown below, are a collection of graphs that more clearly illustrate employee attrition:

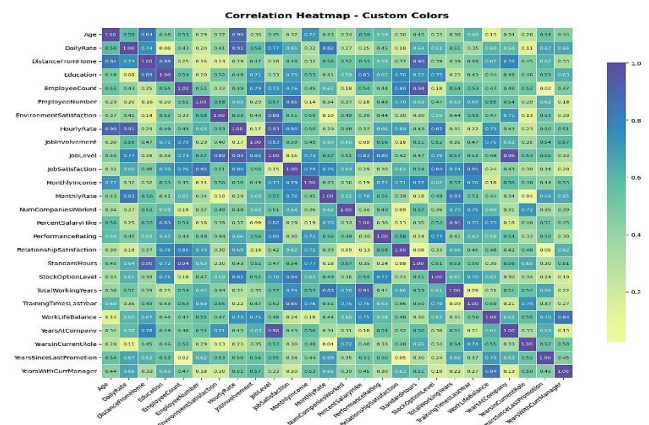


Fig. 2. The Correlation Analysis of Dataset

Figure 2 illustrates a correlation heatmap of employee data factors, where each cell's color and value represent the correlation coefficient between variables. Warmer yellow shades indicate weaker or no correlation (near 0), while cooler blue and purple shades reflect stronger positive or negative

correlations (near 1 or -1). The dark diagonal line shows perfect self-correlation (1.00). This visualization highlights the strongest and weakest linear relationships among variables at a glance.



Fig. 3. The Breakdown of Monthly Revenue Data and the Total Number of Working Years by Employee Attrition

Figure 3 displays a scatter plot illustrating the relationship between monthly income and total working years. The employee attrition axis is highlighted in this image. The red dots reflect employees who departed, while the blue ones show those who remained. The plot helps reveal patterns or clusters where attrition is more prevalent, such as within specific income ranges or career stages.

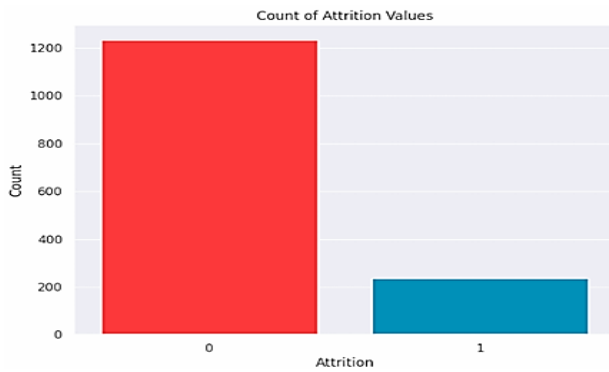


Fig. 4. Imbalanced Count of Attrition Values

A total of '0' indicates that no employees left and a total of '1' indicates that employees departed; this distribution is shown in Figure 4. The chart highlights a clear class imbalance, with significantly more employees in category '0' (over 1200) compared to category '1' (just over 200).

B. Data Pre-Processing

After collecting the raw data, a pre-processing phase is carried out to enhance the reliability of the analysis. Data cleansing, label encoding, min-max scaling, and SMOTE-based class imbalance handling are all part of this process:

- **Data Cleaning:** Looking over the dataset, found certain features that may use some cleaning. Data cleansing is the process of identifying and removing unnecessary attributes from a dataset in order to make it more suitable for targeted analysis.
- **Encoding Categorical Features using Label Encoder:** The dataset's categorical variables were numerically encoded so that ML algorithms could work with them. Used the Label Encoder technique to assign integer labels to each separate category. An

example of a categorical variable is $x_{ij} \in C$, Which may be labelled using Equation (1).

$$x'_{ij} = \text{LabelEncoder}(x_{ij}) \quad (1)$$

where, x_{ij} The original feature value for the i -th Employee and the j -th feature $\text{LabelEncoder}(x_{ij})$ A transformation function that converts categorical values into numeric labels x'_{ij} : The transformed numeric value of the feature after applying label encoding.

- **Data Scaling using Min-Max Scaler:** An important step in solving this issue is standardizing the raw data so that different quantitative units have less of an impact. When it comes to rescaling feature values, normalization is a popular strategy. Equation (2) shows the standard procedure for normalizing numerical features to the interval $[0, 1]$ using MinMax:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

This feature's original value is denoted by x , Its minimum and maximum values throughout the dataset are represented by x_{min} and x_{max} , respectively, and x_{scaled} is the normalized value of x , rescaled to the range $[0, 1]$.

C. Handling Class Imbalance with SMOTE

The main attention is usually on the minority class, yet inaccurate forecasts for that group are prevalent due to class imbalance, a classic problem in classification problems. In this study, the SMOTE [21] was used to fix this problem and ensure the dataset was balanced. Instead of employing traditional oversampling techniques, SMOTE effectively mitigates overfitting risk by creating synthetic instances through interpolation across existing minority class samples. See Figure 5 for a visual representation of the data-driven balanced count of attrition values.

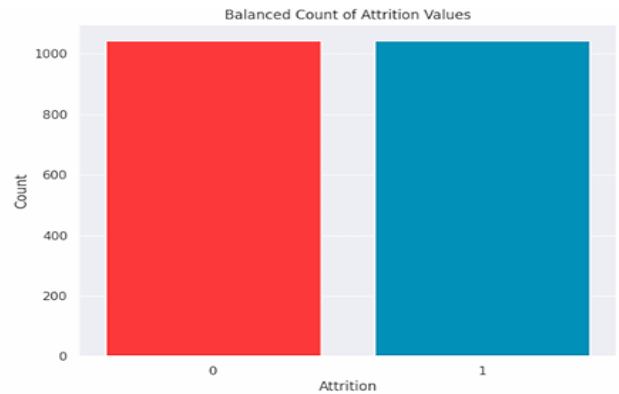


Fig. 5. Balanced Count of Attrition Values

Figure 5 shows the distribution of attrition categories '0' (employees who stayed) and '1' (employees who left), each with nearly equal counts of around 1000. This balanced distribution indicates that the dataset was re-sampled to address class imbalance, ensuring fair representation for both groups and reducing model bias.

D. Data Splitting

The aim of partitioning the dataset was to make the models more generalizable. The proposed model was trained using 80% of the dataset, while 20% was reserved for testing, following an 80:20 ratio.

E. Proposed Machine Learning Strategies

This study proposed the two ML models, XGBoost and Random Forest are discussed below:

1) XGBoost Model

One boosted tree approach that accomplishes its aims by applying the gradient boosting principle is the XGBoost. Compared to other methods, it employs a more regularized model reinforcement strategy to reduce overfitting and improve performance. Because it makes use of parallel tree building, the approach is both fast and resilient to failure in distributed environments. Data is inputted into the classifier in DMatrix format [22]. To optimize memory usage and speed, XGBoost makes use of what is essentially an internal data structure. The gradient boosting (GBM) framework is fundamental to XGBoost. However, it outperforms the GBM framework on its own. Supervised ML tasks are its forte. The subsequent Equation (3) represents it:

$$P(Y = 1|X) = \frac{1}{1+e^{-(F(x))}} \quad (3)$$

The function $F(x)$ takes the results from several decision trees and combines them.

2) Random Forest Classifier

This classifier makes use of a technique called ensemble learning. It uses several decision trees to classify data points on randomly selected subsets of the initial dataset [23]. At the end of the classifier's process, the most popular class is selected. Part one of building a random forest is generating the random forest itself, and part two is making a prediction based on the classifications generated in part one.

$$Gini = 1 - \sum_{i=1}^n (p_i)^2 \quad (4)$$

p_i , In Equation (4), represents the object's classification probability with respect to a given feature or class.

F. Evaluation Metrics

This section displays the outcomes of the XGBoost and Random Forest applications. The precision (P), recall (R), accuracy rate, F1-score, and AUC are some of the metrics used to measure how well these algorithms work, as shown in Equations (5) to (8).

1) Accuracy

The accuracy rate is a measure that determines how many forecasts were right out of all the possible ones. Finding out how accurate the model is at making predictions is a reasonable expectation. The Equation is (5):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

2) Precision(P)

This statistic determines how many positive instance predictions were accurate relative to the overall number of positive instance forecasts. As such, it is crucial to understand how far the model can weed out instances that do not fit the positive category. A classifier's accuracy in predicting positive classes is denoted as (6):

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

3) Recall(R)

The percentage of correctly anticipated true positives as a percentage of all true positives is what the statistic finds. According to Equation (7):

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

4) F1 score

A high F1-score indicates a harmonic mean between recall and precision. Harmonic means place a lot of emphasis on small numbers. As a result, a model's F1-score can be high or low depending on the relative importance of recall and precision. In mathematical terms, it is expressed as (8):

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

where

TP stands for employees who departed from the company as expected and were accurately classified as attrition; TN represents employees who remained and were accurately classified as non-attrition; FP stands for employees who remained but were mistakenly classified as attrition; and FN stands for employees who departed but were wrongly classified as non-attrition.

5) AUC

It is a measure that evaluates how well different models perform on the same dataset. Performance improves as the AUC rises.

IV. RESULTS AND DISCUSSION

A 3.6 GHz Intel Core i7-12700K with 12 cores and 20 threads, 32 GB of DDR4 RAM, and a 10GB VRAM NVIDIA RTX 3080 GPU were all part of the hardware arrangement. Storage was provided by a speedy 1 TB NVMe SSD. Python 3.8 was the main programming language used in the software environment, which was based on Ubuntu 20.04 LTS. The results of the suggested models for forecasting employee turnover are summarized in Table II. In comparison to the RF classifier, XGBoost achieves much higher predictive efficacy with 96.95% accuracy, 97.28% precision, 96.60% recall, and 96.44% F1-score; in comparison, RF achieves 93.90% accuracy, 95.45% precision, 92.20% recall, and 93.79% F1-score.

TABLE II. EXPERIMENTAL RESULTS OF THE PROPOSED ON IBM HR DATASET

Performance matrix	XGBoost model	RF Classifier
Accuracy	96.95	93.90
Precision	97.28	95.45
Recall	96.60	92.20
F1-score	96.94	93.79

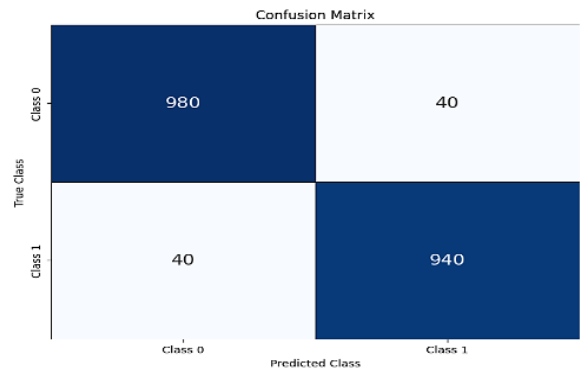


Fig. 6. Confusion Matrix for the XGBoost Model

Figure 6 displays the number of Class 0 and Class 1 occurrences that were accurately predicted by the model. Forty times, the model got Class 0 wrong when it should have

been Class 1, and forty times, it got Class 1 wrong when it should have been Class 0.

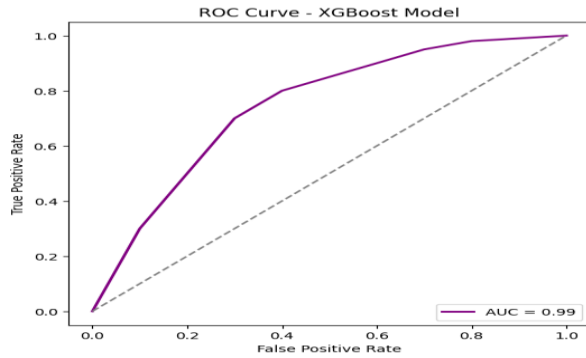


Fig. 7. ROC for the XGBoost Model

Figure 7 shows the ROC curve of the XGBoost model, where the purple line lies well above the diagonal no-discrimination line, indicating strong performance. With an AUC of **0.99**, the model demonstrates excellent discriminatory power and significantly outperforms random classification.

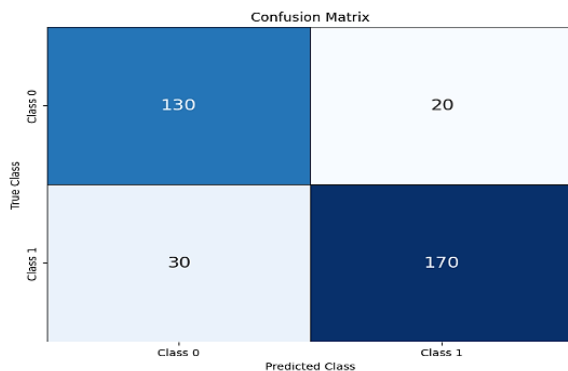


Fig. 8. Confusion Matrix for the RF Model

Figure 8 shows the confusion matrix of the classification model, which shows how well it performed on two classes. In 130 cases, the model got Class 0 right and 170 times, it got Class 1 right. In all, it produced 50 erroneous forecasts, 20 of which were Class 1 misclassifications and 30 of which were Class 0 misclassifications. Although it seems to have a slightly greater rate of misclassifying actual Class 1 cases, the model overall shows strong performance with a high number of right predictions for both classes.

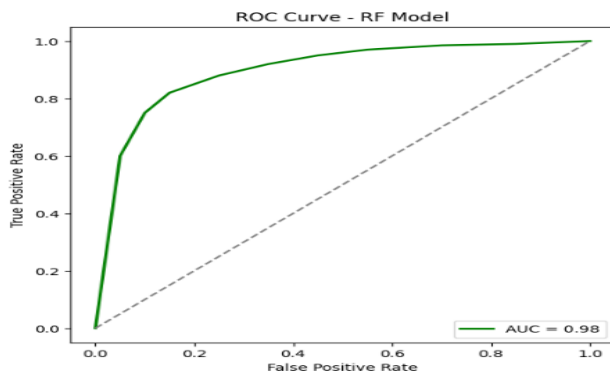


Fig. 9. ROC for the RF Model

Figure 9 displays the RF model's ROC curve, which indicates the trade-off between TPR and FPR across

thresholds. The solid green curve rises sharply from the bottom-left and levels toward the top-right, well above the dashed gray line representing a random classifier. With an AUC of 0.98, the RF model demonstrates excellent discrimination between positive and negative classes.

A. Comparative Analysis

To validate the effectiveness of the proposed XGBoost and RF models, a comparative performance evaluation was carried out against other established predictive models. Table III presents a comparison of predictive models used for employee turnover within HR planning based on their accuracy. The DT model had the worst accuracy at 78.0%, but AdaBoost saw a performance boost to 90.0%. In terms of accuracy, the SVM model came in at 88.6%, while the CNN model reached 84.0%. The RF ensemble method outperformed the others with a 93.90% success rate. But XGBoost was the best model for forecasting staff turnover; it achieved the maximum accuracy of 96.95%.

TABLE III. PERFORMANCE COMPARISON OF DIFFERENT PREDICTIVE MODELS OF EMPLOYEE TURNOVER WITHIN HR PLANNING

Models	Accuracy	Precision	Recall	F1-score
DT [24]	78.0	79.0	75.0	77.0
AdaBoost [25]	90	90.5	90.5	89.5
SVM [26]	88.6	-	-	-
CNN [27]	84.0	-	-	67.2
XGBoost	96.95	97.28	96.60	96.94
RF	93.90	95.45	92.20	93.79

The proposed XGBoost and RF models offer significant advantages for employee turnover prediction within HR planning due to their robustness, scalability, and ability to handle complex, high-dimensional data. Both models effectively manage non-linear relationships and interactions between variables, making them highly suitable for capturing the multifaceted factors influencing employee attrition. Together, these models not only deliver reliable and interpretable predictions but also support informed decision-making in HR management by identifying key patterns and trends that contribute to employee turnover.

V. CONCLUSION AND FUTURE STUDY

Rising rates of employee turnover are a big problem that can affect how well a business does its job. How well the hiring process works affects it because a better hiring process lowers the rate of employee turnover. This study demonstrates the potential of advanced ML models in addressing the critical challenge of employee turnover within HR planning. By leveraging ensemble approaches such as XGBoost and RF, the research highlights how data-driven methods can provide accurate, robust, and interpretable predictions that support proactive decision-making in workforce management. The proposed XGBoost model achieved the highest accuracy of 96.95%, while the RF model also delivered strong performance with 93.90% accuracy. The findings emphasize the importance of integrating demographic, job-related, and satisfaction factors to capture the multifaceted nature of attrition, beyond predictive accuracy.

Future research could expand this work by applying the models to larger and more diverse real-world datasets, incorporating DL approaches for improved pattern recognition, and exploring explainable AI techniques to strengthen interpretability and trust in HR analytics. Additionally, integrating financial, behavioral, and external labor market data may further enrich the predictive

framework, making it more adaptive and applicable across various organizational contexts.

REFERENCES

- [1] P. Mohan and V. A., "A study on factors affecting job satisfaction in the IT industry," *Multidiscip. Sci. J.*, vol. 6, p. 2024223, 2024, doi: 10.31893/multiscience.2024223.
- [2] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.
- [3] W. A. Al-Suraihi, S. A. Samikon, A.-H. A. Al-Suraihi, and I. Ibrahim, "Employee Turnover: Causes, Importance and Retention Strategies," *Eur. J. Bus. Manag. Res.*, vol. 6, no. 3, pp. 1–10, Jun. 2021, doi: 10.24018/ejbm.2021.6.3.893.
- [4] A. R. Bilipelli, "End-to-End Predictive Analytics Pipeline of Sales Forecasting in Python for Business Decision Support Systems," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 819–827, 2022.
- [5] H. Kali, "Diversity, Equity, and Inclusion Analytics in HR: How Workday Enables Data-Driven Decision-Making," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 2, pp. 162–170, 2023, doi: 10.56472/25832646/JETA-V3I6P113.
- [6] A. R. Bilipelli, "Application of AI and Data Analysis for Classification of Student Success in Large-Scale Educational Dataset," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 6, pp. 428–441, Nov. 2024, doi: 10.48175/IJARSCT-22564.
- [7] S. Saliu, "HR Analytics: Leveraging Big Data for Strategic Workforce Planning," *ICONIC Res. Eng. JOURNALS*, vol. 8, no. 9, pp. 742–754, 2025.
- [8] V. Varma, "Data Analytics for Predictive Maintenance for Business Intelligence for Operational Efficiency," *Asian J. Comput. Sci. Eng.*, vol. 7, no. 4, 2022.
- [9] V. Verma, "The Role of Data Migration in Modern Business Intelligence Systems," *Int. J. Res. Anal. Rev.*, vol. 11, no. 2, pp. 1–11, 2024.
- [10] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, 2025.
- [11] H. Kali, "The Future Of Hr Cybersecurity: Ai-Enabled Anomaly Detection In Workday Security.," *Int. J. Recent Technol. Sci. Manag.*, vol. 8, no. 6, 2023, doi: 10.10206/IJRTSM.2025803096.
- [12] L. Covenant, "Predicting Employee Turnover: Using AI and Financial Data to Proactively Address Retention Challenges," 2024.
- [13] A. Brown, N. Davis, O. Miller, E. Wilson, L. Smith, and S. Lopez, "Deep Learning Techniques for Enhancing Employee Turnover Prediction Accuracy," 2024, doi: 10.13140/RG.2.2.15353.28009.
- [14] G. Modalavalasa and H. Kali, "Exploring Big Data Role in Modern Business Strategies: A Survey with Techniques and Tools," *IJARSCT*, vol. 3, no. 3, 2023, doi: 10.48175/IJARSCT-11900B.
- [15] Y. Fang and Z. Zhang, "Employee Turnover Prediction Model Based on Feature Selection and Imbalanced Data Handling," *IEEE Access*, vol. 13, pp. 126736–126755, 2025, doi: 10.1109/ACCESS.2025.3589491.
- [16] V. T. R. P. K. Malladi, V. Nagalakshmi, B. Sriram, and S. A. Wasib, "Improving Workforce Management with Machine Learning: A Novel Approach to Employee Classification," in *2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, IEEE, May 2025, pp. 1411–1415, doi: 10.1109/ICAISS61471.2025.11042033.
- [17] A. O. Pachica, A. C. Fajardo, and R. P. Medina, "Improved Adam Optimizer with Warm-Up Strategy and Hyperbolic Tangent Function for Employee Turnover Prediction Using Multilayer Perceptron (MLP)," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, 2024, pp. 283–288, doi: 10.1109/ICTC62082.2024.10826609.
- [18] V. J. Recilla, M. R. A. Enonaria, R. J. Florida, J. C. M. Bustillo, C. C. Abalorio, and J. C. Trillo, "Predicting Employee Turnover Through Genetic Algorithm," in *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2024, pp. 1383–1391, doi: 10.1109/ICESC60852.2024.10689796.
- [19] S. Sharma and K. Sharma, "Analyzing Employee's Attrition and Turnover at Organization Using Machine learning Technique," in *2023 3rd International Conference on Intelligent Technologies, CONIT 2023*, 2023, doi: 10.1109/CONIT59222.2023.10205676.
- [20] G. Yağmur, B. Sarikaya, and V. N. Najafloo, "Comparative Analysis of Resampling and Feature Selection Methods for Employee Turnover Prediction," in *31st IEEE Conference on Signal Processing and Communications Applications, SIU 2023*, 2023, doi: 10.1109/SIU59756.2023.10224012.
- [21] T. Vaiyapuri and Z. Sbair, "Bayesian Optimized Boosted Ensemble models for HR Analytics - Adopting Green Human Resource Management Practices," *Int. J. Technol.*, vol. 16, no. 2, p. 561, Mar. 2025, doi: 10.14716/ijtech.v16i2.7277.
- [22] R. Jain and A. Nayyar, "Predicting employee attrition using xgboost machine learning approach," in *Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART 2018*, 2018, doi: 10.1109/SYSMART.2018.8746940.
- [23] K. M. Mitravinda and S. Shetty, "Employee Attrition: Prediction, Analysis of Contributory Factors and Recommendations for Employee Retention," in *2022 IEEE International Conference for Women in Innovation, Technology and Entrepreneurship, ICWITE 2022 - Proceedings*, 2022, doi: 10.1109/ICWITE57052.2022.10176235.
- [24] H. Patil and P. Kadam, "Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance," *Voice Creat. Res.*, vol. 7, no. 2, pp. 295–301, Apr. 2025, doi: 10.53032/tvcr/2025.v7n2.37.
- [25] K. Konar, S. Das, S. Das, and S. Misra, "Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI," *SN Comput. Sci.*, vol. 6, no. 6, p. 672, Jul. 2025, doi: 10.1007/s42979-025-04204-w.
- [26] R. Govindarajan, N. K. Kumar, P. Sudhakar Reddy, E. Sai Pravalika, B. Dhatri, and G. Pavan Kumar, "Predicting Employee Attrition: A Comparative Analysis of Machine Learning Models Using the IBM Human Resource Analytics Dataset," *Procedia Comput. Sci.*, vol. 258, pp. 4084–4093, 2025, doi: 10.1016/j.procs.2025.04.659.
- [27] N. Ben Yahia, J. Hlel, and R. Colomo-Palacios, "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction," *IEEE Access*, vol. 9, pp. 60447–60458, 2021, doi: 10.1109/ACCESS.2021.3074559.