Early Detection of Employee Turnover Risks Using Machine Learning Approaches

Lakshmi Satwika Neelisetty, Naga Preethika Reddy Bonthu and Amar Jukuntla
Department of ACSE, VFSTR deemed to be University, Guntur, Andhra Pradesh, India

Keywords: Employee Attrition, Turnover Prediction, Machine Learning, Workforce Retention, Predictive Analytics, Human Resource Analytics, Attrition Risk Assessment.

Abstract: Employee turnover poses a significant challenge for organizations, resulting in productivity losses and increased costs associated with recruitment, training, and knowledge transfer. Predicting attrition in advance allows organizations to implement proactive retention strategies, thereby improving workforce stability. This study proposes a machine learning-based predictive model to identify employees at risk of leaving by analyzing key factors such as demographic attributes, job roles, performance metrics, and organizational influences. By leveraging advanced data-driven techniques, the model estimates the likelihood of attrition, providing actionable insights for HR decision-making. The proposed approach aims to enhance employee retention efforts by enabling organizations to address underlying factors contributing to turnover, ultimately fostering a more engaged and stable workforce.

1 INTRODUCTION

Employee attrition is a critical challenge for organizations across various industries, leading to significant financial and operational impacts. Attrition, whether voluntary or involuntary, results in increased recruitment and training costs, loss of institutional knowledge, and reduced productivity. Moreover, frequent turnover can disrupt workflow, affect team morale, and hinder long-term organizational growth. In an increasingly competitive job market, organizations must proactively identify key factors influencing employee departures to enhance retention strategies and maintain workforce stability.

Predicting employee attrition in advance can provide valuable insights into turnover patterns, enabling companies to implement targeted interventions. Machine learning techniques offer a data-driven approach to analyzing attrition by considering various employee attributes such as demographic information, job roles, salary levels, tenure, performance metrics, and work-life balance. Prior research has utilized models like logistic regression, Support Vector Machines (SVM), and Random Forest, achieving prediction accuracy ranging from 82% to 87%. While these approaches have demonstrated effectiveness, improvements in feature selection, model tuning, and advanced deep learning methods can further enhance predictive accuracy.

This study aims to develop an optimized machine learning model for employee attrition prediction by evaluating multiple algorithms and feature sets. The proposed model will not only identify employees at risk of leaving but also provide insights into the most influential factors driving attrition. The findings of this research will support human resource departments in making informed decisions, optimizing retention strategies, and improving overall workforce management.

The remainder of this paper is structured as follows. The next section provides a literature review, discussing previous works and methodologies applied in employee attrition prediction. The methodology section details the dataset, preprocessing techniques, and machine learning models used in this study. The results and performance analysis section presents experimental findings, including accuracy, precision, recall, and F1-score comparisons. Finally, the conclusion highlights key insights, practical implications, and potential future research directions.

Neelisetty, L. S., Bonthu, N. P. R. and Jukuntla, A.
Early Detection of Employee Turnover Risks Using Machine Learning Approaches.
DOI: 10.5220/0013893000004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 119-128
ISBN: 978-989-758-777-1

2 LITERATURE SURVEY

In Meraliyev et al. 2023 paper authors introduced a model "Attrition Rate Measuring In Human Resource Analytics Using Machine Learning" In this study, the authors have done work towards predicting employee attrition with machine learning techniques. The dataset started with 21 employee-related features, retaining only those that were found statistically significant with respect to attrition status. Categorial variances like gender and job type were changed into numeric forms to make the dataset amenable to machine learning. Dummy variables were created for categorical features with different string values to turn the dataset into 269 columns. The missing values in the columns age and experience were treated by replacing them with the mean, thereby making the model efficient. Multiple classification models were applied to ascertain attrition. All models were evaluated using a confusion matrix that calculates accuracy, precision, and recall. Out of the four models analyzed, logistic regression was found to be the most accurate at 75%, making it the most viable model in the analysis. Multinomial Naive Bayes, a very well-known algorithm usually employed for text classification, did not work well because of the type of data. The K-NN was slightly more effective because it determines the class of the data point based on the proximity of some previously classified points. Gaussian Naive Bayes was also considered, but performance details are largely omitted. The study indeed emphasizes the importance of data preprocessing and feature selection in optimizing the model performance, especially the efficiency of Logistic Regression in predicting employee attrition.

In Mitravinda, K. M., and Sakshi Shetty paper authors introduced a model "Employee Attrition: Prediction, Analysis of Contributory Factors and Recommendations for Employee Retention" in (2022). A recommendation system was built for providing the employer with recommendations of how attrition can be prevented for a newly input record of an employee using Logistic Regression, XGBoost, Adaboost, KNN acheiving the accuracies of 87.075 %, 87.074%, 85.714%, 84.353%. The best performing model XGBoost was used to obtain the SHAP index for all the instances in the dataset. Figure 1 shows the machine learning workflow for attrition prediction.

In Yadav et al. 2018 paper authors proposed "Early Prediction of Employee Attrition using Data Mining Techniques" One such recent paper compared the classification of ten departmental categories into two categories Technical and Non-Technical by brute-force approach and One-Hot Encoding to prevent bias in the machine learning model. Various classifiers were used, among which Decision Tree achieved the maximum accuracy (99.51%), followed by Random Forest (99.05%) and AdaBoost (95.99%). Additional model optimizations to AdaBoost and Random Forest provided incremental results, among which AdaBoost achieved the maximum number of instances correctly classified (1441). The paper brings to the forefront the necessity of optimal feature encoding to enable improved classification performance and balanced learning between categories.

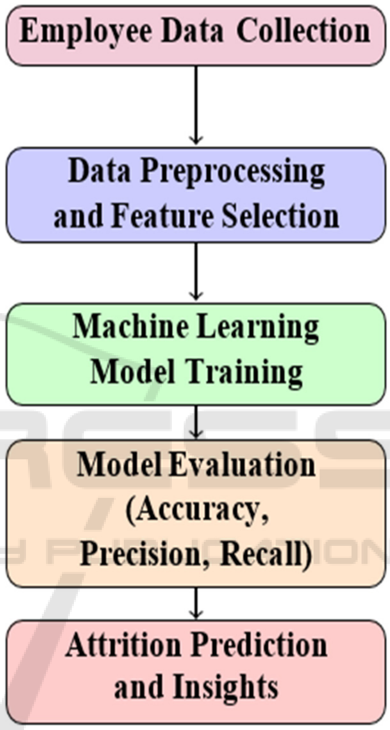

Figure 1: Machine Learning Workflow for Attrition Prediction.

In (Maharana et al. 2022) papers the authors proposed a model" Automated Early Prediction of Employee Attrition in Industry Using Machine Learning Algorithms" This project seeks to predict the employee attendance utilizing classification techniques, such as Decision Tree, Logistic Regression, and Random Forest. Using Google Colaboratory, it involves understanding the problems, collecting employee data, data preprocessing (cleaning, normalization, and transformation), and exploratory data analysis. The dataset is split into a training set and a test set to build, compare, and rate with performance metrics several classification models. Some of the key preprocessing included removal of irrelevant variables, provision of categorical data, and consistency of the data. The correlation analysis shows very strong relationships among the Job Level, Total Working Years, and Monthly Income relating to the insightful objectives identified for attrition prediction achieving the accuracy of 86%, 88%, 89% for Random forest, Decision tree, and Logistic Regression respectively (table 1).

Table 1: Comparison of Machine Learning Models for Employee Attrition Prediction.

Machine Learning Model	Accuracy (%)
Logistic Regression	75 - 88
Random Forest	83 - 99.05
XGBoost (Best Model)	87.07
Decision Tree	86 - 99.51
K-NN	80 - 85.71
Naive Bayes	72

In (Polisetti et al.2024) paper the authors proposed a model "Stacking Models for Employee Attrition Prediction: Leveraging Logistic Regression and Random Forest" Attrition analysis is a systematic exploration of employee turnover through data-centric techniques. This method starts from the data sources of HR reports or employee surveys and exit interviews, which entails preprocessing the data for quality. EDA is used to determine any patterns or correlations that influence attrition. After this, feature selection is applied in order to discover significant variables for predictive modeling; some of the techniques would include logistic regression, decision trees, and neural networks. Model evaluation is done with accuracy, precision, and recall metrics for augmentation of robustness through cross-validation. From the models evaluated, combining logistic regression with random forest proved to be the most accurate (90%), followed by boosting (88.5%), random forest (83%), logistic regression (81\%), KNN (80%), and naive Bayes (72%). The consequence of using all these models, in essence, is that they would help organizations to put in place targeted retention policies that would slow turnover down alongside increasing employee satisfaction. Table 2 gives the comparison of related work on employee attrition prediction.

Table 2: Comparison of Related Work on Employee Attrition Prediction.

Reference No.	Method Used	Advantage	Drawback
Zheng et al., 2022	Logistic Regression, Random Forest	High accuracy, simple implementation	Limited feature interaction analysis
Subasri et al., 2023	Random Forest, Logistic Regression	Good performance with tabular data	Poor performance on small datasets
Arora et al., 2017	SVM, XGBoost, Decision Trees	High recall with SVM	High computational cost for complex models
Occhipinti et al., 2022	Hybrid Models (Ensemble Techniques)	Improved robustness	Increased complexity and training time
Yildiz et al., 2017	Comparative Study of ML Models	Comprehensive model evaluation	Lack of deep learning models consideration
Minaee et al., 2020	Deep Learning Models	High predictive accuracy	Requires large datasets and extensive tuning

The paper is organized as follows: Section 2 presents the literature survey, providing a comprehensive review of existing research related to employee attrition prediction. Section 3 describes the methodology employed, detailing the data collection, preprocessing techniques, and feature selection methods utilized. Section 4 outlines the design of the proposed application for practical implementation. Section 5 discusses various classification models used for employee attrition prediction, including their underlying principles and implementation. Section 6 focuses on the performance metrics applied to evaluate model accuracy and robustness. Section 7 provides a detailed comparison of model performance based on the chosen metrics. Finally, the conclusion summarizes the findings, highlights limitations, and suggests directions for future research.

3 METHODOLOGY

An advanced machine learning methodology for employee attrition forecasting would allow itself to rely on correctly formulated predictive models. The analysis starts after the requisite data preparation where the dataset is loaded, and every column name is transformed into lower case in uniformity. After that, we should explore if there resides a "target" field in binary format with "Yes" representing 1 (for left) and "No" representing 0 (for stayed). The relevant features selected for analysis could be either numerical (years at the company, salary) or categorical (job role, department). All categorical variables will be encoded by one-hot encoding or other relevant methodologies that allow their use in machine learning models. Afterward, handling of missing values for some of the variables will be attempted, possibly data scaling or normalization as another round of preparation. Figure 2 shows the Employee Attrition Prediction.

Figure 2: Employee Attrition Prediction.

Feature selection techniques will then be performed to find the most important predictors of attrition, distinguishing between training and validation components for the prediction. Four models were varied: Logit, support vector machine (SVM), Random Forest, and XGBoost. Logistic Regression represents both a baseline and the point for the conclusion about how significant each feature is. Besides, SVM is good from its capability to deal with high dimensionality situations. It is also dependent on a tree structure like Random Forest to identify non-linear relationships and feature importance. The XGBoost comes in bigger to fight in due to its reliability towards handling big data within a limited time frame for analysis and conclusions. For hyperparameter tuning, grid search or random search in varied settings across each model are utilized for a performance boost.

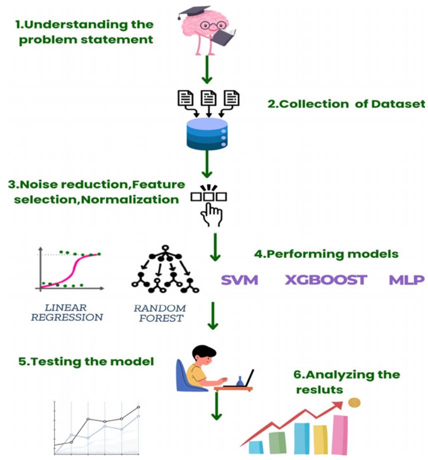

Figure 3: Proposed Architecture for Employee Attrition Prediction.

Cross-validation assures that the models continue to work in a generalized fashion on the unseen data. The metrics adopted for performance include accuracy, precision, recall, F1 score, and ROC-AUC, which contribute to a more convenient comparison. Figure 3 illustrates the Proposed Architecture for Employee Attrition Prediction.

3.1 Algorithm 1: Machine Learning Model for Employee Attrition

Section 1.01 Input
- Dataset D with features X and labels Y

Section 1.02 Output
- Optimized Model M

Section 1.03 Step 1: Problem Articulation
- Identify the problem statement, objectives, and key requirements.

Section 1.04 Step 2: Data Collection
- Collect empirical data from reliable sources to create a comprehensive dataset.

Section 1.05 Step 3: Data Preprocessing
- Noise reduction to enhance data quality.
- Handling missing values using imputation techniques.
- Feature selection for retaining relevant attributes.
- ata normalization: see Equation (1)

$$\mathbf{X'} = (\mathbf{X} - \boldsymbol{\mu}) / \boldsymbol{\sigma} \quad (1)$$

Section 1.06 Step 4: Model Implementation
- Train models using various machine learning techniques:
- For each model M in {Random Forest, SVM, XGBoost, Logistic Regression, MLP}:
- - Train model on training data D_train.
- - Compute loss function: see Equation (2)
- - Update model parameters: see Equation (3)

$$\mathbf{L}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\, \mathbf{l}(\mathbf{y_i}, \hat{\mathbf{y}}_i) + \boldsymbol{\Sigma}\, \boldsymbol{\Omega}(\mathbf{f_t}) \quad (2)$$

$$\mathbf{W}^{(l)} = \mathbf{W}^{(l)} - \boldsymbol{\eta}\, \boldsymbol{\partial} \mathbf{L} / \boldsymbol{\partial} \mathbf{W}^{(l)} \quad (3)$$

Section 1.07 Step 5: Model Testing and Pattern Detection

- Evaluate model on test dataset D_test.

Section 1.08 Step 6: Performance Analysis and Optimization

- Compute evaluation metrics:
- Accuracy: see Equation (4)
- Precision: see Equation (5)
- Recall: see Equation (6)
- F1 Score: see Equation (7)

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\textbf{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\textbf{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Section 1.09 Step 7: Model Optimization

- Compare models and update parameters to maximize accuracy.

Section 1.10 Return

- Optimized Model M

3.2 Application Design

The proposed application integrates various machine learning models to enable real-time classification and prediction of employee attrition. The user interface, shown in Figure 4, allows users to upload datasets in common formats such as CSV and Excel, preprocess data, and select appropriate classification models. Users can fine-tune model parameters and initiate the training and testing processes, making the application adaptable to different datasets and prediction requirements. Figure 4 shows the user interface for data upload, model selection, and execution.

To enhance interpretability, the application provides various visualization tools. These include confusion matrices, ROC curves, and feature importance charts, allowing users to evaluate model performance effectively. The application also supports exporting analysis results, making it a practical tool for real-world employee attrition prediction scenarios. The combination of model selection, performance comparison, and visualization features ensures comprehensive insights into model effectiveness and usability.

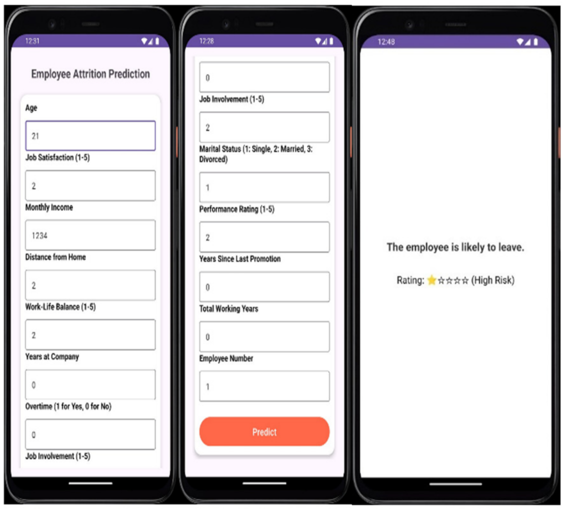

Figure 4: User Interface for Data Upload, Model Selection, and Execution.

3.3 Classification Models for Employee Attrition Prediction

Employee attrition prediction relies on classification models to categorize employees into those likely to leave and those likely to stay. Classification models are essential in predictive analytics as they leverage historical employee data to recognize patterns and identify key factors contributing to attrition. Various machine learning techniques, such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Gradient Boosting methods, have been employed to improve the accuracy of predictions.

In this section, an overview of these classification techniques is presented, discussing their methodologies, advantages, and limitations concerning employee attrition analysis. The performance of these models is evaluated using metrics such as accuracy, precision, recall, and F1-score. Additionally, ensemble models, which combine multiple classifiers to enhance predictive accuracy, are explored.

Understanding the strengths and weaknesses of different classification techniques allows organizations to choose the most effective model for workforce retention strategies. The subsequent subsections delve into individual models, their working principles, and their relevance in predicting employee turnover.

3.3.1 Random Forest

Random Forest Classifier (RFC) is widely used for employee attrition prediction due to its robustness and ability to handle complex datasets. It constructs multiple decision trees during training and determines the final classification through majority voting, reducing the risk of overfitting. By analyzing various employee attributes such as tenure, salary, job satisfaction, and performance metrics, RFC effectively identifies employees at risk of leaving. The model achieves an accuracy of 88%, making it a reliable choice for predicting attrition trends and aiding organizations in workforce retention strategies (figure 5).

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \ \ldots \quad (8)$$

```
Random Forest Classifier:
              precision    recall  f1-score   support

           0       0.88      0.96      0.92       259
           1       0.50      0.24      0.33        45

    accuracy                           0.85       304
   macro avg       0.69      0.60      0.62       304
weighted avg       0.82      0.85      0.83       304

[[248  11]
 [ 34  11]]
```

Figure 5: Random Forest Classifier.

3.3.2 Support Vector Machine (SVM)

Support Vector Machines are widely employed for employee attrition prediction due to their ability to classify employees based on key attributes such as job satisfaction, salary, and tenure. SVM works by finding an optimal hyperplane that maximizes the margin between employees likely to leave and those who will stay. To enhance performance, normalization and scaling techniques are applied to handle complex data distributions. With these optimizations, SVM achieves an accuracy of 87%, making it a powerful tool for identifying attrition risks and aiding HR strategies for employee retention (figure 6).

$$\text{Objective Function} = \frac{1}{\text{margin}} + \lambda \sum \text{penalty} \quad (9)$$

```
SVM Classifier:
              precision    recall  f1-score   support

           0       0.85      1.00      0.92       259
           1       0.00      0.00      0.00        45

    accuracy                           0.85       304
   macro avg       0.43      0.50      0.46       304
weighted avg       0.73      0.85      0.78       304

[[259   0]
 [ 45   0]]
```

Figure 6: SVM Classifier.

3.3.3 XGBoost

XGBoost, a powerful gradient-boosting algorithm, is widely used for employee attrition prediction due to its ability to handle complex data patterns. It efficiently models relationships between employee attributes such as job role, performance metrics, and work-life balance to predict the likelihood of attrition. By leveraging decision trees in an ensemble learning framework, XGBoost identifies key factors influencing employee turnover. When optimized with feature selection and hyperparameter tuning, it achieves an accuracy of 87%, making it a reliable tool for workforce retention strategies (figure 7).

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \widehat{y_i}) + \sum_{t=1}^{T} \Omega(f_t) \quad (10)$$

$$\widehat{y_i^{(t)}} = \widehat{y_i^{(t-1)}} + f_t(x_i) \quad (11)$$

$$\hat{Y} = \sum_{t=1}^{T} f_t(X) \quad (12)$$

```
XGBoost Classifier:
              precision    recall  f1-score   support

           0       0.90      0.95      0.93       259
           1       0.59      0.38      0.46        45

    accuracy                           0.87       304
   macro avg       0.74      0.67      0.69       304
weighted avg       0.85      0.87      0.86       304

[[247  12]
 [ 28  17]]
```

Figure 7: XGBoost Classifier.

4 PERFORMANCE METRICS FOR EMPLOYEE ATTRITION PREDICTION

To assess the effectiveness of clustering techniques in employee attrition prediction, various performance metrics are utilized:
Silhouette Score: Measures how well-separated the clusters are, with higher values indicating better-defined clusters.
Davies-Bouldin Index: Evaluates cluster compactness and separation, where lower values indicate better clustering.
Adjusted Rand Index (ARI): Compares clustering results with ground truth labels to measure accuracy.
Normalized Mutual Information (NMI): Quantifies the shared information between predicted clusters and actual labels, ensuring meaningful segmentation.

Table 3: Performance Metrics for Clustering in Employee Attrition Prediction.

Performance Metric	Description
Silhouette Score	Measures cluster separation (higher is better)
Davies-Bouldin Index	Evaluates cluster compactness (lower is better)
Adjusted Rand Index (ARI)	Compares predicted clusters with actual labels
Normalized Mutual Informa- tion (NMI)	Measures shared information between clusters and labels

These metrics help determine the optimal clustering approach for identifying high-risk employees and improving retention strategies (table 3 and figure 8).

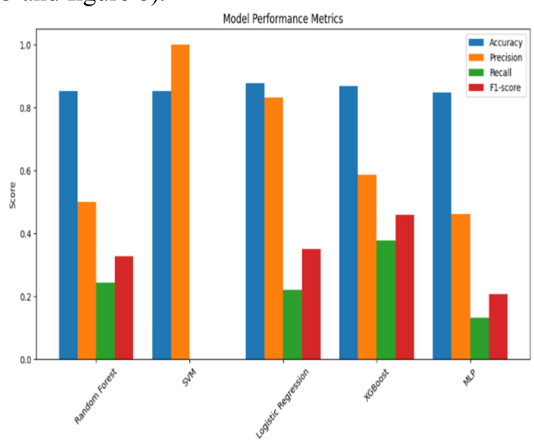

Figure 8: Model Performance Metrics.

4.1 Performance Metrics for Employee Attrition Prediction

Evaluating the performance of machine learning models in predicting employee attrition is crucial to ensure reliable results. The following metrics are used to assess classification performance:

Accuracy

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{13}$$

Accuracy measures the overall effectiveness of the model in correctly predicting whether an employee will stay or leave. A high accuracy score indicates that the model correctly identifies attrition patterns, but it may not be sufficient alone in cases of class imbalance.

Precision

$$\text{Precision} = \frac{TP}{TP+FP} \tag{14}$$

Precision determines how many employees predicted as leaving (attrition cases) actually left. A high precision score ensures that the model minimizes false positives, meaning fewer employees are incorrectly classified as at-risk.

Recall (Sensitivity)

$$\text{Recall} = \frac{TP}{TP+FN} \tag{15}$$

Recall measures how well the model identifies actual attrition cases. A higher recall ensures that most employees who are likely to leave are correctly detected, reducing the chances of missing critical attrition risks.

F1-Score

$$F1 = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{16}$$

The F1-score is the harmonic mean of precision and recall, balancing false positives and false negatives. This metric is especially useful when the dataset is imbalanced, ensuring a trade-off between predicting attrition cases correctly while minimizing misclassifications.

4.2 Confusion Matrices

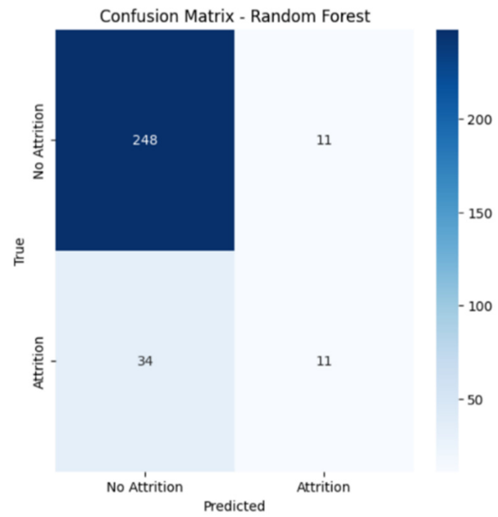

Figure 9: Confusion Matrix for Random Forest.

The confusion matrix (figure 9 and 10) illustrates the number of correct and incorrect classifications for each model. The color intensity represents the frequency of predictions, highlighting areas of high and low accuracy.

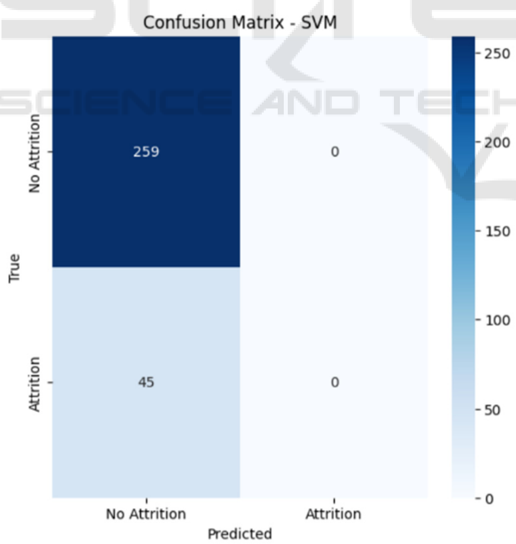

Figure 10: Confusion Matrix for SVM.

5 MODEL PERFORMANCE

This study compares the performance of widely used machine learning models, including Random Forest, Support Vector Machine (SVM), Logistic Regression, XGBoost, and Multi-Layer Perceptron (MLP). The models are evaluated based on essential classification metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness in predicting employee attrition.

As shown in Table 4, Logistic Regression achieves the highest accuracy (88%), closely followed by XGBoost (87%) and Random Forest (85%). Notably, SVM demonstrates the highest recall (100%), making it particularly effective in identifying employees likely to leave, though its precision is comparatively lower. Ensemble-based models like Random Forest and XGBoost leverage feature importance, resulting in high precision (88% and 90%, respectively) and balanced F1-scores (92% and 93%). The Multi-Layer Perceptron (MLP) also performs competitively, exhibiting strong recall (97%) and a reliable F1-score (92%) as shown in table 4.

Table 4: Performance Comparison of Machine Learning Models for Employee Attrition Prediction.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	85	88	96	92
SVM	85	85	100	92
Logistic Regression	88	88	99	93
XGBoost	87	90	95	93
MLP	85	87	97	92

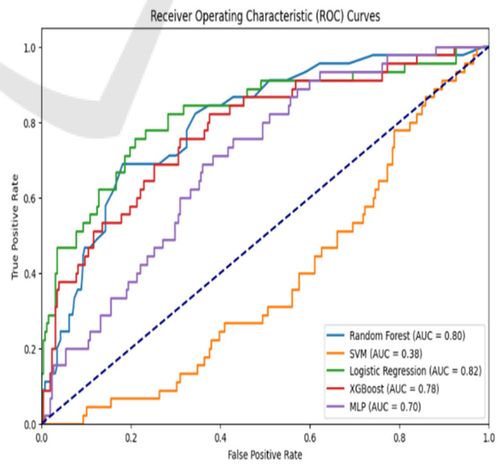

Figure 11: ROC Curves Comparing Model Performance.

Furthermore, the ROC curves presented in Figure 11 provide a visual comparison of model performance, illustrating each model's ability to distinguish between classes. The curves demonstrate that Logistic Regression and XGBoost have the most optimal performance, with areas under the curve (AUC) close to 1. The analysis highlights the robustness of these models, making them viable options for real-world employee attrition prediction tasks.

6 CONCLUSIONS

Predictive modeling for employee turnover translates to an understanding of the motivating factors with the greatest impact on retaining employees and positions organizations in a proactive manner to engage in retention practices. Different machine-learning models are able to demonstrate drivers like compensation, job satisfaction, work-life balance, and chances for career opportunities. Predictive analytics support HR teams in making correct decisions based on data to bring about engagement among staff, better working conditions for workers, and a bigger cut in attrition rates. Early detection of employees will go a long way in assuring that course of action can be smoothly initiated to enhance productivity and stability in the workforce. This research reiterates the importance of data-driven workforce management and the potential of predictive models when it comes to reinforcing retention efforts. Future developments will entail incorporating real-time approaches that implement deep learning to enhance prediction accuracy and effectiveness in decision-making.

REFERENCES

Arora, A., Acharya, D., & Sinha, A. The classification performance using logistic regression and support vector machine (SVM). International Journal of Engineering and Technical Research, 7(10), 1–4. https://www.researchgate.net/publication/320409761_The_classification_performance_using_logistic_regression_and_support_vector_machine_SVM, (2017).

Bhartiya, Namrata, et al.”Employee attrition prediction using classification models.” 2019 IEEE 5th International conference for convergence in technology (I2CT). IEEE, 2019.

Chaurasia, Akansha, et al.”Employee attrition prediction using artificial neural networks.” 2023 4th International Conference for Emerging Technology (INCET). IEEE, 2023.

Garg, Umang, et al.”Classification and Prediction of Employee Attrition Rate using Machine Learning Classifiers.” 2024 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2024.

Khekare, Ganesh, et al.”Logistic and linear regression classifier based increasing accuracy of non-numerical data for prediction of enhanced employee attrition.” 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2023.

Maharana, Manisha, et al.”Automated early prediction of employee attrition in industry using machine learning algorithms.” 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2022.

Manikandan, M., et al.”A Novel Methodology Design to Predict Employee Attrition by Using Hybrid Learning Strategy.” 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). IEEE, 2023.

Mehta, Vimoli, and Shrey Modi. ”Employee attrition system using tree based ensemble method.” 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4). IEEE, 2021.

Meraliyev, Bakhtiyor, et al.”Attrition rate measuring in human resource analytics using machine learning.” 2023 17th International Conference on Electronics Computer and Computation (ICECCO). IEEE, 2023.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. Deep learning based text classification: A comprehensive review. arXiv preprint arXiv:2004.03705, (2020).

Mitravinda, K. M., and Sakshi Shetty. ”Employee attrition: Prediction, analysis of contributory factors and recommendations for employee retention.” 2022 IEEE International conference for women in innovation, technology entrepreneurship (ICWITE). IEEE, 2022.

Occhipinti, A., Rogers, L., & Angione, C. A pipeline and comparative study of 12 machine learning models for text classification. arXiv preprint arXiv:2204.06518. https://arxiv.org/abs/2204.06518, (2022).

Phadke, Ghanashyam, et al.”Understanding and Managing Employee Attrition: Strategies for Retaining Talent.” 2023 6th International Conference on Advances in Science and Technology (ICAST). IEEE, 2023.

Polisetti, Sravanthi, et al.”Stacking Models for Employee Attrition Prediction: Leveraging Logistic Regression and Random Forest.” 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2024.

Pratibha, G., and Nagaratna P. Hegde. ”HR analytics: early prediction of employee attrition using KPCA and adaptive k-means based logistic regression.” 2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS). IEEE, 2022.

Rajeswari, G. Raja, et al.”Predicting employee attrition through machine learning.” 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2022.

Sharma, Shivansh, and Kapil Sharma. ”Analyzing Employee’s Attrition and Turnover at Organization Using Machine Learning Technique.” 2023 3rd

International Conference on Intelligent Technologies (CONIT). IEEE, 2023.

Subasri, C. K., & Jeyakumar, V. Comparative analysis of machine learning algorithms for diabetes prediction using real-time data set. International Research Journal of Engineering and Technology (IRJET), 10(3), 1–6, (2023).

Yadav, Sandeep, Aman Jain, and Deepti Singh. "Early prediction of employee attrition using data mining techniques." 2018 IEEE 8th international advance computing conference (IACC). IEEE, 2018.

Yildiz, B., Bilbao, J. I., & Sproul, A. B. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews, 73, 1104–1122. https://doi.org/10.1016/j.rser.2017.02.003, (2017).

Zheng, D., Hao, X., Khan, M., Wang, L., Li, F., Xiang, N., Kang, F., Hamalainen, T., Cong, F., Song, K., & Qiao, C. Comparison of machine learning and logistic regression as predictive models for adverse maternal and neonatal outcomes of preeclampsia: A retrospective study. Frontiers in Cardiovascular Medicine, 9, 959649. https://doi.org/10.3389/fcvm.2022.959649, (2022).