Employee Attrition Prediction Using Supervised Machine Learning Techniques

Voleti
Data Science
Fordham University
New York City, USA
av106@fordham.edu

***Abstract*—Employee attrition presents a persistent challenge for organizations due to its financial cost, operational disruption, and impact on workforce morale. Proactively identifying employees who are at risk of leaving can enable organizations to implement targeted retention strategies before turnover occurs. This project formulates employee attrition as a supervised binary classification problem using real-world human resources data. The IBM HR Analytics Employee Attrition dataset from Kaggle is used to train and evaluate multiple machine learning models, including Logistic Regression, Random Forest, and XGBoost. Model performance is evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics to account for class imbalance. Results show that tree-based models achieve higher overall accuracy, while Logistic Regression demonstrates stronger recall for attrition cases. Feature importance analysis reveals that attrition is influenced by a combination of workload, compensation, job satisfaction, tenure, and work-life balance. These findings support data-driven, targeted retention strategies rather than uniform organizational policies.**

Keywords—Employee Attrition, Machine Learning, Classification, Human Resources Analytics, Model Interpretation

I. INTRODUCTION

Employee retention has become a critical concern across industries as organizations face increasing competition for skilled talent. High attrition rates lead to increased recruitment and onboarding costs, loss of institutional knowledge, and reduced productivity. Beyond financial impact, frequent employee turnover can negatively affect team morale and long-term organizational stability.

Traditional approaches to addressing attrition often rely on retrospective analysis or generalized human resource policies. However, these approaches may fail to identify at-risk employees early enough to intervene effectively. Advances in machine learning provide an opportunity to shift from reactive analysis to proactive prediction by leveraging historical employee data.

This project applies supervised machine learning techniques to predict employee attrition using real-world HR data. Attrition is formulated as a binary classification problem, where the objective is to predict whether an employee is likely to leave the organization based on demographic, job-related, and satisfaction-based features. In addition to predictive accuracy, emphasis is placed on interpretability and business relevance to ensure that model outputs can support actionable retention strategies.

II. PROBLEM FORMULATION AND OBJECTIVE

The core problem addressed in this project is the prediction of employee attrition prior to employee departure. From a machine learning perspective, this problem is framed as a binary classification task with two possible outcomes: attrition (“Yes”) or retention (“No”).

a. Objectives

The primary objective of this project is to develop a supervised machine learning model capable of predicting employee attrition using historical human resources data. To achieve this goal, multiple classification models, including Logistic Regression, Random Forest, and XGBoost, are trained and compared to evaluate their relative performance. Model evaluation is conducted using performance metrics that appropriately account for class imbalance, ensuring that attrition cases are accurately identified. In addition to predictive performance, model outputs are interpreted to identify the most influential factors contributing to employee attrition. Finally, the technical findings are translated into practical business insights and actionable retention recommendations, ensuring that the results are both interpretable and applicable in real organizational settings.

III. DATASET DESCRIPTION

The dataset used in this project is the IBM HR Analytics Employee Attrition and Performance dataset, obtained from Kaggle. The dataset contains 1,470 employee records and 35 features capturing a combination of demographic attributes, job-related characteristics, compensation details, and satisfaction metrics. The target variable is employee attrition, represented as a binary outcome indicating whether an employee left the organization.

The dataset was selected because it represents real-world HR data and is well-suited for supervised classification tasks. It includes a diverse set of features such as age, job role, monthly income, overtime status, work-life balance, and years at the company. These variables provide meaningful signals for modeling employee behavior and attrition risk. Table I provides a representative snapshot of the IBM HR Analytics Employee Attrition dataset, illustrating the structure and diversity of employee attributes used in this project.

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisf
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	

lours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
80	0	8	0	1	6	4	0	5
80	1	10	3	3	10	7	1	7
80	0	7	3	3	0	0	0	0
80	0	8	3	3	8	7	3	0
80	1	6	3	3	2	2	2	2

Table I. *Sample Records from the IBM HR Analytics Employee Attrition Dataset*

IV. EXPLORATORY ANALYSIS & DATA PREPROCESSING

Initial exploratory data analysis revealed a significant class imbalance in the target variable, with approximately 16 percent of employees labeled as having left the organization. This imbalance motivated the use of evaluation metrics beyond accuracy and informed model design decisions. Several non-informative or constant features, such as employee count and standard hours, were removed to reduce noise.

Categorical variables were encoded using one-hot encoding, while numerical features were scaled using standardization to ensure compatibility across models. The dataset was split into training and testing sets using a stratified approach to preserve class proportions. This preprocessing pipeline was implemented using scikit-learn's ColumnTransformer and Pipeline utilities to ensure reproducibility and consistency across models.

Three supervised learning models were selected for comparison. Logistic Regression was chosen as a baseline due to its interpretability and effectiveness in binary classification problems. Random Forest was selected as a tree-based ensemble method capable of capturing nonlinear relationships and feature interactions. XGBoost was included as a gradient boosting model known for strong predictive performance on structured tabular data. All models were trained using the same preprocessing pipeline to ensure a fair comparison.

Class imbalance was addressed through model-specific weighting strategies. Model evaluation focused on both discriminatory power and the ability to correctly identify attrition cases. To further mitigate the impact of class imbalance, stratified sampling was applied during the train–test split to ensure that both the training and evaluation datasets preserved the original attrition distribution. This approach prevents biased evaluation outcomes that can arise when minority class instances are underrepresented in the test set. In addition, model configurations were adjusted to account for imbalance by applying class weighting where supported, allowing the learning algorithms to place greater emphasis on correctly identifying attrition cases during training.

The preprocessing pipeline was designed to be modular and model-agnostic, enabling consistent transformations across all classification algorithms. By encapsulating encoding, scaling, and feature selection within a unified pipeline, the risk of data leakage was minimized and reproducibility was ensured. This structured preprocessing approach supports fair model comparison and aligns with best practices for supervised machine learning workflows involving imbalanced real-world datasets. Figures 1 and 2 provide context for subsequent preprocessing decisions and model evaluation.

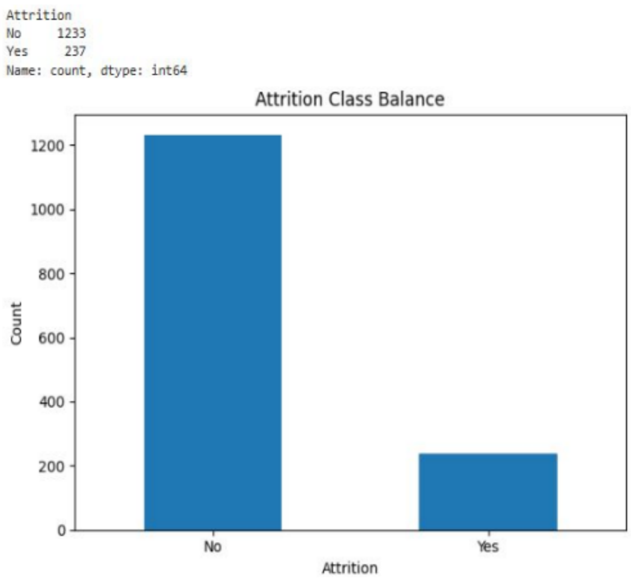

Figure 1. *Class distribution of employee attrition outcomes*

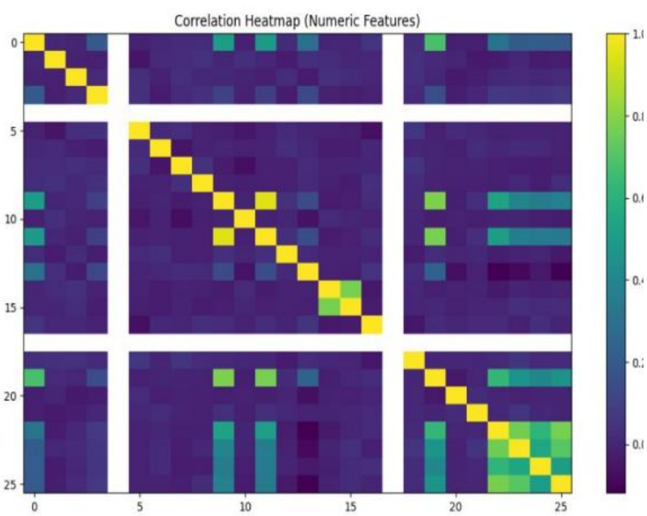

Figure 2. *Correlation heatmap of numerical HR features*

V. MODEL EVALUATION AND RESULTS

Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. XGBoost achieved the highest overall accuracy at approximately 86 percent, with a ROC-AUC close to 0.80, indicating strong discriminative capability. Random Forest demonstrated similar accuracy but showed lower recall for attrition cases, suggesting a tendency to favor majority class predictions. Logistic Regression achieved lower overall accuracy but provided the highest recall for attrition cases.

This highlights a trade-off between maximizing overall accuracy and identifying at-risk employees early. Confusion matrices and ROC curves were used to visually compare model behavior and threshold sensitivity. These results demonstrate that no single model is universally optimal and that model selection should depend on organizational priorities, such as early risk detection versus overall prediction accuracy. In addition to overall performance metrics, the results highlight meaningful differences in how each model balances false positives and false negatives.

Logistic Regression, while less accurate overall, demonstrated stronger sensitivity to attrition cases, making it more suitable in scenarios where identifying potential departures early is prioritized over minimizing false alarms. In contrast, tree-based models, particularly Random Forest, exhibited more conservative behavior by correctly classifying the majority of non-attrition cases but missing a larger proportion of employees who eventually left. This distinction underscores the importance of aligning model choice with the specific risk tolerance and intervention strategy of an organization.

The ROC-AUC scores across all three models further support the conclusion that each approach offers comparable discriminatory power despite differing classification behaviors. The relatively consistent ROC-AUC values indicate that the models are similarly effective at ranking employees by attrition risk, even though their default classification thresholds produce different precision and recall outcomes. This suggests that threshold tuning could be applied in future work to better tailor model outputs without requiring changes to the underlying modeling approach.

Overall, the evaluation results emphasize that predictive performance should not be assessed using a single metric in isolation. While accuracy provides a general measure of correctness, recall and precision offer more nuanced insight into a model's practical utility for workforce planning. The combined use of quantitative metrics and visual diagnostics, such as confusion matrices and ROC curves, enables a more comprehensive assessment of model behavior and supports informed decision-making regarding deployment in real-world HR contexts.

Figures 3–8 present confusion matrices and ROC curves for each model, illustrating differences in classification behavior and threshold sensitivity. Table II summarizes comparative performance across all evaluation metrics.

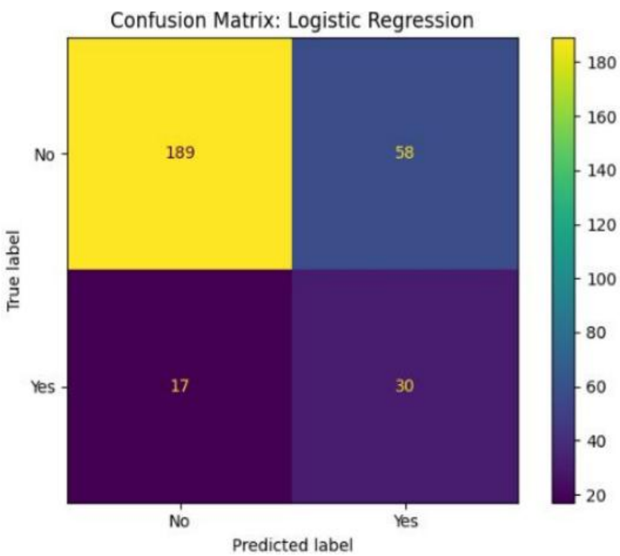

Figure 3. *Confusion matrix for Logistic Regression*

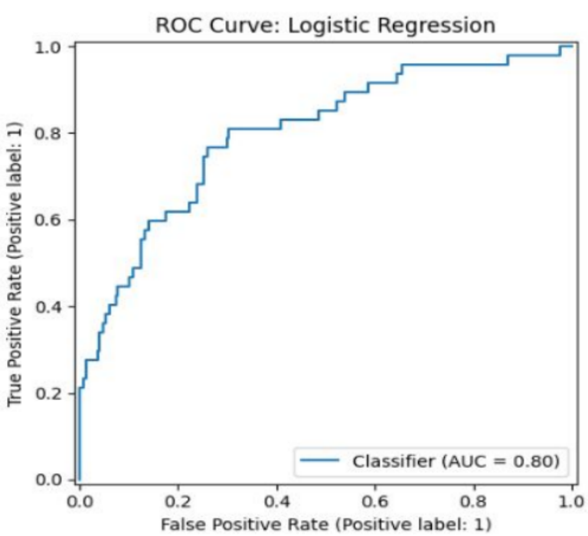

Figure 4. *ROC curve for Logistic Regression*

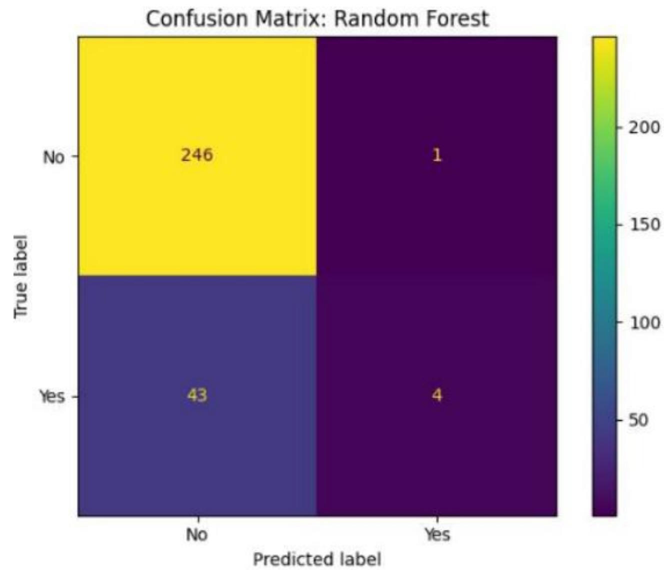

Figure 5. *Confusion matrix for Random Forest*

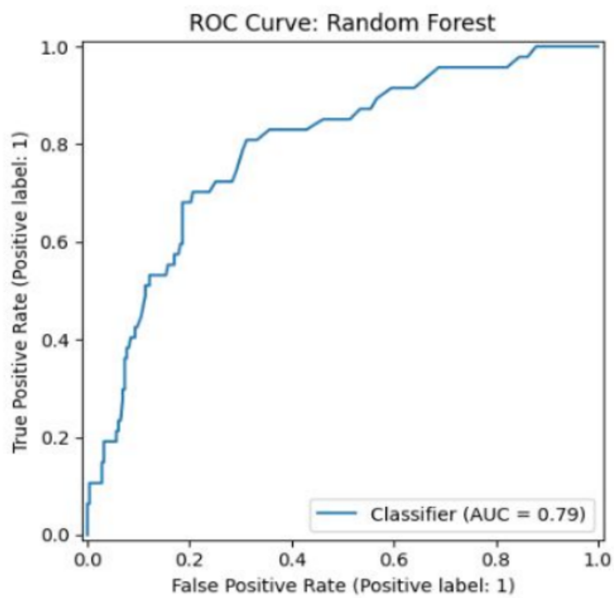

Figure 6. *ROC curve for Random Forest*

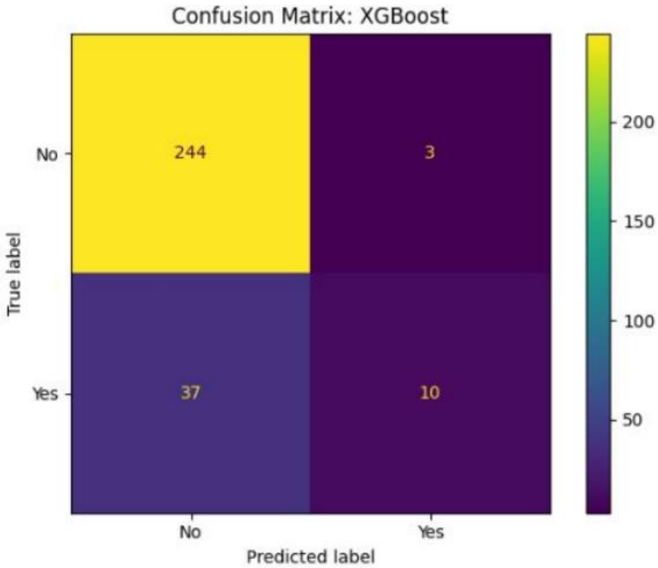

Figure 7. *Confusion matrix for XGBoost*

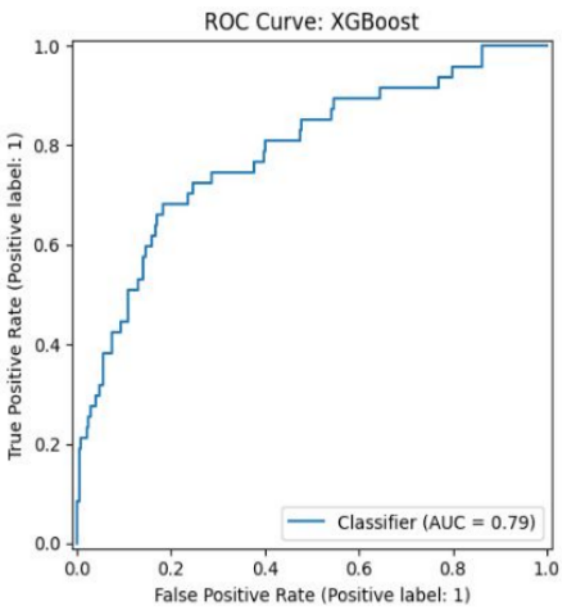

Figure 8. *ROC curve for XGBoost*

	Model	Accuracy	Precision	Recall	F1	ROC_AUC
0	Logistic Regression	0.744898	0.340909	0.638298	0.444444	0.797829
1	Random Forest	0.850340	0.800000	0.085106	0.153846	0.790335
2	XGBoost	0.863946	0.769231	0.212766	0.333333	0.785511

Table II. *Performance comparison of classification models*

VI. FEATURE IMPORTANCE AND INTERPRETAION

Feature importance analysis revealed that employee attrition is driven by a combination of workload, compensation, satisfaction, and tenure-related factors rather than a single dominant variable. Monthly income emerged as one of the strongest predictors, indicating that compensation plays a critical role in employee retention. Overtime status was also a key driver, suggesting that workload and burnout significantly contribute to attrition risk.

Job satisfaction and work-life balance scores were consistently associated with attrition, reinforcing the importance of employee engagement and supportive work environments. Tenure-related features, such as years at the company and years in the current role, indicated that employees with shorter tenure are more likely to leave, highlighting the importance of early-stage retention efforts.

Figure 9 provides a visual ranking of the most influential predictors, while Table III reports the corresponding importance scores for the top features. Together, these visualizations illustrate the relative contribution of each variable to the model's decision-making process and help quantify the strength of key attrition drivers. Presenting both graphical and tabular views enables clearer comparison across features and improves the transparency of the predictive model, supporting meaningful interpretation and actionable insights.

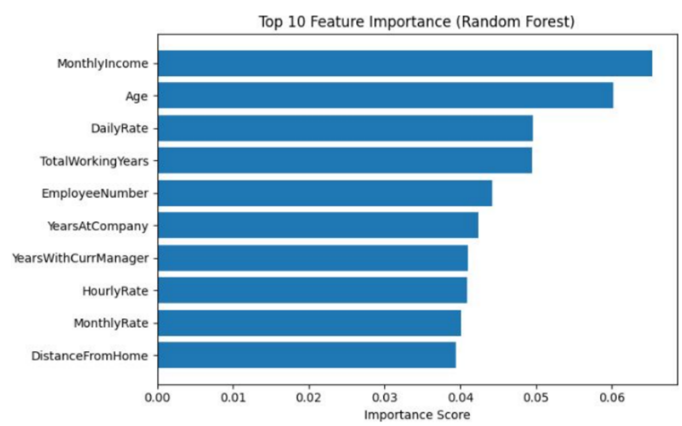

Figure. 9. *Top ten feature importance scores derived from the Random Forest model.*

Feature	Importance
MonthlyIncome	0.065387
Age	0.060171
DailyRate	0.049635
TotalWorkingYears	0.049479
EmployeeNumber	0.044250
YearsAtCompany	0.042412
YearsWithCurrManager	0.041030
HourlyRate	0.040927
MonthlyRate	0.040143
DistanceFromHome	0.039446
NumCompaniesWorked	0.035097
OverTime_No	0.031430
StockOptionLevel	0.030501
YearsInCurrentRole	0.028155
PercentSalaryHike	0.028084

Table III. *Numerical feature importance scores for the top predictors identified by the Random Forest model.*

VII. BUSINESS INSIGHTS AND RECOMMENDATIONS

The results of this study suggest that employee attrition is a multifactorial phenomenon that requires targeted, data-driven interventions. Organizations should prioritize early-tenure employees for retention programs, as attrition risk is highest during the initial years of employment. Monitoring overtime patterns can serve as a leading indicator of burnout, enabling proactive workload adjustments.

Compensation and engagement initiatives should be aligned with high-risk employee segments rather than applied uniformly across the workforce. By combining predictive modeling with interpretable insights, organizations can design more effective retention strategies that balance cost, employee well-being, and organizational performance.

VIII. CONCLUSION

This project highlights the value of applying machine learning techniques to complex human resource challenges such as employee attrition. Rather than treating attrition as an unpredictable outcome, this project demonstrates that historical HR data can be used to meaningfully assess attrition risk and support proactive workforce planning. The comparative modeling approach confirms that different algorithms provide distinct strengths, reinforcing the importance of aligning technical choices with organizational objectives.

While predictive performance varied across models, the analysis highlights that effectiveness should not be judged solely on accuracy. The ability to identify employees at elevated risk of leaving has greater strategic value than optimizing for overall correctness. The observed differences in recall across models emphasize that attrition prediction is inherently a risk management problem, where early detection enables intervention, even at the cost of some false positives. This perspective is particularly relevant for organizations seeking to reduce turnover related to burnout, disengagement, or early tenure dissatisfaction.

Beyond predictive outcomes, the interpretability of model results provides actionable insight into how workplace conditions shape employee behavior. The prominence of factors related to workload, compensation, satisfaction, and tenure suggests that attrition reflects cumulative experience rather than isolated events. For employers, this indicates that retention strategies are most effective when implemented as coordinated initiatives that address multiple dimensions of the employee experience. Data driven prioritization can help organizations focus retention resources on high impact areas instead of relying on generalized retention policies.

From an operational standpoint, the findings support the integration of predictive analytics into ongoing HR processes. Attrition risk scores can be used to inform targeted engagement efforts, guide managerial decision making, and evaluate the potential impact of policy changes over time. When applied responsibly, such models can complement human judgment rather than replace it, enabling organizations to balance workforce stability with employee well-being.

In conclusion, this project demonstrates that supervised machine learning offers both predictive and explanatory value in understanding employee attrition. By combining evaluation with interpretable insights, the approach provides a foundation for evidence-based retention strategies and highlights the broader potential of machine learning applications in organizational decision making.

References

[1] P. Subhash, "IBM HR Analytics Employee Attrition & Performance Dataset," *Kaggle*, 2017. [Online]. Available: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

[2] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available:https://scikit-learn.org/stable/

[3] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794. [Online]. Available: https://xgboost.readthedocs.io/