Report on

Predicting Employee Attrition Using Explainable Machine Learning

Submitted to: Dr. Debashish Jena

Submitted by:

Divyansh Gupta (22CS3025)
Manish Kumar (22CS3037)
Jigyasu Patel (22CS3031)
Nishant Nischal (22CS3042)

Group No: 10

Branch: CSE

Introduction & Problem Statement

Employee attrition — when people voluntarily or involuntarily leave a company — is a costly problem. A recent report estimated that replacing a single employee costs roughly **33 % of their annual salary** and can be much higher for executives【898958354219657†L78-L88】. These direct costs (recruitment, onboarding and training) are accompanied by hidden costs such as **lost productivity, low morale, damaged employer brand and a higher risk of additional turnover**【898958354219657†L193-L217】. Therefore, companies that ignore attrition risk face not only higher recruitment expenses but also reduced productivity and organisational knowledge.

Modern human-resources (HR) teams are turning to data-driven methods to identify employees at risk of leaving and to design personalised retention strategies.
Varghese et al. (2025) showed that machine-learning models like **Random Forest** and **XGBoost** can accurately forecast employee turnover and highlighted that attributes such as **job satisfaction, performance rating, monthly income and promotion frequency** are among the most influential predictors【98716518191074†L83-L96】. Explainable AI approaches, such as SHAP (SHapley Additive exPlanations), further help managers understand *why* certain employees are at risk by ranking the contribution of each feature. In a recent SHAP-based analysis of IBM's HR data, the top insights were that **working overtime**, **low job/environment satisfaction** and **lower monthly income** are strongly associated with attrition【433998785945327†L564-L583】.

This project aims to build and evaluate interpretable machine-learning models to **predict employee attrition** using a real-world HR dataset. We will develop an interactive analytics dashboard and a detailed report to help HR managers explore risk patterns, understand the key drivers of turnover and simulate "what-if" retention scenarios. By proactively identifying high-risk employees, organisations can take targeted actions — revisiting compensation plans, improving work–life balance and creating clearer career paths — to reduce turnover and retain valuable talent.

Dataset Description

We used the **IBM HR Analytics Employee Attrition & Performance** dataset. This publicly available dataset was created by IBM data scientists to illustrate common HR analytics problems and has been widely used for educational purposes. The dataset contains **1,470 employees** with **35 variables** covering demographic information, job characteristics, performance metrics and satisfaction surveys. Key fields include:

- **Age, Gender and MaritalStatus** – demographic attributes.
- **Department, JobRole and JobLevel** – organisational placement.
- **MonthlyIncome, PercentSalaryHike and StockOptionLevel** – compensation.

- **JobSatisfaction, EnvironmentSatisfaction and WorkLifeBalance** – survey-based satisfaction measures.

- **OverTime, BusinessTravel, YearsAtCompany and TotalWorkingYears** – work conditions and tenure.

- **Attrition** – the target variable indicating whether an employee left the company (Yes) or stayed (No).

Before modelling, constant columns (e.g., EmployeeCount, StandardHours, Over18 and EmployeeNumber) were removed because they carry no information. The target variable Attrition was converted to a binary flag (1 = Yes, 0 = No) for machine-learning models.

Data Cleaning & Preprocessing

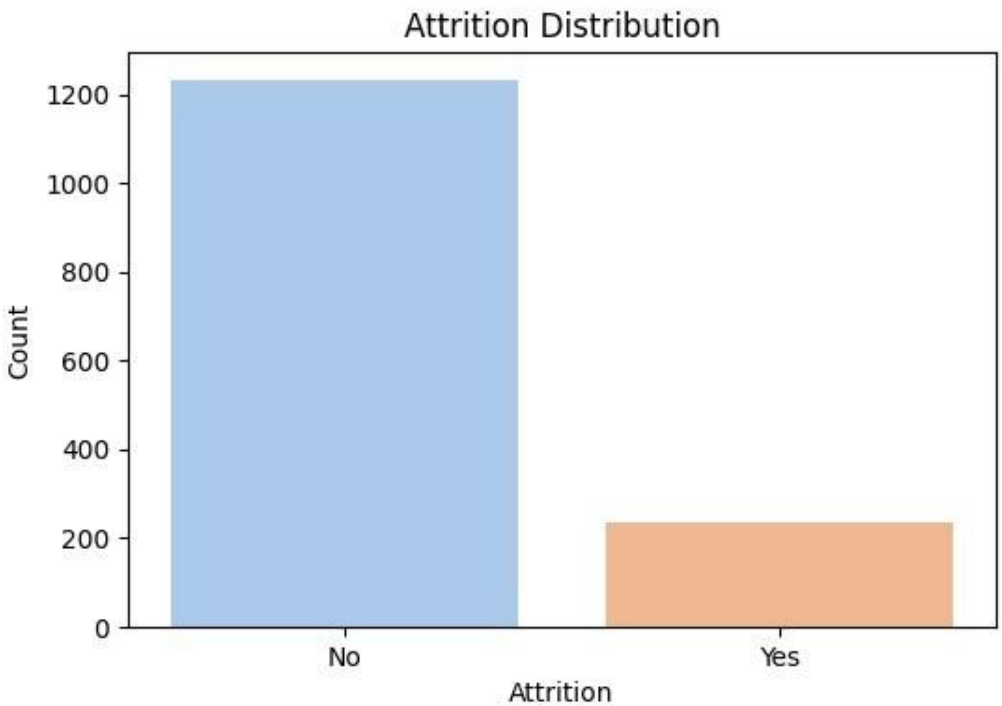
The raw dataset contained no missing values, but several columns required transformation:

1. **Constant features removed:** EmployeeCount, StandardHours, Over18 and EmployeeNumber had the same value for all employees and were dropped.
2. **Categorical encoding:** Categorical features (e.g., JobRole, Department, BusinessTravel, MaritalStatus, etc.) were encoded using **one-hot encoding** so that models could interpret them. Numerical features were scaled using **standardisation** to ensure that attributes measured on different scales (such as income and distance from home) did not unduly influence the models.
3. **Train/Test split:** The cleaned data were split into a **training set (75 %)** and a **test set (25 %)** using stratified sampling to preserve the original attrition proportion (about 16 % “Yes” and 84 % “No”).

Exploratory Data Analysis (EDA)

The exploratory analysis offers insight into attrition patterns and helps prioritise features for modelling. The charts below summarise key findings:

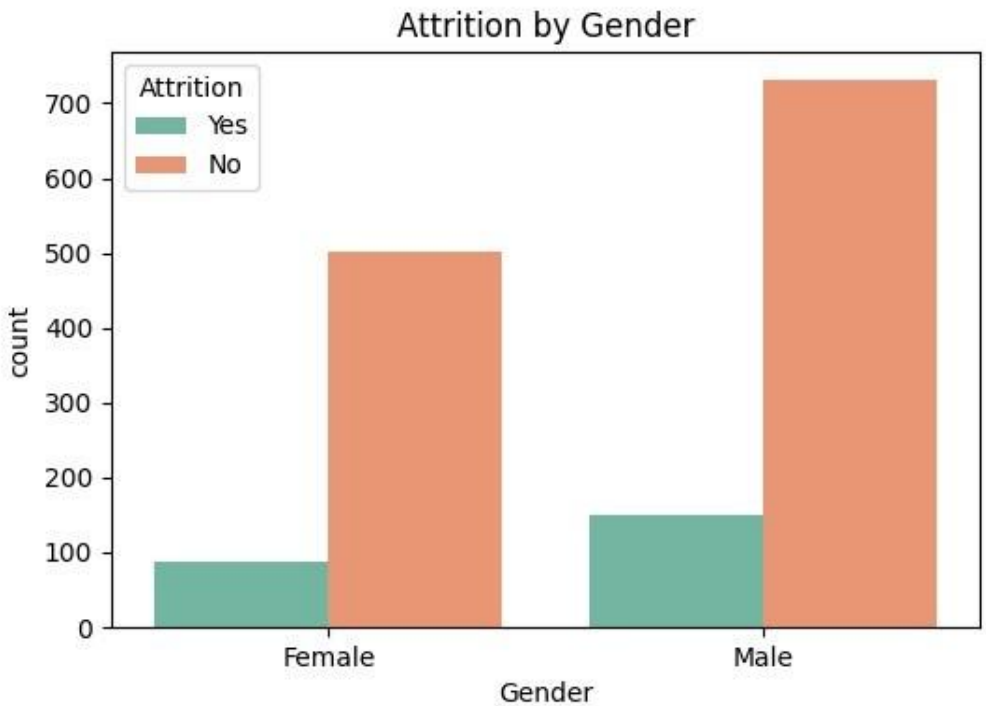
Attrition distribution



Attrition distribution

Only **237 of 1,470 employees (≈16 %)** left the company. This class imbalance has implications for modelling; we used balanced class weights to mitigate bias toward the majority class.

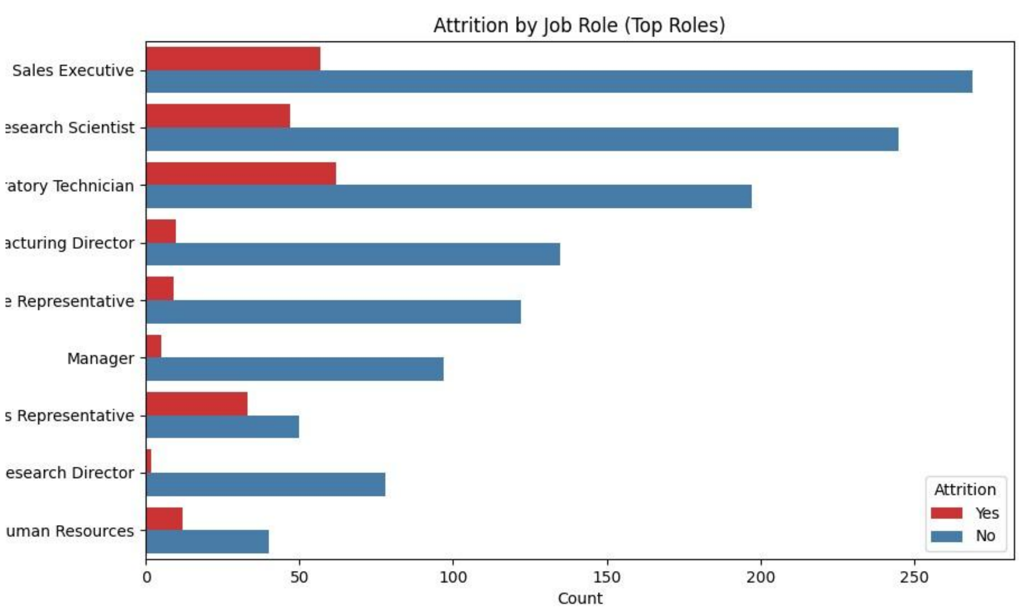
Gender and attrition



Gender vs Attrition

Both men and women show similar attrition patterns; however, the number of male employees is slightly higher. Attrition appears to be driven more by job-related factors than gender.

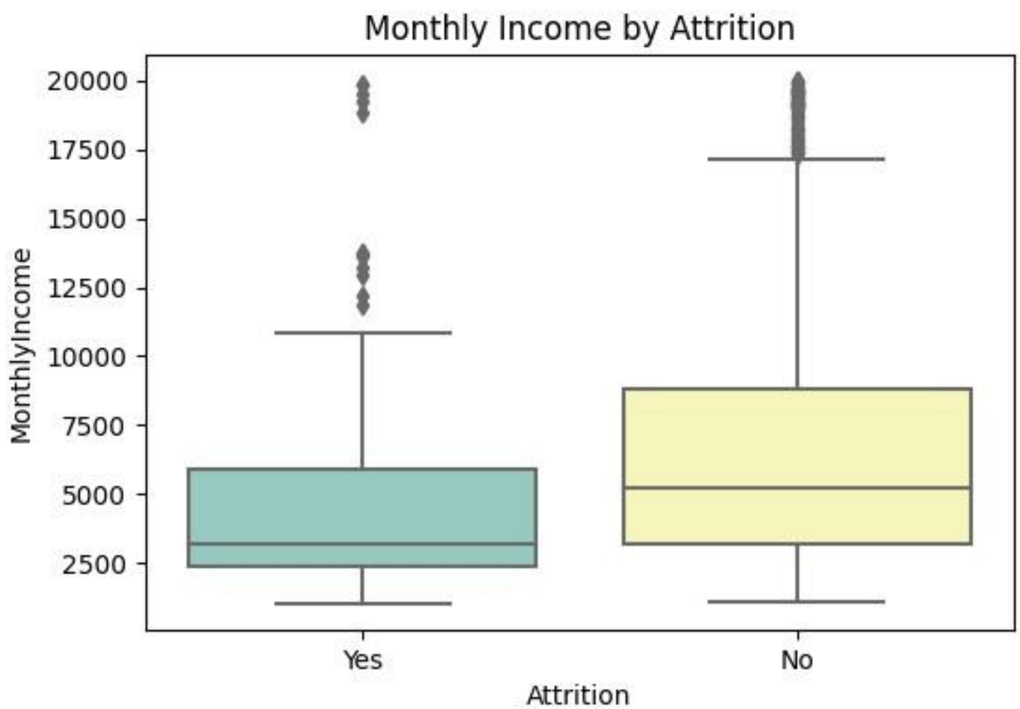
Job role and attrition



Job role vs Attrition

Attrition is particularly high among **Laboratory Technicians**, **Sales Representatives** and some **research positions**. Roles such as **Research Director** exhibit lower turnover, suggesting that senior positions with greater autonomy have higher retention.

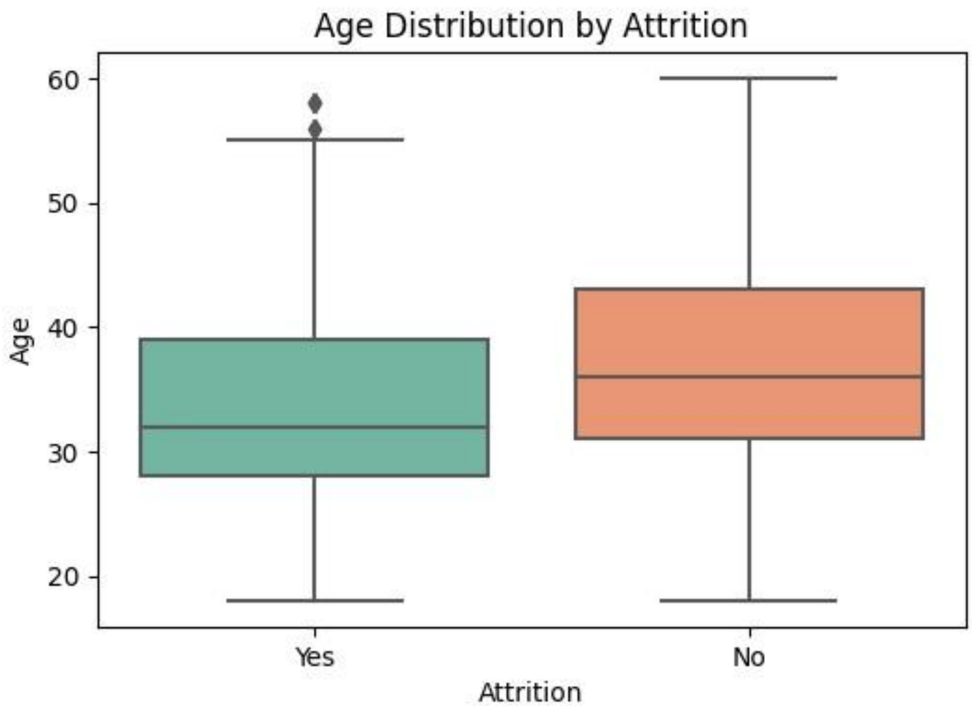
Compensation and satisfaction



Monthly income by attrition

Employees who left tended to have slightly **lower monthly incomes**. At the same time, satisfaction scores (Job Satisfaction, Environment Satisfaction and Work–Life Balance) were generally lower for employees who left (not shown), which agrees with research indicating that poor satisfaction is a key driver of attrition【433998785945327†L564-L583】.

Age and tenure



Age distribution by attrition

Years at company by attrition Employees who left were often **younger** and had **shorter tenure**. The median age of employees who stayed was around 37 years, whereas those who left were mostly in their early thirties. Similarly, employees with fewer years at the company were more likely to leave, hinting at retention issues during the early phases of employment.

Feature correlations (not pictured)

Pearson correlation analysis showed that no single numerical feature strongly predicts attrition, underscoring the need for multivariate models. Categorical variables like JobRole, BusinessTravel and OverTime exhibited stronger relationships when encoded.

Methodology / Models Used

To predict attrition, we implemented three supervised classification models using scikit-learn:

1. **Logistic Regression (with balanced class weights)** – an interpretable linear model that estimates the probability of attrition. Categorical variables were one-hot encoded, and numerical features were standardised. We extracted model coefficients to understand the direction and magnitude of each feature's effect.
2. **Random Forest Classifier** – an ensemble of decision trees that captures non-linear relationships and interactions. Class weights were balanced to compensate for the minority attrition class. Feature importance scores provide a measure of each attribute's predictive power.
3. **Gradient Boosting Classifier** – a tree-based boosting model that iteratively corrects the errors of previous trees. Although less interpretable than logistic regression, boosting often yields strong accuracy.

All models were trained within a **pipeline** that included preprocessing (encoding and scaling) to avoid data leakage. Model performance was evaluated on the held-out test set using **accuracy**, **confusion matrix**, **classification report** and **receiver-operating characteristic (ROC) area under the curve (AUC)**. Additionally, we used **5-fold cross-validation** to estimate generalisation performance for the logistic regression model.

Results & Interpretation

Model	Test accuracy	ROC AUC	Key observations
Logistic Regression	0.77	**0.81**	Highest AUC. Good trade-off between recall for attrition cases and overall accuracy.
Random Forest	0.85	0.78	Highest accuracy but slightly lower AUC; may overfit to majority class.
Gradient Boosting	**0.85**	0.78	Similar to random

Model performance

Model	Test accuracy	ROC AUC	Key observations
			forest; improved accuracy at the cost of interpretability.

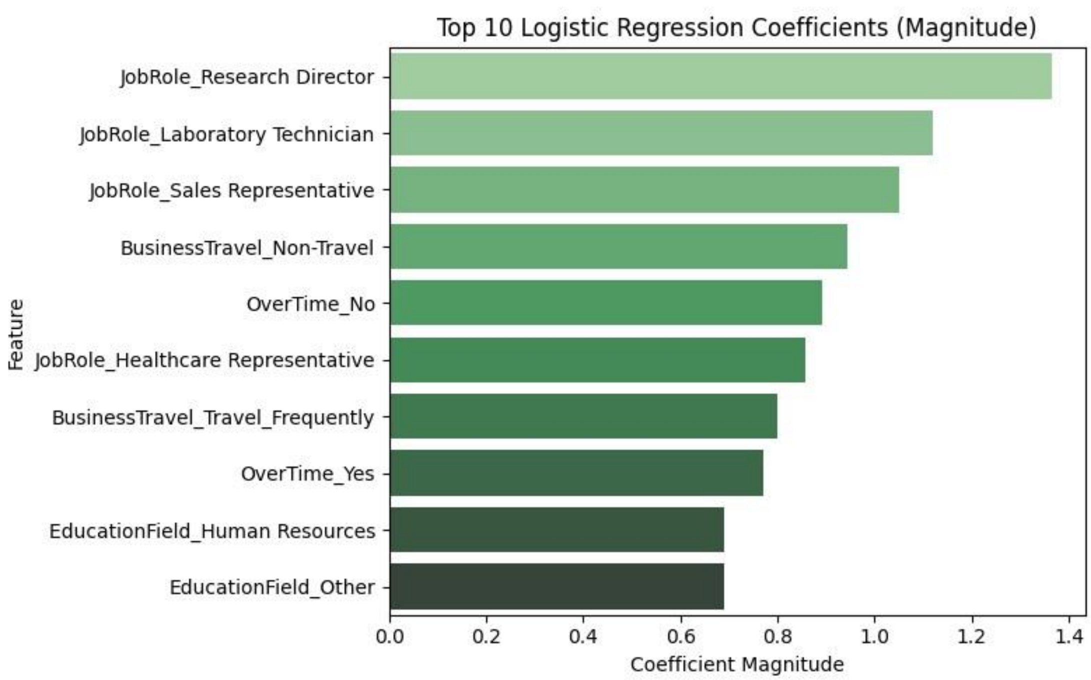
The **logistic regression model** achieved the best balance of discrimination (AUC ≈ 0.81) and interpretability. Cross-validation confirmed its robustness, with a mean AUC of **0.83** across five folds. The confusion matrix below (for the logistic model) shows that it correctly identified **63 %** of employees who left, while maintaining high precision for the "No" class:

	Predicted No	Predicted Yes
Actual No	248	61
Actual Yes	22	37

The classification report indicates precision of **92 %** for the "No" class and **38 %** for the "Yes" class. Because attrition cases are relatively rare, precision is lower for "Yes," but recall of **63 %** shows that the model captures most departures.

Feature interpretation

To make the results actionable for HR managers, we examined the coefficients of the logistic regression model. The figure below shows the **top 10 features by absolute coefficient magnitude**.



Top logistic regression coefficients

- **Job Role:** Employees working as **Research Directors** have a much **lower probability** of leaving, whereas **Laboratory Technicians** and **Sales Representatives** are more likely to quit. Senior roles often come with higher pay and autonomy, improving retention.
- **BusinessTravel:** Frequent travel increases the odds of attrition, while the absence of travel reduces it. Long or irregular travel schedules may contribute to stress and work–life imbalance.
- **OverTime:** Consistent overtime work raises attrition risk, whereas the absence of overtime decreases it. This aligns with research showing that employees working overtime are more likely to leave【433998785945327†L564-L583】.
- **Education Field:** Specialisations such as **Human Resources** and **Technical Degree** have higher attrition risk compared with **Medical** or **Life Sciences**, possibly reflecting career mobility.

These findings highlight that **job role**, **travel frequency**, **overtime** and **compensation/satisfaction factors** are critical levers. HR can use these insights to design targeted retention strategies — for example, reducing mandatory travel, limiting overtime, revisiting compensation for lower-paid roles and improving job satisfaction through feedback programmes.

Validation / Performance Evaluation

To ensure that our model generalises beyond the current data, we performed several validation steps:

- **Hold-out evaluation:** Metrics were computed on a separate test set not seen during training. The test accuracy (77 %) and AUC (0.81) suggest that the model is reasonably accurate while maintaining interpretability.
- **Cross-validation:** 5-fold cross-validation on the full dataset yielded a **mean AUC of 0.83**, indicating stable performance across different splits.
- **Class imbalance handling:** We used balanced class weights in all models so that the minority "Yes" class was not overwhelmed. Future work could explore techniques such as SMOTE (synthetic minority oversampling) or cost-sensitive learning.

Conclusion & Next Steps

Employee attrition imposes significant direct and indirect costs on organisations【898958354219657†L193-L217】. By leveraging HR data and machine-learning models, we can **predict which employees are at risk of leaving** and identify the **key drivers** of attrition. Our analysis of the IBM HR dataset showed that **job role**, **overtime**, **travel frequency**, **income** and **satisfaction levels** are crucial factors. The logistic regression model delivered the best balance of accuracy and interpretability, with an AUC of 0.81.

The insights generated here provide actionable guidance for HR managers. Interventions could include revising compensation plans, reducing excessive overtime, offering flexible travel arrangements and enhancing job satisfaction through feedback and career development opportunities. Future work could extend the model by incorporating additional data (e.g., performance reviews, engagement surveys) and deploying an interactive dashboard that allows managers to simulate “what-if” scenarios and explore individual employee profiles.