**Employee Attrition Prediction Analysis Report**

**1. Introduction:**

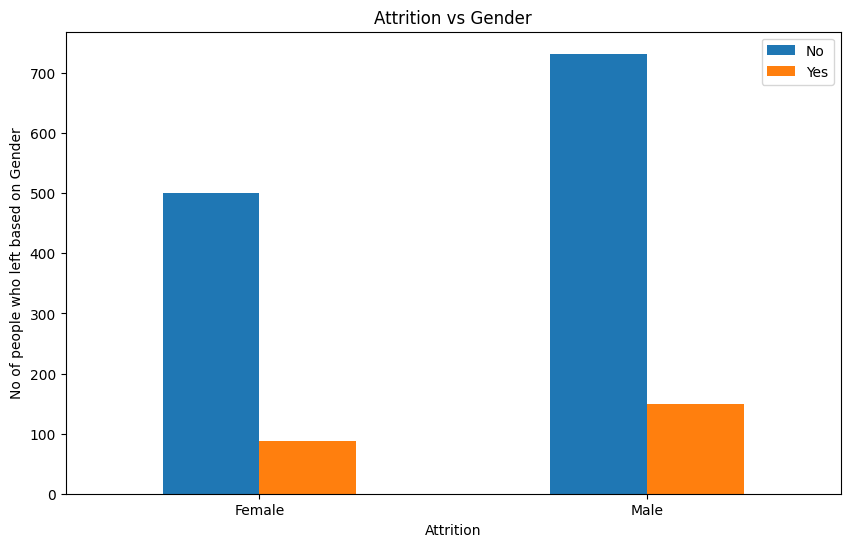
Employee attrition is a critical concern for organizations as it can lead to decreased productivity, loss of institutional knowledge, and increased recruitment costs. Predicting employee attrition using machine learning techniques can help organizations identify at-risk employees and take proactive measures to retain them. In this report, we summarize the findings and insights gained from analyzing the IBM HR Analytics Employee Attrition & Performance dataset.

**2. Dataset Analysis and Preprocessing:**

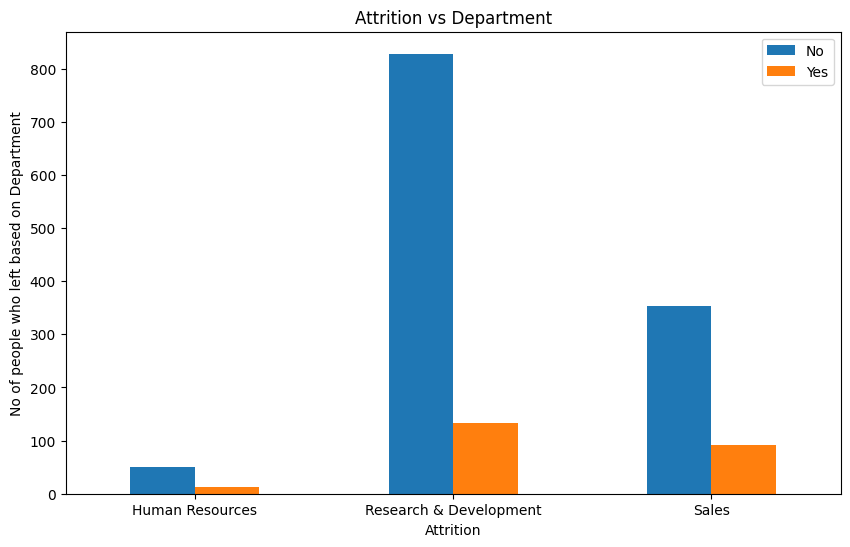
* **Dataset Description:** The dataset contains various attributes related to employee demographics, job roles, satisfaction levels, performance ratings, etc., along with a target variable indicating whether an employee has left the company (Yes or No).
* **Data Exploration:** We explored the dataset to understand its structure, features, and distribution. This involved examining descriptive statistics, checking for missing values, and visualizing relationships between variables.
* **Preprocessing Steps:** We handled missing values, encoded categorical variables using one-hot encoding, and split the dataset into training and testing sets for model development.

**2.1 Data Analysis:**

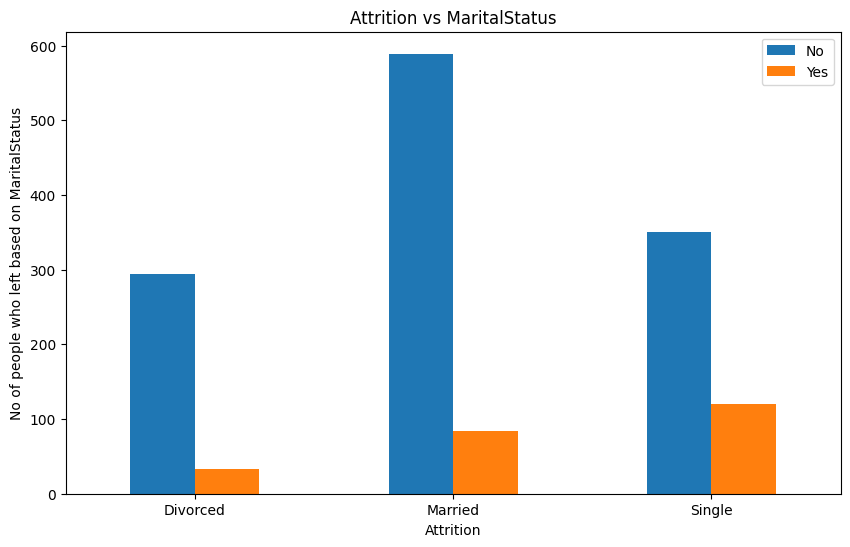
**Gender Vs Attrition:**

****

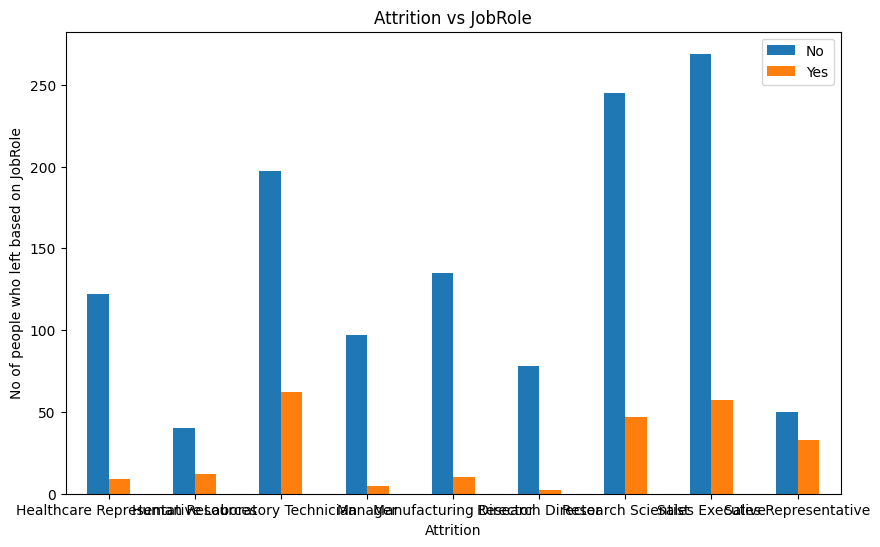
**Attrition Vs Department:**

****

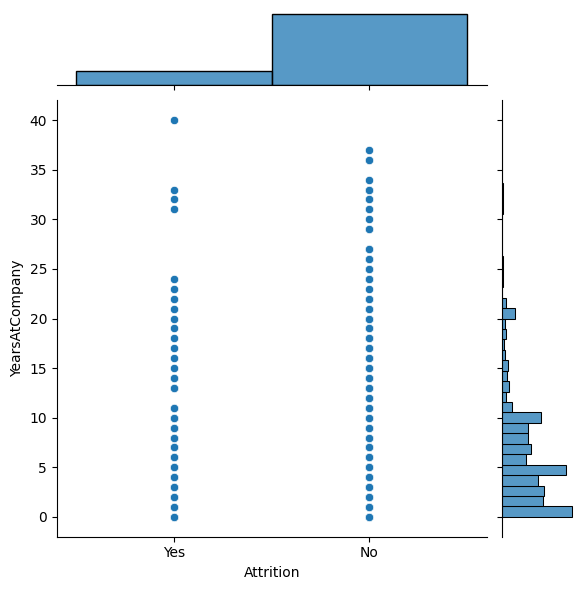
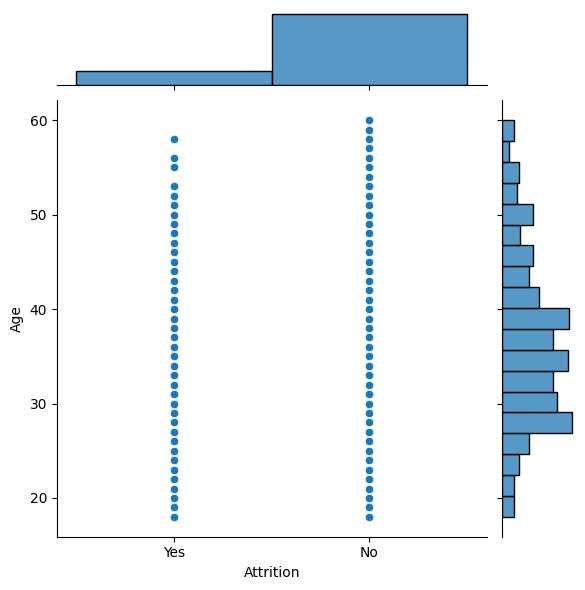
**Attrition Vs Marital Status:**

****

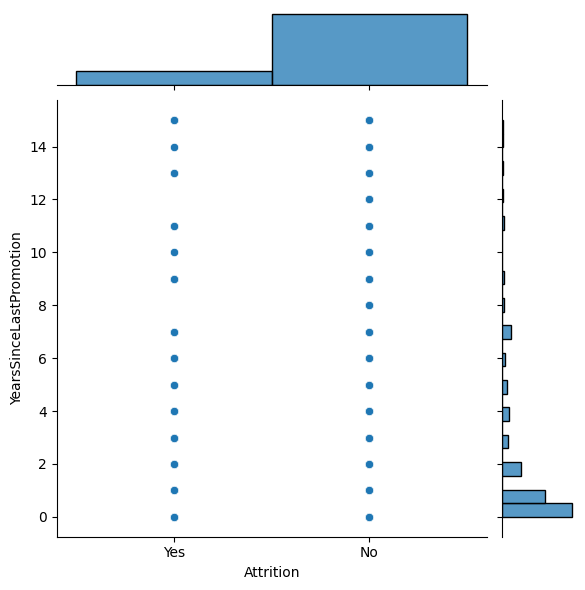
**Attrition Vs Job Role:**

****

**Attrition Vs Age: Attrition Vs Years At Company:**

****

**Attrition Vs Years Since Last Promotion:**

****

**3. Model Development:**

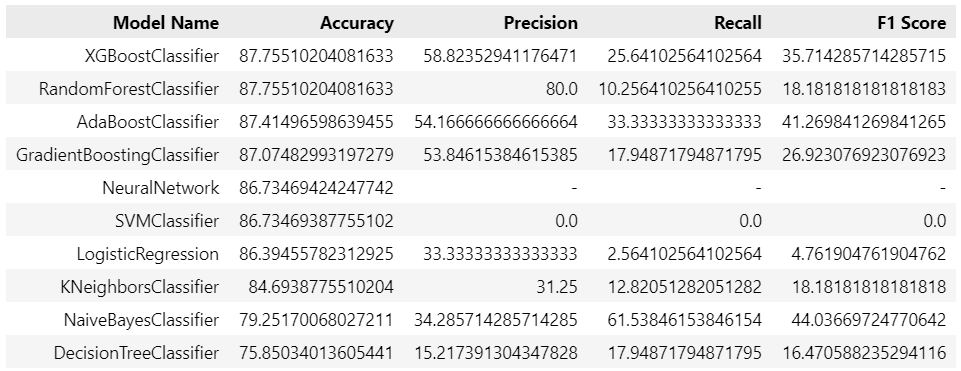
We experimented with various machine learning algorithms for binary classification, including Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost, AdaBoost, Decision Tree, K-Nearest Neighbors (KNN), Gradient Boosting, Neural Network, and Naive Bayes.

**4. Model Evaluation and Optimization:**

* We evaluated each model's performance using metrics such as accuracy, precision, recall, and F1-score on the test data.
* Optimization techniques like hyperparameter tuning and ensemble methods were applied to improve model performance.
* Insights were gained into the significance of various features and their impact on predicting employee attrition.

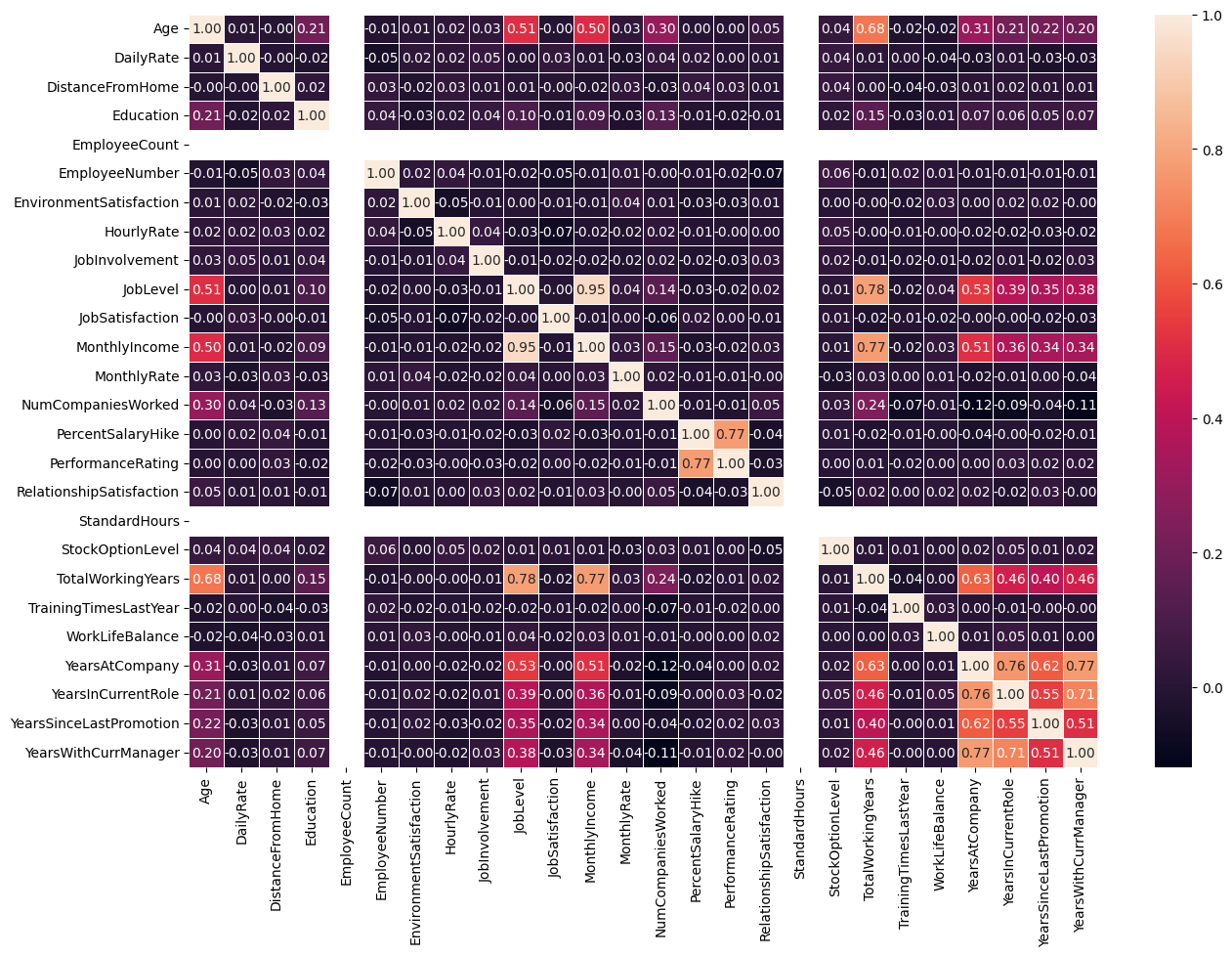
**5. Evaluation Results:**

* XGBoost, Random Forest, and AdaBoost emerged as the best-performing models based on their accuracy and F1-score.
* Neural Network also showed competitive accuracy, but its precision, recall, and F1-score were not provided for comprehensive evaluation.
* Other models like SVM, Logistic Regression, KNN, Naive Bayes, and Decision Tree exhibited varying levels of performance, with some models showing poor precision and recall.



**6. Insights Gained:**

* Factors such as job satisfaction, work-life balance, and job role were identified as significant predictors of employee attrition.
* Models like XGBoost, Random Forest, and AdaBoost provided valuable insights into feature importance and their contributions to predicting attrition.



**7. Recommendations for Reducing Employee Attrition:**

* Implement proactive measures based on predictive models to identify at-risk employees and intervene early.
* Focus on improving job satisfaction, work-life balance, and career development opportunities to increase employee retention.
* Regularly retrain and update predictive models with new data to ensure effectiveness in capturing evolving trends in attrition.

**8. Conclusion:**

Predicting employee attrition is a complex but important task for organizations to manage their workforce effectively. By leveraging machine learning techniques and analyzing relevant factors, organizations can gain valuable insights into attrition patterns and take proactive steps to reduce attrition rates, improve employee satisfaction, and enhance overall organizational performance.

**Submitted by: Ayush Verma**

**Mail: ayushuv135@gmail.com**