**Employee Attrition Analysis: Insights and Predictive Modelling**

**By-**

**Anusuya Bera**

**Preface**

In today's competitive business environment, understanding the factors leading to employee attrition is crucial for organizations to improve retention strategies. This project aims to analyse employee attrition, uncovering key trends and predicting turnover using machine learning techniques. We utilized a comprehensive dataset, explored insights through Power BI visualizations, and developed predictive models to support HR decision-making processes.

**Problem Statement**

The primary objective of this project is to analyse employee attrition and predict which employees are likely to leave the company. By identifying the key factors that influence attrition, the organization can take proactive steps to improve employee retention and engagement.

**Pre-requisites**

1. **Python**
2. **Power**
3. **Microsoft - Excel**

**Dataset**

The dataset used for this analysis is obtained from the 'IBM HR Employee Attrition' dataset, which contains information on 1470 employees, including their demographics, job satisfaction, salary, and performance details. Key columns include: Age, Gender, JobRole, MonthlyIncome, TotalWorkingYears, JobSatisfaction, Attrition, and more.

**Steps Involved**

**Step 1: Data Preprocessing**

1. Load the CSV file using pandas.
2. Handle missing values and remove unnecessary columns.
3. Encode categorical variables using OneHotEncoding.
4. Split the dataset into training and testing sets.

**Step 2: Exploratory Data Analysis (EDA)**

1. Summary statistics to understand key trends.
2. Correlation analysis to find relationships between factors.
3. Visualizations in Power BI

**Step 3: Model Building**

1. Implement machine learning models including Logistic Regression, SVM, Random Forest, etc.
2. Train and evaluate models using accuracy, confusion matrix.
3. Fine-tune models using RandomizedSearchCV to identify the best hyperparameters.

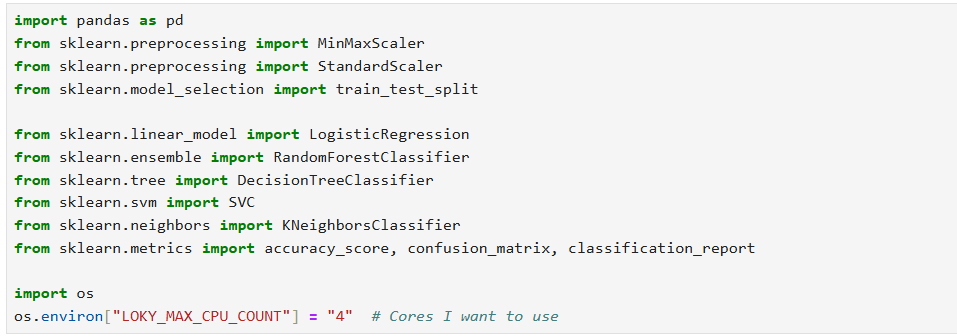
**Step 4: Final Model Selection**

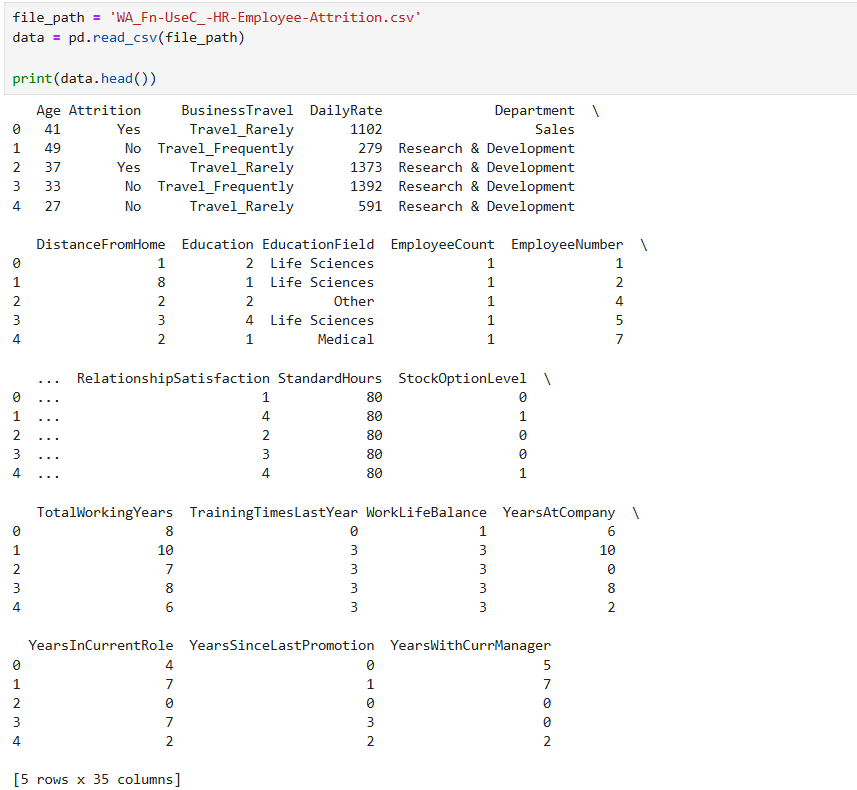
1. Compare all models based on performance metrics.
2. Select the model with the highest accuracy and best generalization.

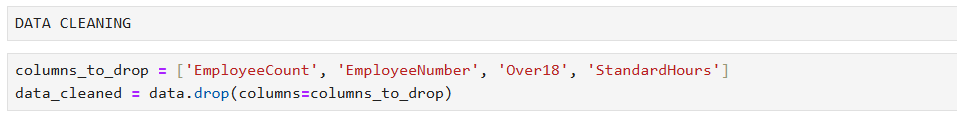
**Step 5: Power BI Visualization**

1. Import the cleaned dataset into Power BI.
2. Create individual visualizations, including pie charts, scatter plots, and stacked bar charts, etc.
3. Develop a comprehensive dashboard displaying key insights.

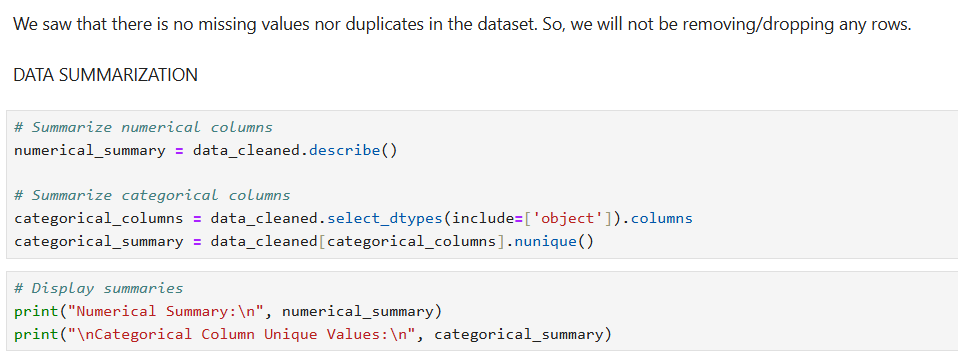
**Python Code**



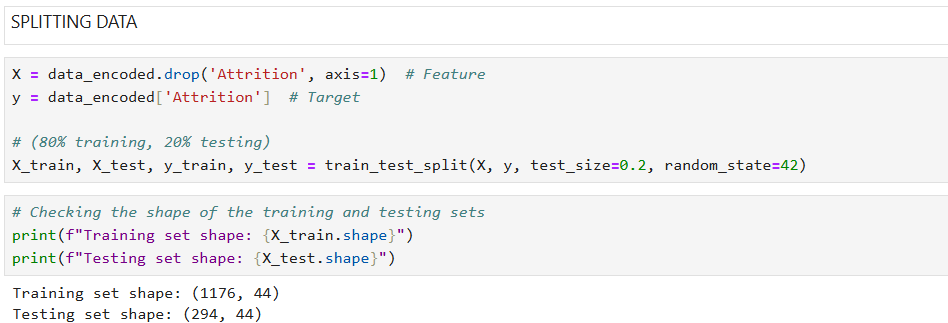


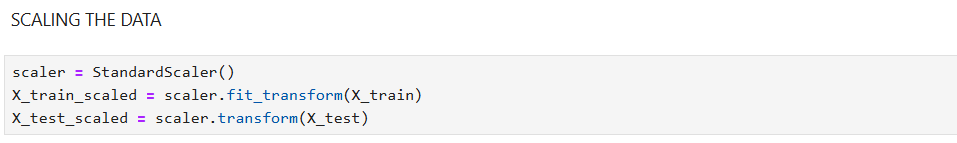


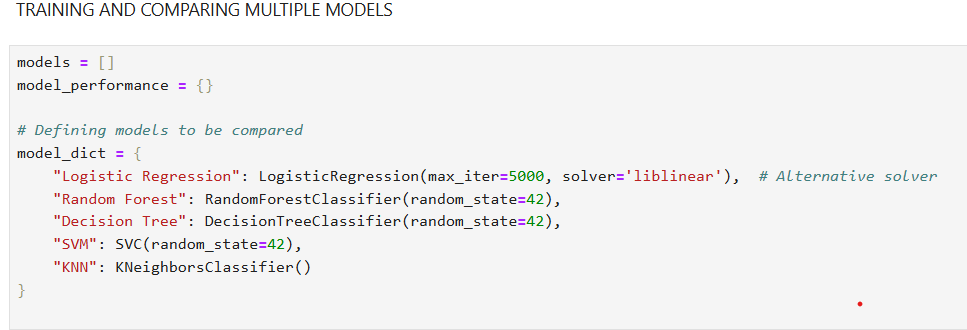




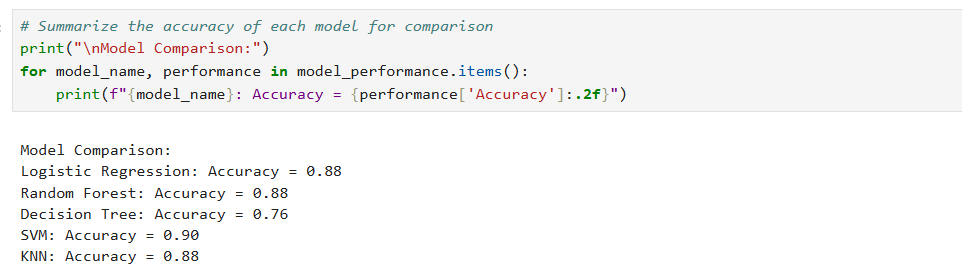


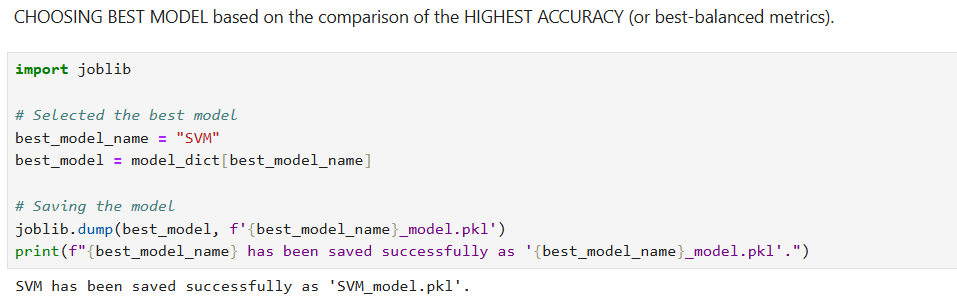




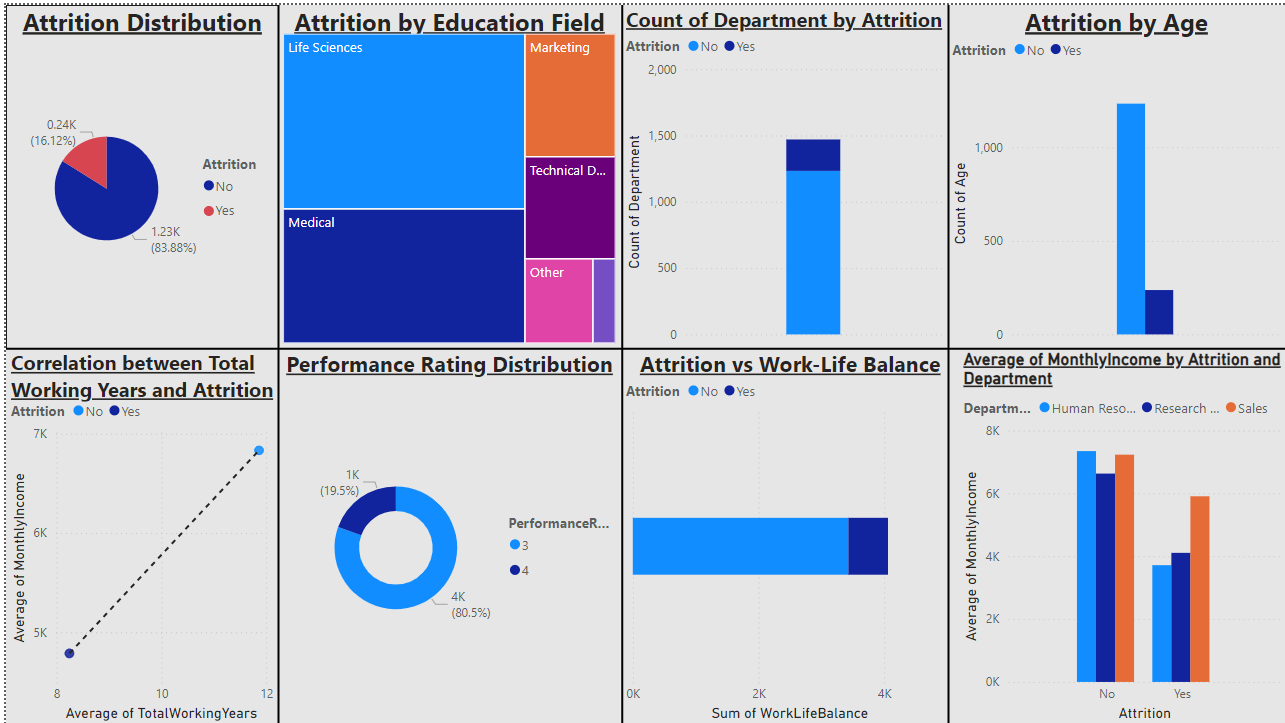








**Power BI Visualization**



**Conclusion**

Through this project, we have successfully analysed employee attrition trends and built machine learning models that predict which employees are at risk of leaving. The visualizations in Power BI provided actionable insights into key factors, allowing HR teams to make data-driven decisions to improve retention.

**References**

1. IBM HR Employee Attrition dataset: -<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset?resource=download>
2. Scikit-learn Documentation: -

<https://scikit-learn.org/>

1. Power BI Tutorials: -

<https://docs.microsoft.com/power-bi>