***IBM HR Analytics Employee Attrition & Performance***

Submitted by: ABHIJIT SINHA

Email: sinhaabhijit12@yahoo.com

UNID: UMID08062541597

**Abstract**

This project focuses on analyzing employee attrition using IBM HR Analytics dataset. The goal is to uncover factors influencing attrition, perform exploratory data analysis (EDA), integrate SQL queries, and build machine learning models for prediction. A Streamlit dashboard was also developed for interactive exploration.

**Dataset Description**

**Source:** IBM HR Analytics dataset (fictional dataset created by IBM data scientists).

**Rows:** 1470 employees

**Columns:** 35 features (demographics, job satisfaction, income, work-life balance,

etc.)

**Target Variable:** Attrition (Yes/No)**A table with numbers and letters

AI-generated content may be incorrect.**

**Technology Stack**

- Python (pandas, NumPy, matplotlib, seaborn, scikit-learn)

- SQL (MySQL) – to query attrition by department, gender, etc.

- Machine Learning Models: Logistic Regression, Decision Tree, Random Forest,

Gradient Boosting

- Stream lit – Interactive dashboard for visualization & prediction

- Excel Export – Business-friendly reports

# Project Workflow and Implementation Steps

# 1.Data Preprocessing

* Dropped irrelevant columns (`Employee Count`, `Employee Number`,

`Standard Hours`, `Over18`)

* Checked for missing values (none found)
* Encoded categorical variables using Label Encoding
* Scaled numerical features with `Standard Scaler`

**2. Exploratory Data Analysis (EDA)**

* Distribution of Attrition (16% attrition rate)
* Correlation Heatmap
* Attrition by Department, Gender, Job Role, etc.

**3. SQL Queries**

* Overall Attrition Rate
* Attrition by Department
* Attrition by Gender

**4. Machine Learning Models**

* Logistic Regression
* Decision Tree
* Random Forest (Best performance)
* Gradient Boosting
* Random Forest provided the most reliable results.
* Feature importance showed Over Time, Monthly Income, Job Satisfaction, and Work Life Balance as key drivers.

**5. Stream lit Dashboard**

* Dataset Overview
* Attrition Distribution & Correlation Heatmap
* Attrition by Department (interactive)
* Model Accuracy Comparison
* Predict Attrition for a New Employee (user input form)

# Analysis & Results

**1. The Preprocessing Flow Diagram.**

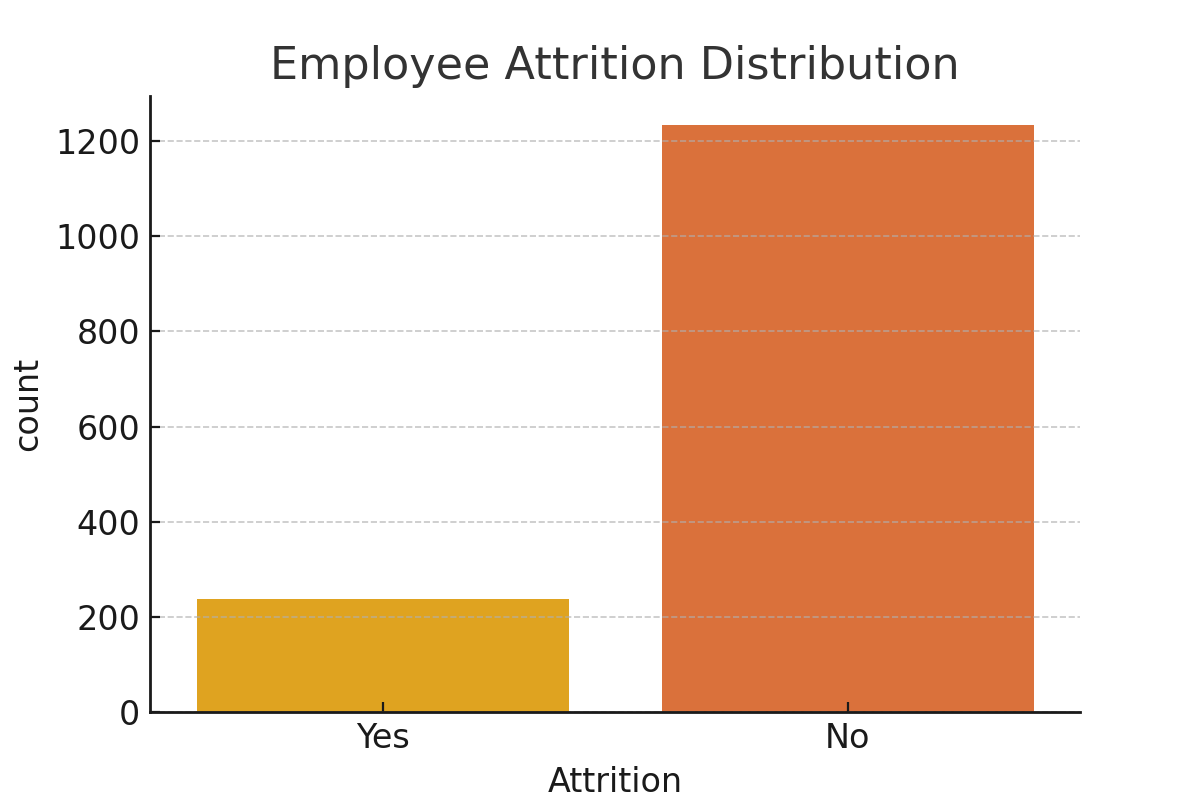
The preprocessing pipeline consisted of the following steps: removing irrelevant columns, checking missing values, encoding categorical variables, and scaling numerical features.

A diagram of steps to a step

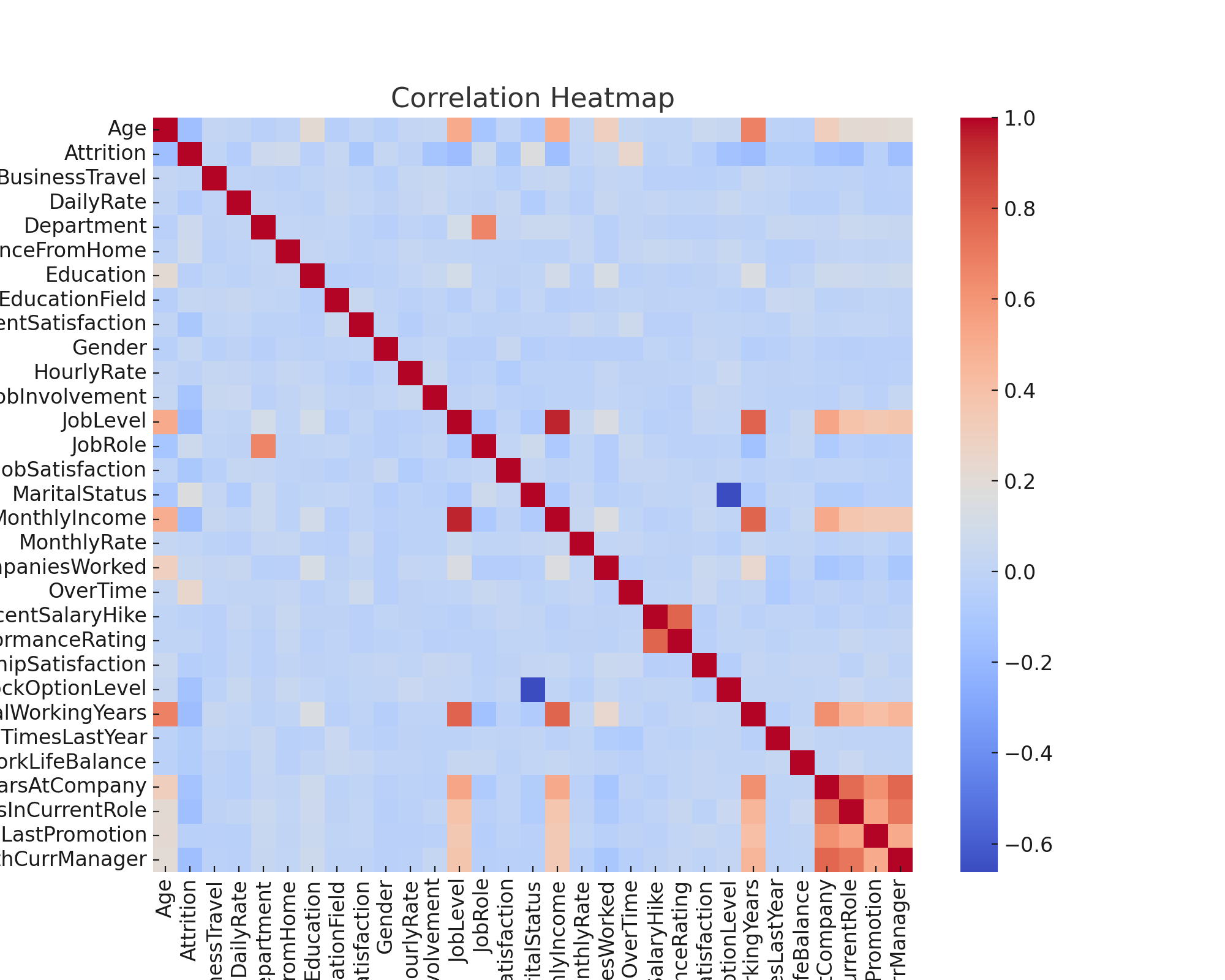
AI-generated content may be incorrect.

# 2. Exploratory Data Analysis charts

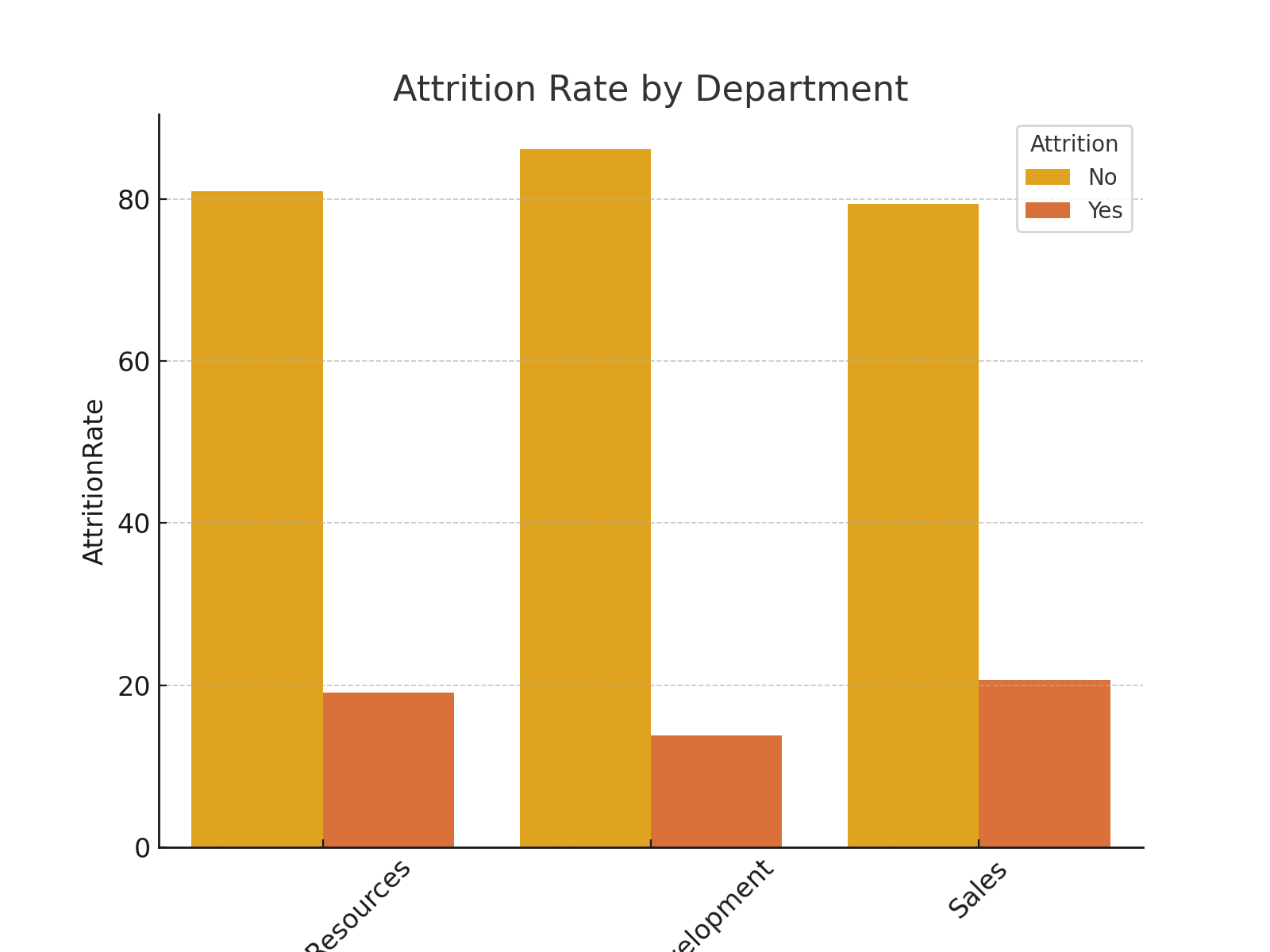
**Attrition Distribution** — About 16% of employees have left the company, indicating a moderate attrition rate.



**Correlation Heatmap** — Reveals relationships between features, highlighting key dependencies like Job Satisfaction, Monthly Income, and Over Time.



**Attrition by Department** — Highlights that attrition is highest in Sales, followed by Human Resources.



# 3. SQL Query Results

A graph with a bar and a number of text

AI-generated content may be incorrect.

A graph of a graph showing different colored squares

AI-generated content may be incorrect.A graph with orange and white bars

AI-generated content may be incorrect.**Table outputs (as returned by SQL queries):**A close-up of a graph

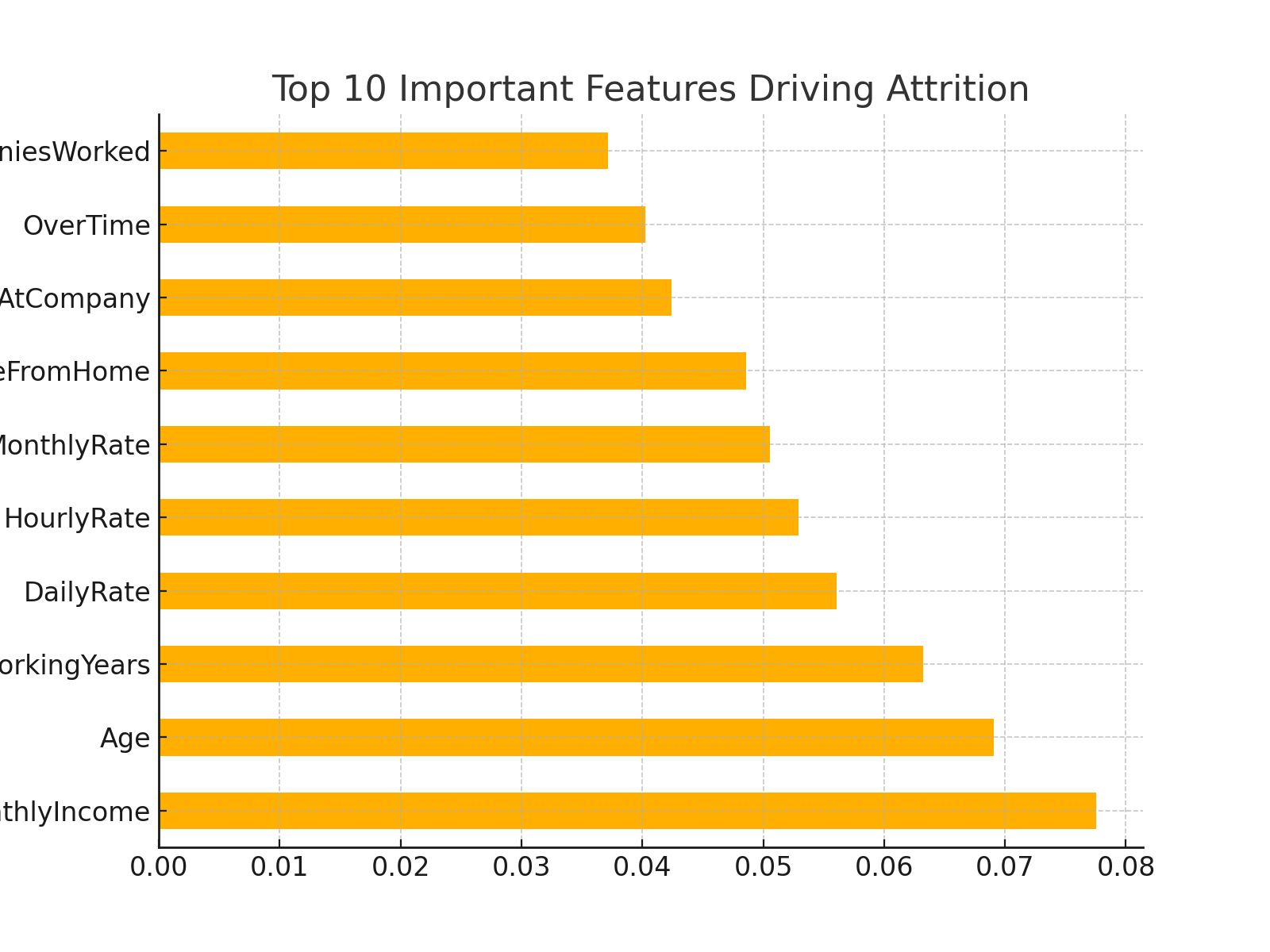
AI-generated content may be incorrect.

A close-up of a number

AI-generated content may be incorrect.

# 4. Machine Learning Models Visualizations

Multiple models were trained including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Random Forest achieved the highest accuracy. Feature importance analysis is shown below.



# 5. Stream lit App Screenshots

A screenshot of a data analysis

AI-generated content may be incorrect.

# Results and Insights

* **Attrition Rate:** ~16% of employees left the company.
* **Key Drivers of Attrition:** Overtime, low Job Satisfaction, low Monthly Income, poor Work-Life Balance.
* **Model Accuracy:** Random Forest & Gradient Boosting performed the best.
* **Stream lit Dashboard:** Provided HR teams with interactive tools for analysis and prediction.

# Conclusion

The analysis demonstrated that attrition is influenced by multiple factors including OverTime, JobSatisfaction, MonthlyIncome, and WorkLifeBalance.

Machine learning models helped in predicting attrition with high accuracy.

Stream lit app provides an interactive way to explore insights.

# References

**Dataset:** C:/Users/Hp/OneDrive/Desktop/python/IBMHRANALYTICS/WA\_Fn-UseC\_-HR-Employee-Attrition.csv  
**GitHub Project:** https://github.com/abhi-1009/IBM-HR-Analytics-Employee-Attrition-Performance