Project Report: HR Analytics & Employee Attrition Prediction

Name: SHAWN THOMAS **Date:** November 24, 2025

1. Abstract

Employee attrition poses a significant financial and operational challenge to organizations. This project aims to analyze historical employee data to identify key factors contributing to resignation and to build a predictive model that identifies employees at risk of leaving. Using a Logistic Regression model and an interactive Power BI dashboard, this project provides actionable insights to help the HR department improve retention strategies.

2. Introduction

"Attrition" refers to the phenomenon of employees leaving a company. High attrition rates lead to increased hiring costs, loss of institutional knowledge, and reduced team morale. The objective of this project is to transition from reactive HR management to proactive retention. By leveraging the "IBM HR Analytics" dataset, we answer two critical questions:

1. **Prediction:** Can we accurately predict which employees are likely to quit?
2. **Insight:** What are the primary drivers (e.g., salary, age, job satisfaction) behind these decisions?

3. Tools & Technologies Used

- **Language:** Python 3.9
- **IDE:** Jupyter Notebook (Anaconda)
- **Libraries:**
  - **Pandas:** For data manipulation and cleaning.
  - **Seaborn/Matplotlib:** For Exploratory Data Analysis (EDA).
  - **Scikit-Learn (sklearn):** For building the Logistic Regression model, scaling data (StandardScaler), and evaluating performance (confusion_matrix).
- **Visualization:** Microsoft Power BI (for the executive dashboard).

4. Steps Involved

The project was executed in five distinct phases:

1. **Data Loading & Cleaning:** The dataset was loaded into Python. Missing values were checked, and the target variable 'Attrition' was converted from binary text (Yes/No) to numerical format (1/0).
2. **Exploratory Data Analysis (EDA):** Visualizations were created to understand data distribution. Key findings showed higher attrition among younger employees and those with lower monthly incomes.

3. **Data Preprocessing:** Categorical variables (e.g., Department, JobRole) were converted to numbers using LabelEncoder. Numerical features (e.g., Age, Income) were scaled using StandardScaler to ensure the Logistic Regression model performed optimally.
4. **Model Building:** The data was split into training (80%) and testing (20%) sets. A Logistic Regression model was trained to classify employees into "Risk of Leaving" vs. "Likely to Stay."
5. **Dashboard Creation:** A Power BI dashboard was built to visualize the key performance indicators (KPIs) and allow dynamic filtering by department and demographics.

5. Model Performance & Results

The Logistic Regression model was evaluated on the test dataset.

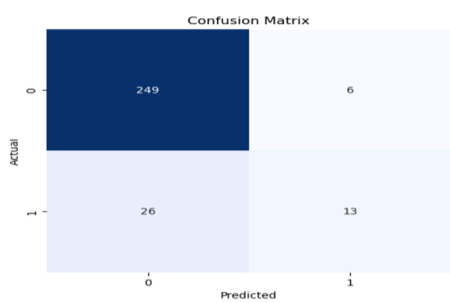
- **Model Accuracy:**

```
--- Model Accuracy Report ---
              precision    recall  f1-score   support

           0       0.91      0.98      0.94       255
           1       0.68      0.33      0.45        39

    accuracy                           0.89       294
   macro avg       0.79      0.65      0.69       294
weighted avg       0.88      0.89      0.87       294
```

- **Confusion Matrix:**



6. Strategic Recommendations (Conclusion)

Based on the data analysis, the following strategies are recommended to reduce attrition:

1. **Address Income Disparities:** Data indicates a correlation between lower income and attrition. A market salary benchmark review is recommended for high-risk roles like Sales Representatives.
2. **Manage Overtime:** Employees working frequent overtime show higher resignation rates. Implementing strict overtime caps or hiring support staff could alleviate burnout.
3. **Career Progression:** Employees stagnating in the same role for over 3 years are at high risk. Implementing a mandatory "Career Roadmap" discussion for employees at the 2-year mark is advised.
4. **Targeted Mentorship:** Since younger employees (18-25) have the highest turnover, a mentorship program pairing them with senior leaders could improve engagement and loyalty.

7. Dashboard Snapshot

Below is the final interactive dashboard created for HR Managers: