Employee Performance and Retention Analysis Report

This report analyses employee attrition patterns using a dataset containing demographic, performance, and employment details. The goal is to identify key drivers of attrition and develop predictive models to assist HR in retention planning. Using statistical analysis, probability concepts, and machine learning models, we extracted actionable insights and recommendations.

---

1. Dataset Overview

The dataset contains the following attributes for each employee:

- **Employee ID** – Unique identifier for each employee.
- **Name** – Employee's name.
- **Age** – Age of the employee.
- **Department** – Department where the employee works.
- **Salary** – Current salary of the employee.
- **Years at Company** – Tenure in the organization.
- **Performance Score** – Employee's performance rating.
- **Attrition (Yes/No)** – Whether the employee left the company.

---

2. Data Cleaning & Preprocessing

- **Missing Values** – No major missing values were found after preprocessing.
- **Duplicates** – Duplicate entries were removed to maintain accuracy.
- **Inconsistent Data** – Standardized department names and categorical labels.
- **Label Encoding** – Converted categorical fields such as Performance Score into numerical form for modelling.

---

3. Exploratory Data Analysis (EDA)

Key findings from the descriptive statistics and visual exploration:

- **Attrition Rate** – 39% of employees have left the company, while 61% remain.
- **Performance and Attrition** – Lower performance scores are strongly linked to higher attrition probability.
- **Tenure and Attrition** – Employees with shorter tenure are more likely to leave.
- **Department Trends** – Some departments have significantly higher attrition rates, indicating possible cultural or workload issues.
- **Salary Insights** – Lower salaries correlate with higher attrition, but salary alone is not the strongest predictor.
- **Age Group Trends** – Younger employees are more likely to leave, possibly due to career changes or better opportunities.

---

4. Probability & Statistical Analysis

- **Probability of Leaving** – Estimated probability of an employee leaving based on department and performance scores showed that underperformers have a much higher risk.
- **Bayes' Theorem Application** – Calculated conditional probabilities confirmed that performance score significantly impacts attrition likelihood.
- **Hypothesis Testing** – Statistical tests confirmed that mean performance scores differ significantly across departments, suggesting varying evaluation standards or work conditions.

---

5. Machine Learning Models

Several predictive models were built and evaluated:

- **Random Forest Classifier**
    - Achieved strong accuracy after preprocessing (~95%).
    - Identified performance score, years at company, and department as top predictors.

- **Deep Learning Model (Keras)**
    - Achieved comparable performance with the Random Forest.
    - Well-suited for capturing non-linear relationships.
- **Comparison**
    - Random Forest provided high interpretability.
    - Deep Learning offered flexibility for scaling with larger datasets.

---

6. Key Insights

1. **Performance Score is the strongest predictor** of attrition, followed by years at the company and department.
2. Employees with low performance ratings are **significantly more likely to leave**.
3. Certain departments face higher attrition risk, suggesting **localized retention challenges**.
4. Younger employees and those with shorter tenure **leave at higher rates**.
5. Salary impacts attrition, but **performance and tenure are stronger drivers**.