Project Report

on

HR Analytics - Predict Employee Attrition

1. Introduction

Employee attrition is a critical challenge faced by organizations across industries, directly affecting workforce stability, team morale, and overall productivity. High turnover not only leads to increased recruitment and training costs but also disrupts project continuity and erodes institutional knowledge. Recognizing this, the primary objective of this project is to systematically analyze employee demographic and performance data to uncover the underlying factors that contribute to voluntary resignations. By leveraging data analytics and machine learning, a predictive model has been developed to identify employees who are at higher risk of leaving, enabling HR departments to take timely and informed interventions.

2. Abstract

In this project, an in-depth Exploratory Data Analysis (EDA) was performed on a comprehensive HR dataset containing various attributes such as employee demographics, departmental distribution, salary ranges, promotion history, job roles, and work-life balance indicators. This analysis helped uncover meaningful patterns and trends related to employee attrition, such as which departments and roles experience higher resignation rates, how factors like age and salary bands influence turnover, and the impact of promotions and work conditions on retention.

To further strengthen the insights, a Random Forest Classifier—a robust machine learning model—was designed and trained to predict the probability of an employee leaving the organization. The model achieved an impressive test accuracy of **87%**, demonstrating strong overall performance, with perfect precision when predicting resignations. However, the model's lower recall rate highlighted the importance of complementing predictions with continuous human oversight and additional qualitative factors.

3. Tools Used

- **Python:** Pandas, Seaborn, Scikit-learn (Random Forest, SHAP)
- **Power BI:** Interactive dashboard for visual analytics
- **Excel:** Data preprocessing and manual inspection
- **Jupyter Notebook:** Code development and testing

4. Steps Involved in Building the Project

1. **Data Collection & Cleaning:**
   The first step involved gathering a reliable HR dataset containing detailed information about employees, including demographics, departmental assignments, job roles, income levels, years at the company, and resignation status. The raw data was then cleaned by removing duplicates, handling missing values, and ensuring consistency in data types to maintain data integrity for further analysis.

2. **Exploratory Data Analysis (EDA):**
   A comprehensive EDA was conducted to understand the distribution and relationship of various factors with employee attrition.
3. **Feature Engineering:**
   Relevant new features were created to strengthen the predictive power of the model. Categorical variables like departments, education fields, and job roles were encoded into numerical values using label encoding. Additional features, such as promotion gap years and ratios related to work-life balance, were engineered to provide deeper insights into potential attrition drivers.

4. **Model Building:**
   A robust Random Forest Classifier was developed as the machine learning model of choice due to its effectiveness in handling classification problems with complex feature interactions. The cleaned and feature-engineered dataset was split into training and testing sets to validate performance.

5. **Model Evaluation:**
   The performance of the trained model was evaluated using key metrics, including accuracy, precision, recall, F1-score, and a confusion matrix. leaving.

6. **Visualization & Dashboard:**
   An interactive Power BI dashboard was designed and developed to present the findings visually. The dashboard allows HR managers to monitor attrition trends in real time, filter insights by department or job role, and identify high-risk areas quickly.

7. **Recommendations Report:**
   Finally, an evidence-based recommendations report was prepared. This document summarizes key findings and outlines practical strategies for HR teams to proactively reduce attrition.

Conclusion

The HR Attrition Analysis project provides powerful insights into the root causes of employee turnover and demonstrates how machine learning and data visualization can support HR decisions. By implementing the recommended strategies — such as career growth opportunities, competitive compensation, work-life balance improvements, and an early-warning system — organizations can significantly reduce attrition rates and build a more engaged workforce.