HR Analytics: Employee Attrition Prediction

Project Completion Report

Introduction

Employee attrition represents a significant challenge for organizations, impacting productivity, morale, and bottom-line results. This project leveraged data analytics and machine learning to identify key factors driving employee turnover and build predictive models to proactively identify at-risk employees. By analysing a comprehensive HR dataset containing 1,480 employee records with 38 attributes, we developed data-driven insights and recommendations to help reduce unwanted attrition, which currently stands at 16.1% across the organization.

Abstract

This project applied data science techniques to predict employee attrition using HR analytics data. We conducted extensive exploratory data analysis to uncover patterns in employee turnover, developed machine learning models (Logistic Regression and Random Forest) achieving 85.8% accuracy, and created interactive visualizations to communicate insights effectively. Our analysis revealed that monthly income, overtime requirements, age demographics, and distance from home were the strongest predictors of attrition. The project delivered three key outputs: an interactive Power BI dashboard displaying attrition trends, a detailed model accuracy report with performance metrics, and evidence-based recommendations for reducing employee turnover.

Tools Used

- **Python 3.x**
    - Pandas for data manipulation and analysis
    - NumPy for numerical operations
    - Scikit-learn for machine learning model development
    - Matplotlib and Seaborn for data visualization
    - Jupyter Notebook as development environment
- **Power BI**
    - Dashboard creation and interactive visualizations
    - Metric calculation and KPI tracking
    - Department and demographic filtering capabilities
- **Supporting Tools**
    - Git/GitHub for version control
    - Microsoft Office for documentation

Steps Involved in Building the Project

1. **Data Preparation and Exploration**
    - Loaded and cleaned the HR dataset containing 1,480 employee records

- Handled missing values, particularly in YearsWithCurrManager column
- Generated descriptive statistics and visualization of key variables
- Analyzed attrition patterns across departments, age groups, and job roles

2. **Feature Engineering and Selection**
    - Applied label encoding to transform categorical variables
    - Identified significant features through correlation analysis
    - Selected relevant variables based on business context and statistical significance
    - Prepared training and testing datasets with 80/20 split
3. **Model Development**
    - Implemented Logistic Regression model (84.8% accuracy)
    - Developed Random Forest Classifier (85.8% accuracy)
    - Evaluated models using precision, recall, F1-score and confusion matrices
    - Identified feature importance ranking for deeper understanding of attrition drivers
4. **Visualization and Dashboard Creation**
    - Designed interactive Power BI dashboard with key metrics
    - Created visualizations for attrition by age, job satisfaction, years at company
    - Implemented filtering capabilities by department and demographics
    - Developed executive summary metrics for at-a-glance understanding
5. **Analysis and Recommendation Development**
    - Interpreted model results to identify key attrition factors
    - Developed targeted recommendations for attrition prevention
    - Created comprehensive documentation of project methodology and findings
    - Prepared final deliverables including dashboard, model report, and prevention strategies

Conclusion

The HR Analytics Attrition Prediction project identified key turnover drivers, including compensation, work-life balance, age, and commute distance. The Random Forest model achieved 85.8% accuracy, highlighting high-risk groups like younger employees and those in specific fields. The interactive dashboard provides HR with actionable insights, with recommendations focused on compensation, overtime, career development, and targeted demographic strategies to reduce the 16.1% attrition rate and cut replacement costs.