

HR Analytics - Predicting Employee Attrition

Introduction

In a competitive business environment, retaining employees is critical. This project aims to understand the key factors influencing employee attrition and build a predictive model to identify those at risk of leaving. The insights help HR departments proactively reduce turnover.

Abstract

Using IBM's HR dataset, we conducted exploratory data analysis (EDA), built classification models, and identified major drivers of attrition. The project combines statistical analysis and machine learning to not only predict attrition but also explain the 'why' behind it. A dashboard-ready dataset was created for HR decision-makers.

Tools Used

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

Jupyter Notebook

Power BI (recommended)

SHAP (tried but not used due to technical limitations)

Steps Involved

1. Data Cleaning
2. Exploratory Data Analysis (EDA)
3. Model Building using RandomForestClassifier
4. Feature Importance Extraction
5. Individual Risk Analysis
6. Export for Dashboard Visualization

HR Analytics - Predicting Employee Attrition

Key Outputs & Visuals

Top Features:

- OverTime_Yes
- MonthlyIncome
- Age
- JobLevel
- TotalWorkingYears

Employee #25 Summary:

- Predicted Attrition: No
- Actual Attrition: No
- Attrition Risk Score: 0.462
- MonthlyIncome: -0.442
- Age: -1.307
- OverTime_Yes: 1.568
- YearsAtCompany: -0.336
- TotalWorkingYears: -0.816

Conclusion

The model identifies employees at risk of attrition with good accuracy and interpretability. OverTime and income were key drivers. The exported results can help HR teams proactively intervene. A dashboard integration with Power BI is recommended for real-time tracking.