HR Analytics – Employee Attrition Prediction Report

Abstract

This project focuses on analyzing employee attrition using data analytics and machine learning techniques. The objective is to identify key factors that influence employees leaving an organization and to predict future attrition. A Decision Tree classification model was built using HR data and the results were visualized using Power BI. Additionally, SHAP explainability was applied to understand and interpret the model predictions, making the solution transparent and useful for HR
decision-making.

Introduction

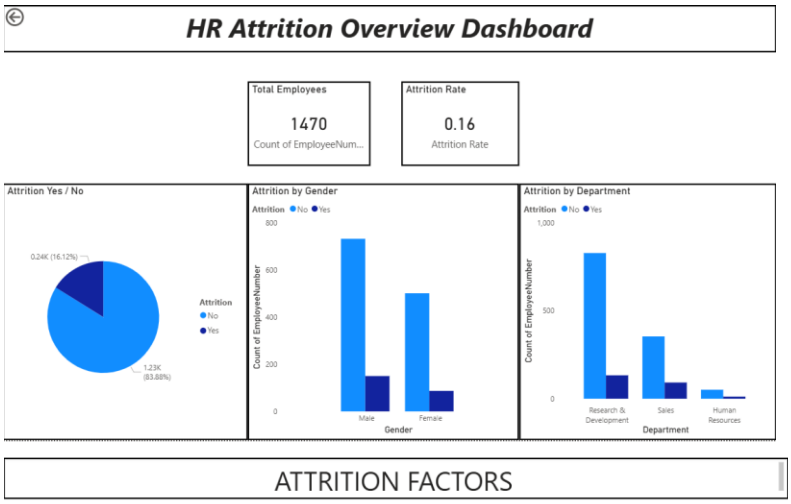
Employee attrition is a critical challenge faced by organizations as it leads to increased recruitment costs, loss of talent, and reduced productivity. HR Analytics helps organizations understand workforce behavior by analyzing employee data. In this project, historical HR data is used to explore patterns related to attrition and build a predictive model that can help HR teams proactively identify employees who are at risk of leaving.

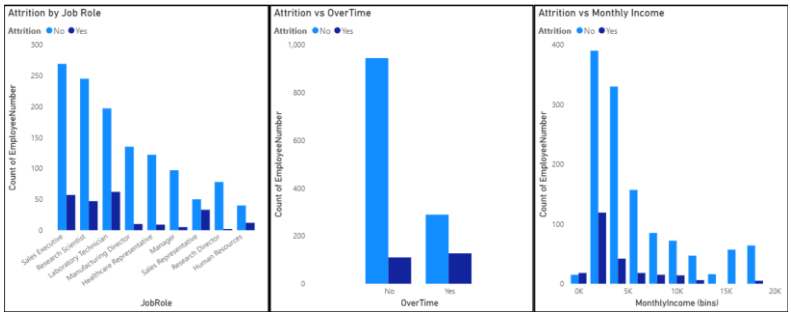
Tools Used

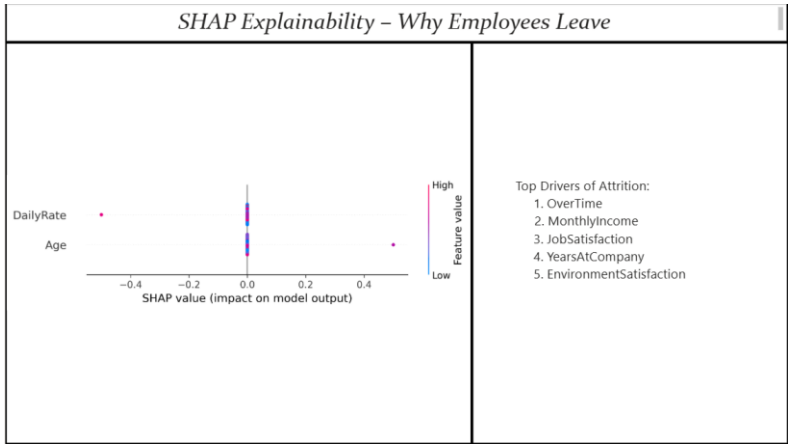
• Python: Used for data preprocessing, exploratory data analysis, and machine learning model development.
• Pandas & NumPy: Used for data handling and manipulation.
• Scikit-learn: Used to build and evaluate the Decision Tree classification model.
• SHAP (SHapley Additive exPlanations): Used for model explainability and understanding feature importance.
• Power BI: Used to create interactive dashboards for visualizing attrition trends and prediction results.

Steps Involved in Building the Project

1. Dataset Collection: An HR employee attrition dataset was collected and loaded into Python.
2. Data Understanding and Cleaning: Data types, missing values, and categorical variables were analyzed and prepared for modeling.
3. Exploratory Data Analysis (EDA): Attrition trends were analyzed based on department, job role, income, overtime, and satisfaction levels.
4. Feature Encoding: Categorical variables were converted into numerical form using one-hot encoding.
5. Model Building: A Decision Tree classifier was trained to predict whether an employee would leave the organization.
6. Model Evaluation: Model performance was evaluated using accuracy, classification report, and ROC-AUC score.
7. Explainability: SHAP values were generated to identify the most influential factors driving attrition.
8. Visualization: Power BI dashboards were created to present insights, predictions, and explainability results.







Conclusion

The HR Attrition Analytics project successfully demonstrates how machine learning and data visualization can support HR decision-making. The Decision Tree model accurately predicted employee attrition, while SHAP explainability provided transparency into the model's behavior. Power BI dashboards made the insights easy to interpret for non-technical stakeholders. This solution can help organizations take preventive actions, improve employee satisfaction, and reduce attrition rates.

By ,

Sree Lakshmi N