HR Analytics – Predict Employee Attrition

Soumya Sworupa Dash

16-05-2025

Abstract

Employee attrition means employees leaving a company, which can cause problems like lower productivity and the need to spend time and money hiring new people. In this project, we used data and simple machine learning techniques to help predict which employees might leave a company. By doing this, HR teams can take early steps to keep valuable employees and make the workplace better for everyone.

Introduction

HR Analytics involves applying data analysis and machine learning techniques to human resources data. This project focuses on predicting whether an employee is likely to leave the company and identifying the key factors influencing this decision. Such predictive insights enable HR teams to implement preventive strategies, ultimately enhancing employee retention and satisfaction.

Tools Used

- **Python (Google Colab)** – Development environment
- **Pandas, NumPy** – Data manipulation
- **Matplotlib, Seaborn** – Data visualization
- **Scikit-learn** – Machine learning model training and evaluation
- **SHAP (SHapley Additive exPlanations)** – Model interpretability and explanation

Steps Involved in Building the Project

1. **Data Loading and Exploration:** The IBM HR dataset was used, featuring information such as age, job role, satisfaction level, and work-life balance. Visualizations were created to understand attrition trends.

2. **Data Preprocessing:** Included handling missing values, encoding categorical variables, and feature scaling to prepare data for model training.

3. **Model Building:** Classification models like Logistic Regression, Decision Tree, and Random Forest were trained. Among them, the Random Forest model achieved the best performance.

4. **Model Evaluation:** Accuracy, precision, recall, F1-score, and confusion matrix were used to evaluate model performance.

5. **SHAP Value Analysis:** SHAP was applied to interpret the model predictions. It helped identify the most influential features (e.g., overtime, job satisfaction, years at company) driving individual predictions. Visual tools like SHAP summary plots and force plots were used to explain the global and local behavior of the Random Forest model.

Conclusion

Through this project, we demonstrated how machine learning can predict employee attrition with good accuracy, and how SHAP values enhance trust and transparency in the model. These insights are valuable for HR managers aiming to reduce turnover and implement data-driven retention policies.