

HR Analytics - Predict Employee Attrition

Introduction

Employee attrition is considered a significant challenge for organisations. It can lead to increased recruitment costs, loss of talent, and disruptions in work productivity. Predicting and preventing attrition can help companies to retain valuable employees and reduce unnecessary costs. This project focuses on data-driven insights to better understand the factors that contribute to employee attrition. By building predictive models that help HR identify employees who are more likely to leave.

Abstract

The objective of this project is to analyze historical employee data to uncover key factors that are contributing to attrition among employees. This project involves data collection, preprocessing, and exploratory data analysis to uncover insights and predict employee attrition by building an attrition prediction model using the employee data and logistic regression classification. Then, developing an interactive Power BI dashboard can help visualise attrition trends.

Tools and Technologies Used

The project utilises a variety of tools and technologies to preprocess the data, perform exploratory data analysis, and develop predictive models.

1. Programming Language & Modules

Python: For data cleaning, data preprocessing, exploratory data analysis, and developing prediction models. These are the required modules from the Python programming language for this project:

Pandas: Data manipulation and transformation

Seaborn: Statistical data visualisation

Matplotlib: General-purpose plotting

Scikit-learn (Sklearn): For building a logistic regression model

2. Machine Learning Model

Logistic Regression Classification: A classification model that can be used to predict the likelihood of employee attrition.

3. Model Interpretability

SHAP: Explain the individual predictions and identify feature importance.

4. Data Storage and Formats

CSV (Comma-Separated Values): Format used for storing and loading the data.

5. Development Environment

Jupyter Notebook: Interactive development environment for coding, visualisation, and documentation.

6. Visualisation and Reporting

Power BI: Create interactive dashboards for visualizing the insights.

Steps Involved in Building the Project

1. **Data Collection:** Gather historical employee data from various sources.
2. **Data Preprocessing:** Handle missing values, inconsistent data, and prepare data for analysis.
3. **Exploratory Data Analysis:** Uncover trends, correlations, and patterns in the data.
4. **Building a Prediction Model:** Train a logistic regression model to predict attrition among employees, then evaluate the accuracy, precision, and recall of the model.
5. **SHAP value analysis:** Analyse model predictions at both global and individual levels.
6. **Power BI Dashboard:** Create visualisations to show trends, patterns and insights.
7. **Employee Attrition Prevention Suggestions:** Based on insights and predictions, provide some suggestions on how to prevent attrition among employees.

Conclusion

This HR analytics project shows how data science can be applied to predict and help reduce employee attrition by identifying key factors that contribute to attrition among employees so that the HR teams can make more informed decisions. The use of SHAP analysis can ensure transparency in predictions, while a Power BI dashboard can provide a better understanding of the gathered insights by visualising the data.