HR Analytics - Predict Employee Attrition

Name: Tetali Durga Venkata Reddy

Email: tetalidurgavenkatareddy@gmail.com

Introduction

Employee attrition is a critical concern for organizations as it directly impacts operational efficiency, team morale, and financial performance. High employee turnover not only increases recruitment and training costs but also disrupts business continuity. Therefore, understanding the key drivers of employee resignation and predicting future attrition is essential for strategic workforce planning.

Abstract:

This project aims to predict employee attrition by analysing the main causes behind resignation. By using various machine learning techniques, we can uncover patterns within employee data, providing businesses with valuable insights for retention strategies.

Tools Used:

- **Python (Pandas, Seaborn)**: For data manipulation and exploratory data analysis.

- **Power BI**: For creating an interactive dashboard for visualizing attrition-related factors.

- **Scikit-learn**: For building and evaluating classification models such as Logistic Regression or Decision Trees.

Steps Involved in Building the Project

1. Data Collection and Preprocessing:

- Import the HR dataset.

- Clean the data by handling missing values, duplicates, and inconsistent formatting.

- Convert categorical variables (e.g., department, salary band) to numerical formats using encoding techniques (Label Encoding or One-Hot Encoding).

2. Exploratory Data Analysis (EDA):

- Department-wise Attrition: Calculate and visualize attrition rates per department using Seaborn bar plots.

- Salary Bands: Analyze how salary bands correlate with attrition by plotting salary bands against attrition.

- Promotions: Explore the relationship between employee promotions and attrition. For example, employees who received promotions vs. those who didn't.

- Other Features: Visualize other potentially significant features (age, tenure, job satisfaction, etc.) in relation to attrition.

3. Feature Selection:

- Identify the most influential features related to employee attrition.

- Use correlation matrices and statistical tests to check feature significance.

4. Building the Classification Model:

- Split the dataset into training and testing sets.

- Train a Logistic Regression or Decision Tree model using the training set.

- Evaluate the model using accuracy, precision, recall, and F1-score.

- Plot the confusion matrix to visualize model performance.

5. Model Interpretation with SHAP Values:

- Use SHAP (SHapley Additive exPlanations) to explain the model's predictions.

- Visualize the impact of each feature on the likelihood of attrition.

6. Visualization in Power BI:

- Create an interactive Power BI dashboard to display key insights:

  o Attrition trends over time.

  o Attrition rates by department, salary band, and promotion status.

  o Factors most affecting attrition, such as age, job satisfaction, or tenure.

  o Predicted attrition rates based on model output.

7. Deliverables:

- Power BI Dashboard: A comprehensive dashboard that presents the visual findings.

- Model Accuracy Report & Confusion Matrix: A report detailing the model's accuracy and evaluation metrics along with the confusion matrix.

- PDF of Attrition Prevention Suggestions: A set of actionable recommendations based on the findings of the analysis

Conclusion:

In this project, we demonstrated how HR analytics and machine learning models can be leveraged to predict employee attrition. By combining statistical analysis and machine learning techniques, businesses can make data-driven decisions to reduce turnover and improve employee retention strategies. The Power BI dashboard serves as an interactive tool for HR teams to monitor and act on these insights effectively.