HR Analytics - Employee Attrition Prediction

Introduction:

The HR Analytics Employee Attrition Prediction project focuses on building machine learning models to predict the likelihood of employee attrition. Understanding which factors influence attrition helps HR teams to design better retention strategies.

Abstract:

Using IBM's HR Analytics dataset, data preprocessing was performed by dropping irrelevant columns and applying Label Encoding for categorical variables. Feature scaling was applied to ensure uniformity. Logistic Regression and Decision Tree classifiers were trained on the dataset to predict employee attrition. Evaluation was done using Accuracy Score, Confusion Matrix, and Classification Report.

Tools Used:

- Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
- Jupyter Notebook

Steps Involved in Building the Project:

1. Data Cleaning: Removed columns 'EmployeeCount', 'Over18', 'StandardHours', 'EmployeeNumber'.
2. Feature Engineering: Applied Label Encoding to convert categorical variables.
3. Data Splitting: Split into training and testing datasets (80/20 split).
4. Feature Scaling: Used StandardScaler for feature normalization.
5. Model Building: Built Logistic Regression and Decision Tree classifiers.
6. Evaluation: Used Accuracy Score, Confusion Matrix, and Classification Report to assess performance.

Conclusion:

Both models provided valuable predictions regarding employee attrition. These insights can help HR departments identify at-risk employees and devise proactive retention strategies.

Created by Susan Chandra Bonkuri