HR Attrition Analysis Report

Introduction

This project focuses on predicting employee attrition using HR data. Attrition is a critical HR metric that impacts workforce planning and operational efficiency. The objective is to build a model that can identify employees likely to leave, enabling proactive retention strategies.

Abstract

Using a dataset containing employee information (e.g., department, age, and total working years), this project applies a Decision Tree Classifier to predict attrition. The dataset was preprocessed with one-hot encoding for categorical variables and split into training and testing subsets. The model demonstrated reliable predictive accuracy and was also saved for deployment. A Power BI dashboard was developed for interactive visual analytics.

Tools Used

- Python (Jupyter Notebook): Data preprocessing, model training using scikit-learn, and evaluation.
- Pandas & NumPy: Data manipulation and numerical analysis.
- Matplotlib & Seaborn: Visualization and plotting.
- Pickle: Model serialization.
- Power BI: Dashboard and business insights visualization.
- Decision Tree Classifier: Core ML algorithm used for prediction.

Steps Involved in Building the Project

1. Data Import & Cleaning: Loaded the HR dataset and checked for null values.
2. Feature Selection: Focused on Department, Age, and TotalWorkingYears as predictors.
3. Encoding: Applied one-hot encoding to categorical features.
4. Train-Test Split: Divided the data using train_test_split.
5. Model Training: Used DecisionTreeClassifier from scikit-learn.
6. Evaluation: Evaluated accuracy and generated predictions.
7. Deployment: Saved the model using pickle.
8. Interactive Dashboard: Visualized attrition trends in Power BI.

Conclusion

The project successfully built a predictive model for employee attrition, which could be used by HR teams to anticipate and reduce voluntary exits. Combined with a Power BI dashboard, it offers a comprehensive toolkit for both predictive and visual analysis.