HR Analytics Dashboard Project Report

Introduction

Employee attrition is a major issue for organizations, as high turnover can impact productivity, morale, and recruitment costs. With increasing access to HR data, companies can now leverage analytics and machine learning to better understand attrition patterns. This project focuses on visualizing employee metrics and using predictive models to identify potential attrition risks. A Power BI dashboard and Python-based models in Google Colab were used to support data-driven HR decisions.

Abstract

This project aims to analyze HR data to uncover patterns in employee attrition and predict individuals who are at risk of leaving the company. We used Power BI for interactive dashboards and Google Colab with Python for implementing machine learning models. The visual dashboard includes demographic, departmental, and performance-based insights. Predictive models including Logistic Regression and Decision Tree Classifier were evaluated using standard metrics like Accuracy, Precision, Recall, and F1-Score. This solution is designed to empower HR professionals with actionable insights to retain key talent.

Tools Used

- Power BI Desktop - Used to create interactive visual dashboards.

- Google Colab - Python-based development environment for ML modeling.

- Python Libraries - pandas, numpy, matplotlib, seaborn, scikit-learn.

- Excel - For initial data cleaning and formatting.

Steps Involved in Building the Project

1. Data Collection & Preprocessing

   - Collected HR data from open-source datasets.

   - Removed nulls, converted data types, handled outliers in Excel and pandas.

2. Exploratory Data Analysis (EDA)

   - Used Python (matplotlib/seaborn) for univariate and bivariate analysis.

   - Visualized attrition rate across departments, education fields, and age bands.

HR Analytics Dashboard Project Report

3. Dashboard Design in Power BI

   - Created KPIs: Total Employees, Attrition Count, Attrition Rate, Average Age.

   - Visual elements: Bar charts, pie charts, tree maps for categorical variables.

   - Integrated filters for interactivity (e.g., gender, education, age bands).

4. Machine Learning Modeling in Google Colab

   - Applied Label Encoding and One-Hot Encoding for categorical features.

   - Split data into training and test sets using train_test_split.

   a. Logistic Regression

      - Used LogisticRegression() from scikit-learn.

      - Achieved 88% accuracy, better precision (0.67), lower recall (0.15).

   b. Decision Tree Classifier

      - Used DecisionTreeClassifier() with entropy criterion.

      - Achieved 83% accuracy, lower precision but balanced classification.

5. Model Evaluation

   - Evaluated both models using confusion matrix, classification report.

   - Compared metrics: Accuracy, Precision, Recall, and F1-Score.

6. Risk Prediction & Output

   - Predicted attrition probability for each employee.

   - Categorized into High Risk (>0.6), Medium Risk (0.3-0.6), Low Risk (<0.3).

   - Added top 20 at-risk employees to the dashboard for HR follow-up.

Conclusion

Through a combination of Power BI visualization and machine learning modeling in Google Colab, this project delivers a powerful HR analytics solution. It not only explains historical attrition trends but also predicts future risks based on employee profiles. The use of Logistic Regression and Decision Tree classifiers gave different perspectives-Logistic Regression excelled in precision, whereas Decision Tree provided better recall.