Employee Attrition Prediction Model - Project Report

Objective

To build a machine learning model that predicts whether an employee is likely to leave the company (attrition), helping HR teams make proactive decisions.

Dataset

- Source: WA_Fn-UseC_-HR-Employee-Attrition.csv
- Shape: ~35 features, 1470 rows
- Target Variable: Attrition (Yes = 1, No = 0)

Preprocessing

- Checked for null values and data types.
- Performed outlier treatment using IQR method.
- Label Encoding on categorical features.
- StandardScaler on numerical features.

Feature Engineering

Selected features: Age, MonthlyIncome, DistanceFromHome, TotalWorkingYears, JobRole, OverTime, Gender, Education.

ML Models Trained

1. Logistic Regression
2. Decision Tree Classifier
3. Random Forest Classifier
4. Gradient Boosting Classifier
5. AdaBoost Classifier

Final chosen model: Random Forest Classifier

Model Evaluation Metrics

Random Forest Classifier:

Employee Attrition Prediction Model - Project Report

- Accuracy: ~0.86
- Precision: High
- Recall: Good
- F1 Score: Balanced
- Cross-Validation (3-fold): Stable

Cross-Validation

Performed 3-fold cross-validation to ensure the models generalizability.

Deployment Interface

- Built using ipywidgets in Jupyter Notebook.
- Inputs for users with guided ranges (e.g., Age: 1860).
- Interactive UI with a Predict button.
- Real-time prediction output:

  Employee is likely to stay

  Employee may leave

Future Enhancements

- Deploy as a web app using Streamlit or Flask.
- Include SHAP or LIME for model explainability.
- Automate data updates from HR systems.