

Employee Turnover Prediction

This project aims to analyze and predict employee turnover using machine learning techniques. The project is divided into several phases, including data intake, cleaning, integration, processing, analysis, and model implementation. Below is an overview of each phase and instructions for running the notebook:

Phases:

1. ****Data Intake and Cleaning****:

- Load datasets from external sources, including employee_performance folder (which consists of multiple CSV files), cities.csv, states_ts.csv, and employees.csv.
- Perform data cleaning tasks such as handling missing values and duplicates.

2. ****Data Integration****:

- Merge multiple datasets to create a unified dataset.

3. ****Data Processing and Feature Engineering****:

- Prepare data for modeling by computing tenure, creating labels for turnover, and generating lag features.
- Perform feature engineering tasks to enhance predictive power.

4. ****Model Input Preparation****:

- Split the dataset into training, validation, and test sets.
- Encode categorical variables and impute missing values.

5. ****Analysis****:

- Analyze correlations between features and turnover.
- Visualize turnover rates over time and compute median/average tenure.
- Examine key performance metrics and their distributions.

6. ****Machine Learning Model Implementation****:

- Implement a Random Forest Classifier to predict employee turnover.
- Train the model, evaluate its performance, and analyze feature importances.

Instructions for Running the Notebook:

1. Clone the repository or download the FORE.ipynb notebook.
2. Download the required datasets:
 - employee_performance folder (which consists of multiple CSV files)
 - cities.csv
 - states_ts.csv
 - employees.csv
3. Update file paths in the notebook to point to the downloaded datasets.
4. Ensure the required libraries (e.g., pandas, scikit-learn, seaborn) are installed.
5. Execute each cell in the notebook sequentially to perform data analysis and model implementation.
6. Review the results and insights generated from each phase.
7. Refer to the README file for additional information and instructions.