Employee Attrition Project

Project Overview

The primary objective of the **Employee Attrition Project** is to analyze and predict why employees leave a company. By identifying the key factors that contribute to employee turnover, organizations can take proactive measures to reduce attrition rates, improve employee retention, and optimize their HR strategies.

Data : https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

Github : https://github.com/shvwkyyy/Employee-Attrition-Prediction

---

Tools and Libraries

The following tools and libraries will be used for this project:

- **pandas**: For data manipulation and analysis.
- **numpy**: For numerical computations.
- **matplotlib**: For data visualization.
- **seaborn**: For advanced statistical visualizations.
- **scikit-learn**: For machine learning model development and evaluation.

Project Team Members

- **Yousef Khaled Shawky** (Team Leader)
- **Ali Fathy Abdelghani**
- **Abdelrahman Mohamed Abdelrazek**
- **Amr Sabry Awad**
- **Fares Essam Mostafa**

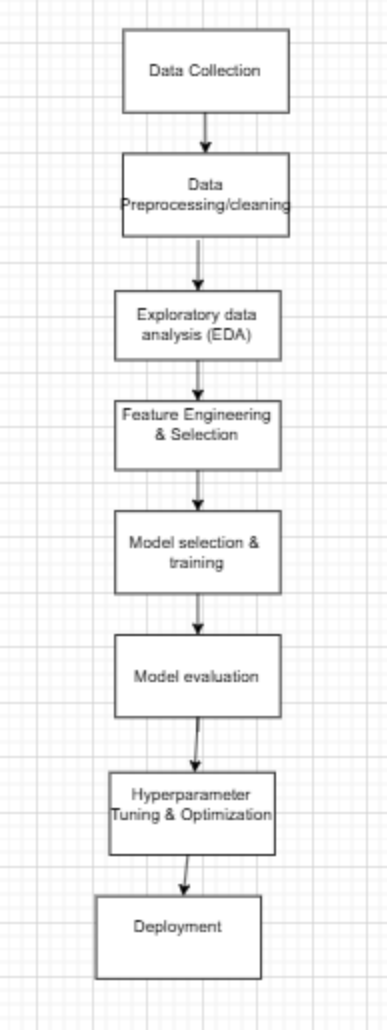
Final Project Deliverables

The final deliverables for this project will include the following:

1. **Report (PDF/Word)**: A well-structured document summarizing the project findings, methodologies, and conclusions.

2. **Presentation (PowerPoint/Google Slides)**: A slide deck for presenting key insights, visualizations, and results to stakeholders.
3. **Jupyter Notebook (.ipynb)**: A notebook containing the code, data analysis, and visualizations.
4. **Dashboard (Power BI/Tableau)**: An interactive dashboard designed for HR professionals to explore attrition insights.
5. **API & Deployment**:
    - **Flask/FastAPI Script**: A backend service to serve model predictions.
    - **Dockerfile**: Instructions for containerizing the application.
    - **Docker Compose File (Optional)**: For managing multiple services (e.g., database, API).
    - **Environment Variables & Configs**: .env file for managing secrets and configuration settings.

Machine Learning Pipeline



Timeline and Deliverables

Week 1-2 (8/2 to 21/2)

Deliverables:

- Define the project scope, objectives, and success criteria.
- Collect and explore the dataset.
- Assign roles and responsibilities to team members.
- Set up a well-structured work environment (e.g., GitHub/GitLab repository).

Team Members' Roles:

- **Ali**: Research dataset sources and clean the data.
- **Fares**: Set up the project repository and documentation.

- **Yousef**: Develop a plan for feature engineering and data preprocessing.
- **Amr**: Research potential machine learning models.
- **Abdelrahman**: Research project requirements, tech stack, and best practices.

Week 3-4 (22/2 to 7/3)

Deliverables:

- Identify and handle missing values and outliers.
- Perform data visualization and correlation analysis.
- Preprocess categorical and numerical features.
- Prepare the cleaned dataset for model training.

Team Members' Roles:

- **Ali**: Detect and handle missing values and outliers (using IQR, Z-score, and imputation techniques).
- **Yousef**: Perform data visualization (e.g., histograms, boxplots, heatmaps) and analyze feature relationships.
- **Fares**: Apply feature engineering techniques (e.g., encoding categorical variables, scaling numerical data).
- **Abdelrahman**: Address class imbalance issues (e.g., check distribution, apply SMOTE if needed).
- **Amr**: Document findings, summarize insights, and save the final processed dataset.