Github Link: https://github.com/abdulraseeth2006

Project Title: Predicting Customer Churn Using Machine Learning to Uncover Hidden Patterns

PHASE-2

- Problem Statement

Customer churn is a major concern for businesses, especially in highly competitive markets.

The ability to predict which customers are likely to leave allows companies to take proactive steps to retain them.

This project aims to build a machine learning model that can identify churn-prone customers based on historical data and behavioral indicators.

By uncovering hidden patterns and key predictors of churn, companies can improve customer satisfaction,

reduce revenue loss, and enhance marketing strategies. This classification problem uses customer demographic

and service usage data to predict the binary target variable: churn (yes/no).

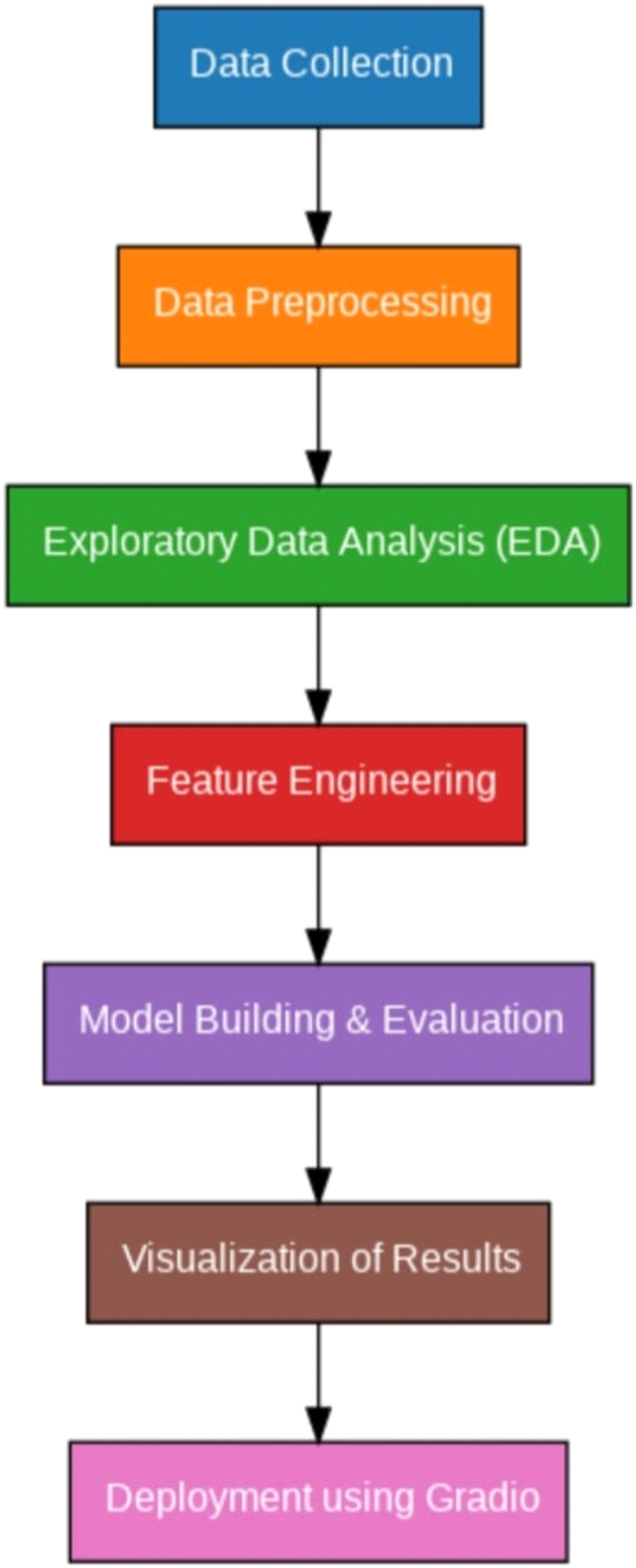
- Project Objectives

• Build a machine learning classification model to predict customer churn.

• Analyze and identify key features that influence churn decisions.

- Provide interpretable insights to help businesses take preventive action.
- Develop a Gradio interface for easy input and prediction testing.
- Ensure high model performance with proper preprocessing and validation.

3. Flowchart of the Project Workflow



- Data Description

- Dataset Name: Telco Customer Churn Dataset

- Source: https://www.kaggle.com/code/bhartiprasad17/customer-churn-prediction

- Type of Data: Structured tabular data

- Target Variable: Churn yes

- Features: Demographic data (age, gender), service usage (internet, phone, support calls), billing details

- Dataset :https://www.kaggle.com/competitions/public-telecom-customer-churn-analysis-and-prediction

- Data Preprocessing

- Checked for and handled missing values.

- Encoded categorical variables using one-hot or label encoding.

- Scaled numerical features using StandardScaler.

- Removed irrelevant or highly correlated features.

- Exploratory Data Analysis (EDA)

- Univariate analysis of churn distribution.

- Correlation matrix for numerical predictors.

- Churn rate breakdown by contract type, tenure, internet service, etc.

- Key finding: Contract type and tenure are strong indicators of churn.

- Feature Engineering

• Derived binary features for high-risk indicators.

• Created interaction terms between payment type and contract length.

• Removed multicollinearity through variance inflation factor analysis.

- Model Building

• Algorithms Used: Logistic Regression, Random Forest Classifier

• Train-Test Split: 80/20 with fixed random_state

• Evaluation Metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC

- Visualization of Results

• Feature importance plotted from Random Forest.

• Confusion matrices and ROC curves compared across models.

• Integrated Gradio interface for live predictions.

- Tools and Technologies

• Language: Python 3

• Environment: Google Colab

• Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, Gradio

- Team Members and Contributions

• M. Abdul Raseeth

• Data preprocessing

- Feature engineering
- Final integration and interface

- P. Abiharan
  - Exploratory Data Analysis
  - Model selection and training
  - Visualization and performance evaluation

- K. Adhithya
  - Problem definition and documentation
  - Data sourcing and cleaning
  - Reporting and presentation