**Predicting customer churn using machine learning to uncover hidden patterns**

STUDENT NAME : AMRINBANU.S

 REGISTER NUMBER : 411823106003

INSTITUTION : RRASE COLLEGE OF ENGINEERING

Date of submission :10-05-2025



**GITHUB LINk:**

<https://github.com/Amrinbanu-sathakathullah/amrin-naan-mudhalvan.git>

1.Problem Statement

The objective of this project is to develop a machine learning- based solution that can

accurately predict customer churn by analyzing historical customer data. By uncovering

underlying patterns and risk factors associated with churn, the system w enable businesses to

peroactively engage at-risk customers, tailor retention strategies, and ultimately improve

customer loyalty and revenue stability.

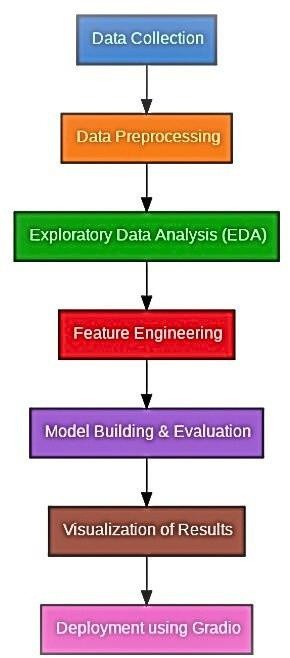
# 2.Project Objectives

* + To analyze customer behavior data.
  + To build a predictive model that classifies whether customer will churn.
  + To uncover hidden patterns through exploratory data

analysis and feature engineering.

* + To provide actionable insights for reducing churn.

# 3.Flowchart of the Project Workflow



4.Data Description

* + Target Variable: Churn (Yes/No)
  + Features: Customer demographics, usage patterns, service subscriptions, tenu billing, and support interactions.- Example Distributions:
    - Gender: 50% Male, 50% Female
    - Internet Service: 70% Fiber Optic, 30% DSL
    - Churn Rate: ~27%

-Datasetlink: https:/[/www.kaggle.com/datasets/blastchar/telco-customer-churn](http://www.kaggle.com/datasets/blastchar/telco-customer-churn)

# 5.Data Preprocessing

- Handling missing values

* + Encoding categorical variables
  + Scaling numerical features
  + Addressing class imbalance using techniques like SMOTE or under-sampling)

# 6.Exploratory Data Analysis(EDA)

- Correlation heatmaps

* + Churn vs. Tenure, MonthlyCharges, Contract type
  + Histograms and boxplots to understand distributions and outliers

# 7.Feature Engineering

* + Creation of new features (e.g., total charges per tenure)
  + Binning of numerical variables
  + Feature selection using mutual information and recursive feature elimination

# 8.Model Building

* + Algorithms used: Logistic Regression, Decision Tree, Random Forest, XGBoost
  + Performance metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC
  + Cross-validation for robust evaluation

# 9.Visualization of Results & Model Insights

* + Confusion matrix
  + ROC Curve
  + Feature importance plots
  + SHAP values for interpretability

# 10.Toolsand Technologies Used

* + Programming Language: Python
  + Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, shap
  + IDE: Jupyter Notebook / VS Code
  + Version Control: Git & GitHub

# 11.Team Members and Contributions

|  |  |
| --- | --- |
| AMRINBANU.S | Contribution |
| UMAMAHESHWARI.R | Data preprocessing, model building |
| AATHISH.S | EDA, visualization, feature engineering |
| DHANUSH.A | Report writing, GitHub management |

-