Phase-1

Student Name: Aswin P.G

Register Number: 410723106005

Institution: Dhanalakshmi college of engineering

Department: Electronics and Communication Engineering

Date of Submission: 28/04/2025

1.Problem Statement

PREDICTING COSTOMER CHURN USING MACHINE LEARING TO UNCOVER HIDDEN PATTERNS.

With rising competition in various industries, retaining existing customers has become crucial for businesses. Traditional methods often fail to predict customer churn effectively due to the complex, hidden patterns influencing customer behavior.

This project uses AI and Machine Learning to predict customer churn and uncover hidden behavioral patterns, helping businesses proactively engage and retain customers.

2.Objectives of the Project

1. Develop a machine learning model to accurately predict customer churn.
2. Uncover hidden patterns and key factors influencing churn.
3. Minimize false positives while maximizing the recall of churn cases.

4. Evaluate and compare multiple ML algorithms for optimal performance.

5. Present findings through intuitive visualizations and (optionally) create a dashboard for real-time prediction simulation.

3.Scope of the Project

☐ Use customer behavior and service usage datasets to train churn ☐prediction models. Apply preprocessing, handle missing values, and manage class imbalance.

☐ Experiment with various ML models like Decision Trees, Random Forest, XGBoost, etc.

☐ Constraints:

  o Limited to publicly available anonymized datasets.

  o No access to real-time customer systems. ₒ Deployment limited to simulation environments (Streamlit app or web demo).

4.Data Sources

- Dataset: Customer Churn Prediction Dataset

- Source: Kaggle, IBM Sample Dataset, or Telco Customer Churn Dataset

- Type: Public

- Nature: Static (downloaded once, not updated in real-time)
  Description: Contains anonymized customer data including demographics, account details, and churn status.

Data Set Link:

https://www.kaggle.com/datasets/bhuviranga/customer-churn-data

5.High-Level Methodology

- Data Collection

  Download the dataset from Kaggle.

- Data Cleaning

  Handle missing values, encode categorical variables, remove duplicates, ensure consistent formatting.

- Exploratory Data Analysis (EDA)

• Use bar charts, heatmaps, and pair plots to explore features. • Visualize churn rates across different customer segments.

- Feature Engineering

☐ Feature scaling and transformation.

☐ Create new interaction features (e.g., tenure vs monthly charges).   Apply
☐ dimensionality reduction (e.g., PCA) if necessary.

● Model Building –   Experiment with:

☐   Logistic Regression

☐   Decision Tree Classifier

☐   Random Forest Classifier

☐   XGBoost Classifier

☐

LightGBM

    (Optionally) Artificial Neural Networks (ANN)

● Model Evaluation –

    Use metrics such as:

·   Confusion Matrix

·   Precision, Recall, F1-Score

·   ROC-AUC

·   Cross-validation

● Visualization & Interpretation

    Visualize churn patterns, feature importance, and model comparison using
matplotlib, seaborn, and plotly..

- Deployment –

Simulate predictions using a Streamlit web application where users can input customer attributes and receive churn risk predictions.

6.Tools and Technologies

● Programming Language: Python

● Notebook/IDE: Google Colab or Jupyter Notebook   ● Libraries:

- Data Processing: pandas, numpy

- Visualization: matplotlib, seaborn, plotly

- Modeling: scikit-learn, xgboost, lightgbm, imbalanced-learn

- Deployment: Streamlit or Flask

7.Team Members and Roles

S.NO	NAMES	ROLE	RESPONSIBILITY
1.	Sudharsan.V.S	Member	Visualization & cleaning
2.	Aswin.P.G	Member	Exploratory data analysis(EDA) Feature engineering
3.	Sudharsan.V.S	Member	Model building,model evalu

| 4. | Yogesh.J | Leader | Data collection & cleaning |