CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

Student Name: V.SIVAVETRIVEL

Register Number: 511023106003

Institution: 5110 - JEI MATHAAJEE COLLEGE OF ENGINEERING

Department: B.E(ECE)

Date of Submission: 25.04.2025

---

1. Problem Statement

To develop a machine learning model that can **predict whether a customer is likely to churn (leave the service)**

2. Objectives of the Project

To predict whether a customer is likely to **churn** (leave a company) using historical data, so that the business can take proactive steps to **retain** them.

3. Scope of the Project

The **scope of Customer Churn Prediction using Machine Learning** is broad and highly valuable across industries

4. Data Sources

i. Customer Demographics

ii. Account Information

iii. Payment and Billing History

iv. Usage and Engagement Data

v. Customer Support Interactions

vi. Marketing & Communication History

5. High-Level Methodology

1. Data Collection
Obtain data from CRM systems, logs, billing systems, or public datasets like the Telco Churn Dataset.

2. Data Preprocessing
- Handle **missing values**
- Convert categorical variables using **encoding**
- Scale numeric features (if needed)
- Clean or drop irrelevant columns (e.g., customer ID)

3. Feature Engineering
- Create new features (e.g., engagement_score, tenure_bucket)
- Transform skewed variables
- Bin continuous features for decision trees

4. Model Selection
Common models used:
- **Logistic Regression** (baseline)
- **Decision Tree / Random Forest**

5. Model Evaluation
- Use train-test split or cross-validation
- Metrics:
    - **Accuracy** (overall correctness)
    - **Precision & Recall** (important for imbalanced classes)
    - **F1 Score**
    - **ROC-AUC** (to evaluate probability outputs)

6. Interpret Results
- Use SHAP or feature importance to explain predictions
- Identify top factors that lead to churn (e.g., short tenure, high charges)

7. Actionable Steps
- Target high-risk customers with:
    - Discounts
    - Personalized offers
    - Loyalty programs
    - Improved service quality

6. Tools and Technologies

1. Data Collection & Storage

These tools help gather and store raw data about customers, transactions, and interactions:

- **SQL / MySQL / PostgreSQL** – For querying customer databases
- **Excel / CSV Files** – For manual or exported data
- **CRM Tools** (e.g., Salesforce, HubSpot) – Customer profiles and behavior
- **Cloud Platforms** (e.g., AWS S3, Google BigQuery) – Scalable storage
- **APIs** – To pull in data from external systems or application

2. Data Preprocessing & Analysis

Used for cleaning, exploring, and preparing data before modeling:

- **Python** – Most popular language for ML tasks
- **Pandas** – For data manipulation and preprocessing
- **NumPy** – For numerical operations
- **Matplotlib / Seaborn** – For data visualization

3. Machine Learning Modeling

Tools and libraries for building and evaluating predictive models:

- **Scikit-learn** – Core ML models (Logistic Regression, Random Forest, etc.)
- **XGBoost / LightGBM / CatBoost** – High-performance gradient boosting libraries
- **TensorFlow / Keras / PyTorch** – For deep learning models

4. Feature Engineering & Selection

Used to create and select the best features to improve model performance:

- **Scikit-learn** – Has built-in tools for feature selection
- **Featuretools** – For automated feature engineering
- **SHAP / LIME** – For interpreting feature importance

5. Model Evaluation & Tuning

Helps fine-tune models and understand their performance:

- **Scikit-learn** – Cross-validation, confusion matrix, ROC curves
- **Optuna / Hyperopt** – For hyperparameter tuning
- **Yellowbrick** – For visual model diagnostics and performance plots

6. Deployment

Used to deploy churn models into production:

- **Flask / FastAPI** – Lightweight Python APIs for serving models
- **Docker** – For packaging ML models as containers
- **Streamlit / Gradio** – For building interactive dashboards

7. Reporting & Business Intelligence

Used to present churn insights to non-technical stakeholders:

Tableau / Power BI – Visual dashboards for business teams
Looker / Metabase – Data exploration for decision-makers
Excel – Still common for quick reports and summaries

7. Team Members and Roles

1. K.ABIKA – (Data collection,cleaning)
2. T.R.VIGNESH – (tools and technology)
3. S.KIRUBANANTHAM – (high level methodology)