



Phase-1 Submission Template

Student Name: RAGUL D

Register Number : 23UIT039

Institution: AVS COLLEGE OF TECHNOLOGY

Department: B.TECH/INFORMATION TECHNOLOGY

Date of Submission:28/04/2025

1. Problem Statement

predicting customer churn using machine learning to uncover hidden patterns

2 .Objectives of the Project:

- *Use historical data to detect patterns that indicate a customer is likely to leave.*
- *Uncover key factors (e.g., low engagement, high complaints, billing issues) contributing to customer churn.*
- *Provide actionable insights to reduce churn by targeting vulnerable customer segments.*
- *Develop and compare machine learning models (e.g., logistic regression, decision trees, random forest, XGBoost) for churn prediction.*

3.Scope of the Project



- **Data Collection and Preprocessing:** Collect historical customer data (demographics, transaction history, service usage, support interactions, etc.) and clean it for analysis.
- **Exploratory Data Analysis (EDA):** Analyze patterns, trends, and correlations to understand factors influencing churn.
- **Feature Engineering:** Create new meaningful features from raw data to improve model performance.
- **Model Development:** Train various machine learning models (e.g., logistic regression, decision trees, random forest, gradient boosting, neural networks) to predict churn.
- **Model Evaluation:** Assess models using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to select the best-performing one.
- **Pattern Discovery:** Use model insights (e.g., feature importance) to reveal hidden patterns and key churn drivers.
- **Deployment and Monitoring:** Optionally deploy the model for real-time prediction and monitor its performance over time.

4.Data Sources

- **CRM data:** Customer profiles, sign-up details.
- **Transactional data:** Purchase history, payment patterns.
- **Support logs:** Number of complaints, resolution time.
- **Behavioral data:** App/website usage frequency, login patterns.

5.High-Level Methodolog

- **Data Collection:** data from CRM, billing, and usage
- **Data Cleaning:** Handle missing values, remove duplicates, and fix inconsistencies.
- **Exploratory Data Analysis (EDA):** Analyze trends, distributions, and churn



patterns.

- **Feature Engineering:** Create meaningful features like tenure, usage frequency, etc.
- **Model Building:** Train ML models (e.g., Logistic Regression, Random Forest, XGBoost).
- **Model Evaluation:** Assess performance using metrics like accuracy, precision, recall, and ROC-AUC.
- **Visualization & Interpretation:** Use graphs to explain key churn drivers and model outputs.
- **Deployment:** Deploy the model into a production environment for real-time

6.Tools and Technologies

- **Programming Language:** Python
- **Notebook/IDE:** Jupyter Notebook or VS Code
- **Libraries:** pandas, NumPy, scikit-learn, matplotlib, seaborn, XGBoost
- **Optional Tools for Deployment:** Flask, Streamlit, Docker, Heroku or AWS

7.Team Members and Roles

TEAM LEADER: RAGUL D

RESEACHAR: AKASH E

DEVELOPER: HARISH S

DESIGNER:SUBASH C

TESTER: FRANKLIN M