Phase-1 Submission

Student Name: Saghana.K.S

Register Number: 410723104069

Institution: Dhanalakshmi college of Engineering

Department: Computer Science and Engineering

Date of Submission: 30-04-2025

1.Problem Statement

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

The project aims to address customer churn in subscription-based business models. High churn rates can lead to significant financial loss and reputational damage. Understanding why customers leave and identifying those at risk is crucial for retention and revenue growth.

2.Objectives of the Project

❖ *Predict churn risk using customer behavioural data.*
❖ *Identify key factors influencing customer attrition.*
❖ *Segment customers for targeted marketing and retention strategies.*
❖ *Improve customer engagement and reduce churn rates.*

3.Scope of the Project

❖ ***Features to Analyse:*** *Customer demographics, membership data, internet usage, transaction behaviour, complaints, feedback.*

❖ **Limitations/Constraints:** *Dataset is pre-defined and static; limited to historical data without real-time updates; deployment scope is not clearly defined in the report.*

4.Data Sources

❖ **Source:** [Dataset link](#)
❖ **Type:** *Static.*
❖ **Content:** *36,992 rows and 25 columns including variables such as customer ID, membership category, login frequency, wallet points, complaint status, and churn risk score.*

5.High-Level Methodology

❖ **Data Collection** – *Dataset obtained from project repository.*
❖ **Data Cleaning** –
o *Type casting date columns.*
o *Handling missing values (region_category=, points_in_wallet).*
o *Correcting erroneous churn scores.*
o *Dropping insignificant or error-prone columns (e.g., customer_id, name, referral_id).*
❖ **Exploratory Data Analysis (EDA)** –
o *Univariate and bivariate analysis.*
o *Count plots for gender, region, membership.*
o *Heatmap for correlation.*
❖ **Feature Engineering** – *Cleaned and engineered base model-ready dataset.*
❖ **Model Building** – *Base model was built but specific algorithms not detailed in the document.*
❖ **Model Evaluation** – *Initial model accuracy presented, but metrics/validation strategies not elaborated.*
❖ **Visualization & Interpretation** -- *Used plots and charts for EDA; insights derived from visual patterns.*
❖ **Deployment** – *Not covered explicitly in the document.*

6.Tools and Technologies

Programming Language – *Python*.

- ❖ **Notebook/IDE** – *not specified (likely Google Colab / Jupyter Notebook)*.
- ❖ **Libraries** – *pandas, numpy, seaborn, matplotlib(assumed based on EDA)*.
- ❖ **Optional Tools for Deployment** – *Not mentioned.*

7.Team Members and Roles

Team Members:	Roles:
Vidhya.S	Team Leader
Santhanayaki.M	Member
Saghana.K.S	Member
Rakshi.D	Member