**Phase-1 Submission Template**

**Student Name:** V.ASWINI

**Register Number:** 511523205010

**Institution:** P. T. Lee Chengalvaraya Naicker College of Engineering and Technology

**Department:** B. Tech Information Technology / 2 year

**Date of Submission**: 26.04.2025

**1.Problem Statement**

* Customer churn is a significant challenge across industries, especially for subscription-based
* services. Losing existing customers leads to revenue loss and increased customer acquisition
* costs. This project aims to predict churn by analyzing behavioral patterns and identifying risk
* factors, enabling businesses to take proactive measures to retain customers.

**2. Objectives of the Project**

* Develop a machine learning model to predict customer churn.
* Uncover patterns and factors contributing to churn using EDA.
* Generate actionable insights to improve customer retention strategies.
* Evaluate multiple models to determine the best-performing one.

**3. Scope of the Project**

* Analyze customer behavior, service usage, demographics, and interaction history.
* Develop classification models (e.g., logistic regression, random forest).
* Limitations: Restricted to publicly available datasets, no real-time deployment in this phase.

**4. Data Sources**

* **Dataset:** [e.g., Telco Customer Churn Dataset from Kaggle]
* **Source:** Kaggle (Public)
* **Type:** Static

**5. High-Level Methodology**

* **Data Collection:**

Download dataset from Kaggle.

* **Data Cleaning:**

Handle missing values, remove duplicates, standardize formats.

* **Exploratory Data Analysis (EDA):**

Use visualizations (e.g., histograms, heatmaps) to explore

correlations and trends.

* **Feature Engineering:**

Convert categorical data, normalize features, create churn-related indicators.

* **Model Building:**

Try logistic regression, decision trees, random forests, and XGBoost.

* **Model Evaluation:**

Use accuracy, precision, recall, F1-score, ROC-AUC for evaluation.

* **Visualization & Interpretation:**

Present results using bar charts, confusion matrix, feature

importance plots.

* **Deployment:**

(Optional) Deploy model using Streamlit for demo purposes.

6. Tools and Technologies

* **Programming Language:** Python
* **Notebook/IDE:** Google Colab
* **Libraries:** pandas, numpy, matplotlib, seaborn, scikit-learn, XGBoost
* **Deployment Tools (Optional):** Streamlit

7. Team Members and Roles

1. **J.JANANI**

Data collection, cleaning, EDA, modeling, documentation

1. **V.ASWINI**

Feature engineering, model evaluation, visualization

1. **K.SAGUNTHALA**

Final report, presentation

1. **R.KEERTHANA**

optional deployment