**Phase-1 Submission Template**

**Student Name:** Nivithrasri.V

**Register Number:** 712523104040

**Institution:** PPG Institute of Technology

**Department:** BE Computer Science and Engineering

**Date of Submission:** 25.04.2025

**1.Problem Statement**

Customer churn is a major challenge for businesses, particularly in industries like telecommunications, retail, and banking. Predicting churn helps businesses identify at-risk customers, enabling proactive retention strategies and improving profitability.

**2.Objectives of the Project**

The goal is to build a machine learning model to predict customer churn, identify key factors influencing churn, and generate insights for business retention strategies.

**3.Scope of the project**

The project will analyze customer data, explore churn-related patterns, and build a predictive model. The focus is on using a specific dataset, and deployment may not be fully implemented in the initial phase.

**4.Data Source**

*The dataset will be sourced from Kaggle and UCI machine learning repository. It is a public, static dataset.*

*Link:* [*Kaggle: Your Machine Learning and Data Science Community*](https://www.kaggle.com/)

**5.High-Level Methodology**

* **Data Collection:** *Data will be downloaded from Kaggle.*
* **Data Cleaning*:*** *Handle missing values, remove irrelevant features, and standardize formats.*
* **EDA**: *Explore data using visualizations to uncover trends*.
* **Feature Engineering:** *Create or transform features like tenure or usage patterns*.
* **Model Building*:*** *Experiment with models like logistic regression, decision trees, and random forests.*
* **Model Evaluation:** *Evaluate using accuracy, precision, recall, F1-score, and ROC-AUC.*
* **Visualization & Interpretation:** *Present insights through visualizations*.
* **Deployment:** *Showcase via a Jupyter Notebook or interactive dashboard* .

**6.Tools and Technologies**

* **Programming Language**: *Python.*
* **Notebook/IDE:** *Visual studio code*.
* **Libraries*:*** *pandas, numpy, seaborn, matplotlib, scikit-learn,tensor flow*.
* **Optional Deployment Tools:** *Streamlit, Flask.*

**7.Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| ***NAME*** | ***ROLE*** | ***DESCRIPTION*** |
| *Krisha.A.P* | *Project Manager* | *Oversees the entire project,*  *Assigns tasks and ensures deadlines are met,*  *Communicates progress to mentors/judges,*  *Manages documentation and final,presentation* |
| *Nivitharasri.V* | *Analyst & Visualizer* | *Evaluates model performance,Uses SHAP/LIME to uncover hidden patterns,Builds graphs, dashboards, and reports for stakeholders* |
| *Ganga.SS* | *Machine Learning Developer* | *Builds and trains ML models,Experiments with algorithms, Tunes hyperparameters for best performance* |
| *Hirudhanya.S* | *Data Engineer* | *Collects and cleans the dataset,*  *Handles missing values, encoding, and feature engineering,Ensures the data is model-ready.* |
| Vijay.S | *Deployment* | *Sets up the project in GitHub or Colab, Deploys the model via Flask/Streamlit , Ensures reproducibility and system integration* |