**Phase-1 Submission**

**Student Name:** Ganga.ss

**Register Number:** 712523104020

**Institution:** PPG Institute of Technology

**Department:** BE Computer Science and Engineering

**Date of Submission:** 26.04.25

**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 Sources**

*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.*
* **Exploratory Data Analysis (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**

[List the tools, programming languages, and libraries you plan to use in your project. Include the following details

* **Programming Language** – *Python.*
* **Notebook/IDE** – *Google Colab, Jupyter Notebook, VS Code.*
* **Libraries** – *pandas, NumPy, seaborn, matplotlib, scikit-learn, TensorFlow.*
* **Optional Tools for Deployment** – *Stream lit, Flask, Gradio, Fast API.*

**7.Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| **Name** | **Role** | **Description** |
| *Krisha A P* | *Team Lead* | *Oversees the entire project, assigns tasks and ensures deadlines are met, Communicates progress to mentors.* |
| *Nivithrasri V* | *Analyst &Visualizer* | *Evaluates model performance, Uses SHAP/LIME to uncover hidden patterns, Builds graphs, dashboards.* |
| *Ganga S S* | *ML Developer* | *Builds and trains ML models, Experiments with algorithms.* |
| *Hirudhanya S* | *Data Engineer* | *Collects and cleans the dataset, handles missing values, encoding, and feature engineering, Ensures the data is model-ready.* |
| *Vijay S* | Testing & *Deployment* | *Ensures reproducibility and system integration* |