Student Name: BHARATHI R

Register Number: 422623106009

Institution: UNIVERSITY COLLEGE OF ENGINEERING,PANRUTI

Department: ECE

Date of Submission: [Insert Date]

1.Problem Statement

- In today's highly competitive business environment, retaining existing customers is more cost-effective than acquiring new ones. However, many companies face challenges in identifying customers who are likely to stop using their services,a phenomenon known as customer churn. This project aims to develop a machine learning-based solution to predict customer churn by analyzing historical customer data. The goal is to uncover hidden patterns and behavioral indicators that contribute to customer attrition

2.Objectives of the Project

3. **Identify factors contributing to churn**: Pinpoint significant variables and patterns that influence customer attrition, such as usage frequency, service quality, and customer demographics.

4. **Build an accurate predictive model**: Develop a machine learning model capable of predicting which customers are at high risk of churning, allowing businesses to take preventive action

5. **Provide actionable insights**: Offer data-driven recommendations to improve customer retention strategies and enhance overall customer satisfaction.

6. **Improve decision-making**: Enable businesses to make informed, proactive decisions regarding customer engagement, loyalty programs, and personalized offerings.

7.Scope of the Project

- This project focuses on analyzing customer data to predict churn using machine learning algorithms. It covers data preprocessing, feature selection, model training, and evaluation, with the goal of identifying key patterns and factors driving customer attrition. The scope is limited to available customer-related data, excluding external market trends or non-quantifiable customer sentiments.

8.Data Sources

- The data for this project was sourced from publicly available telecommunications datasets on the internet, specifically focusing on customer demographics, service usage, billing history, and customer support interactions. These datasets are typically used in the telecommunications sector to analyze customer behavior and predict churn, providing valuable insights into factors that contribute to customer retention and attrition.

9.High-Level Methodology

- **Data Collection** - Gather relevant customer data from multiple internal sources, including demographics, transaction history, customer interactions, and feedback.

- **Data Cleaning** - Clean and preprocess the data to handle missing values, outliers, and categorical variables. Normalize and standardize features as required for model input.

- **Exploratory Data Analysis (EDA)** - The EDA phase involved analyzing the dataset to identify key patterns and relationships between customer attributes and churn. This included investigating the distribution of features, detecting missing values, visualizing correlations, and examining how various factors such as service usage, payment history, and customer demographics influence churn. The insights from this analysis guided the feature engineering process and helped identify important predictors for the churn prediction model.

- **Feature Engineering** - Identify and create meaningful features that can help predict churn, such as customer activity trends, service usage patterns, and engagement metrics.

- **Model building -**The model was built using machine learning algorithms like Logistic Regression and Random Forest to predict customer churn. The data was split into training and testing sets, and models were trained on the training set. Performance was evaluated using metrics like accuracy and F1-score, with the best-performing model selected and optimized for better predictions.

- **Visualization & Interpretation** -Visualizations revealed key patterns in churn, such as higher rates among month-to-month contract customers and those with frequent support interactions. Longer-tenured customers and higher service usage were associated with lower churn, providing insights for targeted retention strategies through chart or bar graph

- **Deployment** – This project will be deployed on **Google Colab**. The trained churn prediction model will be saved and then uploaded to the Colab environment for real-time predictions. Users will be able to input new customer data via the notebook interface, and the model will return churn predictions

10. Tools and Technologies:

- **Programming Language** – python

- **Notebook/IDE -** **Google Colab**: The primary environment for running the project, as it provides easy access to GPU acceleration and collaborative sharing.

  Jupyter Notebook: Integrated within Google Colab for interactive coding and data visualization.

  Google Drive: Used for storing datasets, models, and results for easy access and sharing within Google Colab.

- **Libraries** - ***Pandas***: For data manipulation, cleaning, and Preprocessing

  NumPy*:* For numerical computing and handling arrays

  Matplotlib: For creating static, animated, and interactive visualizations of data

  Seaborn*:* For statistical data visualization and creating informative charts (e.g., heatmaps, count plots).

Scikit-learn*:* For building and evaluating machine learning models (e.g., Logistic Regression, Decision Trees, Random Forest).

XGBoost: A popular gradient boosting library for more advanced predictive modeling.

Pickle*:* For saving and loading machine learning models.

- **Tools for Deployment** – ***Google Colab***: As the deployment platform, where the trained model will be used to make real-time churn predictions.

11. Team Members and Roles

1.Project Manager (SAKTHIVEL T)

Role: Oversee the overall project, manage timelines, and ensure smooth coordination between team members.

1.Define project milestones and ensure deadlines are met.

2.Coordinate team activities and resolve any conflicts or issues.

2. Data Scientist (BHARATHI R)

Role: Lead the data analysis, feature engineering, and model development processes.

1.Collect and preprocess the data (handle missing values, normalize features).

2.Perform exploratory data analysis (EDA) and visualize key patterns.

3.Train and evaluate machine learning models (e.g., Logistic Regression, Random Forest).

3.Software Developer (BHARATHIRAJA A)

Role: Develop and deploy the model within a functional environment.

1.Implement the model in a Google Colab notebook or simple web app.

2.Integrate the trained machine learning model for real-time predictions

3.Handle data input/output and ensure the model can make predictions on new data.

4. Data Analyst (ESAKKIANKEERTHIK S)

Role: Analyze the data, generate insights, and document the entire project.

1.Conduct detailed data analysis and provide actionable insights.

2.Create visualizations to support key findings from the data.

3.Document the project process, model details, and evaluation results.