Project Proposal: Customer Churn Analysis

1. Objective

The primary objective of this project is to analyze the factors contributing to customer churn in a telecommunications company. By identifying key indicators and patterns, we aim to develop strategies to improve customer retention and reduce churn rates.

2. Problem Statement

Customer churn is a significant issue for telecommunications companies, leading to revenue loss and increased customer acquisition costs. Understanding the reasons behind customer churn and identifying at-risk customers can help in implementing targeted retention strategies.

3. Goals

1. **Identify Key Factors**: Determine the main factors influencing customer churn.
2. **Predict Churn**: Develop a predictive model to identify customers at risk of churning.
3. **Provide Insights**: Offer actionable insights to help the company improve customer retention.

4. Deliverables

1. **Data Analysis Report**: A comprehensive report detailing the exploratory data analysis (EDA), including visualizations and summary statistics.
2. **Predictive Model**: A machine learning model capable of predicting customer churn with high accuracy.
3. **Insights and Recommendations**: A set of actionable insights and recommendations based on the analysis and model results.
4. **Presentation**: A presentation summarizing the findings, model performance, and recommendations.

5. Project Scope

- **Data Collection**: Utilize the provided customer churn dataset from Kaggle, which includes customer demographics, service details, account information, and churn status.
- **Data Cleaning and Preprocessing**: Handle missing values, encode categorical variables, and scale numerical features.
- **Exploratory Data Analysis (EDA)**: Perform EDA to understand the distribution of data, identify patterns, and visualize relationships between features.
- **Model Building**: Develop and evaluate multiple machine learning models (e.g., Logistic Regression, Random Forest, Support Vector Classifier) to predict customer churn.
- **Model Evaluation**: Assess model performance using metrics such as accuracy, precision, recall, and F1-score.
- **Insights Generation**: Analyze the model results to identify key factors influencing churn and provide actionable recommendations.

6. Intended Approach

1. **Data Preparation**:
    - Load and inspect the dataset.
    - Clean the data by handling missing values and outliers.
    - Encode categorical variables and scale numerical features.
2. **Exploratory Data Analysis (EDA)**:
    - Generate summary statistics and visualizations.
    - Analyze the distribution of churn and other key features.
    - Create correlation heatmaps to identify relationships between features.
3. **Model Development**:
    - Split the data into training and testing sets.
    - Train multiple machine learning models.
    - Evaluate model performance and select the best model.
4. **Insights and Recommendations**:
    - Interpret the model results to identify key churn indicators.
    - Provide actionable insights and recommendations to reduce churn.
5. **Documentation and Presentation**:

- Compile the analysis, model results, and recommendations into a comprehensive report.
- Prepare a presentation summarizing the project findings and recommendations.

7. Timeline

Phase	**Duration**	**Activities**	**Completion Date**
Topic Selection	1 week	Choose the project topic	12/04/2024
Data Cleaning & Preparation	1 week	Handle missing values, encode variables, scale features	12/09/2024
Exploratory Data Analysis	1 week	Perform EDA, generate visualizations	12/09/2024
Model Development	1 week	Train and evaluate machine learning models	12/11/2024
Deployment	1 day	Deploy the model using Flask and Postman	12/11/2024
Documentation & Presentation	1 week	Compile reports, prepare the presentation	12/11/2024

8. Resources Required

- **Software**: Python, Jupyter Notebook, libraries (pandas, numpy, matplotlib, seaborn, scikit-learn), Flask, Postman
- **Hardware**: Computer with sufficient processing power and memory
- **Data**: Customer churn dataset from Kaggle

9. Team Members and Contributions

- **Devika Dileep (200566996)**: Visualization
- **Sandra Sajimon (200567825)**: Data Cleaning and Preparation
- **John Hanok (200573253)**: Deployment using Flask and Postman

10. Challenges Faced

- **Data Cleaning**: Handling the large dataset and ensuring data quality was challenging.
- **Deployment**: Lack of experience with Flask and Postman led to difficulties during deployment.
- **Time Management**: Coordinating tasks and managing time effectively was a challenge.
- **Group Collaboration**: Ensuring effective communication and collaboration among team members was crucial.

11. Budget

- **Software Licenses**: $0 (using open-source tools)
- **Hardware**: $0 (using existing resources)
- **Personnel**: $10,000 (estimated cost for team members over the project duration)

12. Risk Management

- **Data Quality Issues**: Mitigate by thorough data cleaning and preprocessing.
- **Model Performance**: Ensure robust model evaluation and selection.
- **Timeline Delays**: Regular progress reviews and adjustments to the timeline as needed.