



# Project Proposal

## Telecom Customer Intelligence System

### 1. Introduction

In today's competitive telecom industry, customer retention and revenue optimization are major challenges. Telecom companies generate large volumes of customer data, including demographic details, service usage patterns, and billing information. However, this data is often not fully utilized to extract meaningful insights that can support business decisions.

Machine Learning provides powerful techniques to analyze such data and uncover patterns that help organizations understand customer behavior, predict churn, and forecast revenue. This project aims to build an integrated Machine Learning system that applies clustering, classification, and regression techniques in a unified and scalable manner.

### 2. Problem Statement

Telecom companies face difficulties in identifying different customer segments, predicting customer churn, and estimating customer revenue using traditional analytical methods. There is a need for an automated and reliable Machine Learning system that can perform these tasks efficiently while ensuring consistent data preprocessing and accurate predictions.

### 3. Objectives

The main objectives of this project are:

- To segment telecom customers based on their behavior and service usage
- To predict whether a customer is likely to churn
- To estimate monthly charges to understand customer revenue patterns
- To implement all models using pipeline-based Machine Learning architecture
- To evaluate and compare model performance using appropriate metrics

## 4. Dataset Description

The project uses the **Telco Customer Churn Dataset**, which contains information about telecom customers, including:

- Customer demographics
- Subscription and contract details
- Service usage information
- Billing and payment data

The dataset includes both categorical and numerical features, making it suitable for multiple Machine Learning tasks.

## 5. Methodology

The project follows a structured Machine Learning workflow:

### 5.1 Data Preprocessing

- Handling missing values
- Data type conversion and cleaning
- Feature scaling using `StandardScaler`
- Encoding categorical variables using One-Hot Encoding

### 5.2 Model Development

- **Clustering:** KMeans algorithm for customer segmentation
- **Classification:** Random Forest classifier for churn prediction
- **Regression:** Linear Regression for monthly charge prediction

All models are implemented using **scikit-learn Pipelines** and **ColumnTransformer** to ensure consistency, prevent data leakage, and improve scalability.

### 5.3 Model Evaluation

- Clustering evaluated using Silhouette Score
- Classification evaluated using Accuracy and Classification Report
- Regression evaluated using RMSE and  $R^2$  Score

## 6. Tools and Technologies

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib
- **Techniques:** Machine Learning Pipelines, Feature Engineering, Model Evaluation

## 7. Expected Outcomes

- Identification of meaningful customer segments
- Accurate prediction of customer churn
- Reliable estimation of monthly customer charges
- A clean, reusable, and production-ready Machine Learning system

## 8. Applications and Benefits

- Helps telecom companies reduce customer churn
- Supports targeted marketing and retention strategies
- Improves revenue forecasting and planning
- Enhances data-driven decision-making

## 9. Conclusion

This project proposes an end-to-end Telecom Customer Intelligence System that integrates clustering, classification, and regression techniques using Machine Learning pipelines. The system is designed to be scalable, efficient, and suitable for real-world business applications, making it valuable for both academic and industry use.

