**Telecom Customer Churn Prediction: Case Study Problem Statement**

**Introduction**

The telecom industry is highly competitive, and customer churn is a significant concern. Retaining existing customers is often more cost-effective than acquiring new ones. This case study aims to build a predictive model to identify customers who are likely to churn, allowing the company to take proactive measures to retain them.

**Objective**

The primary objective is to develop a machine learning model that can predict customer churn with high accuracy. The model will use historical data to identify patterns or characteristics of customers who have churned in the past.

**Data**

Two datasets are provided:

1. train.csv: This dataset contains historical data, including whether or not a customer has churned. Features include customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges, and Churn.
2. active\_customers.csv: This dataset contains data for active customers without the 'Churn' label. The objective is to predict the likelihood of these active customers churning in the near future.

**Guidelines**

**Data Exploration**

1. Perform initial data exploration to understand the data types, missing values, and summary statistics.
2. Visualize the data to identify patterns and correlations.

**Data Preprocessing**

1. Handle missing values if any.
2. Convert categorical variables into numerical form.
3. Normalize/Standardize numerical features if necessary.

**Feature Engineering**

1. Create new features that might help improve the model.
2. Select relevant features based on statistical tests.

**Model Building**

1. Split the train.csv data into training and validation sets.
2. Try different algorithms like Logistic Regression, Random Forest, and Gradient Boosting to train the model.
3. Tune hyperparameters for better performance.

**Model Evaluation**

1. Evaluate the model using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.
2. Use cross-validation for more robust model evaluation.

**Prediction**

1. Use the final model to predict the 'Churn' label for the active\_customers.csv dataset.

**Data Dictionary**

1. **customerID**: Unique identifier for each customer
2. **gender**: Gender of the customer (Male/Female)
3. **SeniorCitizen**: Whether the customer is a senior citizen or not (1 for Yes, 0 for No)
4. **Partner**: Whether the customer has a partner (Yes/No)
5. **Dependents**: Whether the customer has dependents (Yes/No)
6. **tenure**: Number of months the customer has been with the company
7. **PhoneService**: Whether the customer has a phone service (Yes/No)
8. **MultipleLines**: Whether the customer has multiple lines (Yes/No/No phone service)
9. **InternetService**: Type of internet service the customer has (DSL, Fiber optic, No)
10. **OnlineSecurity**: Whether the customer has online security feature (Yes/No/No internet service)
11. **OnlineBackup**: Whether the customer has online backup feature (Yes/No/No internet service)
12. **DeviceProtection**: Whether the customer has device protection feature (Yes/No/No internet service)
13. **TechSupport**: Whether the customer has tech support feature (Yes/No/No internet service)
14. **StreamingTV**: Whether the customer has streaming TV feature (Yes/No/No internet service)
15. **StreamingMovies**: Whether the customer has streaming movies feature (Yes/No/No internet service)
16. **Contract**: Type of contract the customer has (Month-to-month, One year, Two year)
17. **PaperlessBilling**: Whether the customer has opted for paperless billing (Yes/No)
18. **PaymentMethod**: Method of payment (Electronic check, Mailed check, Bank transfer, Credit card)
19. **MonthlyCharges**: Monthly charges for the customer
20. **TotalCharges**: Total charges for the customer till date
21. **Churn**: Whether the customer churned or not (Yes is 1/No is 0)