SyriaTel Customer Churn Analysis Report

1. Business Understanding

Customer occurs when clients discontinue their relationship with a service provider. In the telecommunications industry, churn significantly impacts profitability since acquiring new customers is often more expensive than retaining existing ones. For SyriaTel, reducing churn is crucial to sustaining revenues. This project aims to predict potential churners, by leveraging machine learning algorithms allowing telecom companies to implement effective retention strategies.

1.1 Objectives

The project is guided by the following objectives:

Main Objective:

Develop a machine learning model capable of predicting customer churn with high reliability.

Specific Objectives:

- Analyze the impact of customer demographics on churn.
- Examine the effect of account details such as account length on churn behavior.
- Investigate the influence of subscription plans on churn.
- Assess how frequency of customer service calls correlates with likelihood of churn.
- Create a predictive system that flags high-risk customers early to enable targeted interventions.

1.2 Success Criteria

The success of this project is measured through both technical and business outcomes. From a technical perspective, the model should achieve strong performance across accuracy, precision, recall, F1-score, and ROC-AUC. From a business perspective, success means providing actionable insights that SyriaTel can implement to reduce churn, increase customer loyalty and optimize retention campaigns.

2.0 Data Understanding

The dataset used is from Kaggle and consists of 3,333 customer records with 21 attributes. It contains a mix of numeric, categorical, and boolean features, and importantly, no missing values or duplicates.

Key characteristics of the dataset include:

• Target Variable: Churn (binary: yes/no).

• Demographics: state, area_code.

• Account Details: account_length.

• Subscription Plans: international_plan, voice_mail_plan.

• Service Calls: customer_service_calls.

3.0 Data Preparation

Data preparation was a crucial step to ensure data quality and readiness for modeling. Key tasks included: removing irrelevant columns, standardizing categorical values, encoding categorical variables, and validating consistency of data types. Additionally, the target class imbalance was addressed using SMOTE to improve model learning.

Exploratory Data Analysis

Univariate and bivariate analyses were conducted to understand data distributions and relationships. Key findings from EDA include:

• Churn Rate: Approximately 14.5% of customers left the service, confirming significant imbalance.

• Subscription Plans: Most customers did not subscribe to international or voicemail plans, and those who did showed higher churn tendencies.

• Area Codes & States: Customers were evenly distributed across most area codes except for a concentration in area code 415. WV had the highest state-level customer count.

• Usage M: Summary statistics revealed typical telecom usage patterns, though higher day and evening charges appeared correlated with churn.

• Customer Service Calls: High frequency of service calls was strongly associated with churn, suggesting dissatisfaction or unresolved issues.

4. Modelling and Evaluation

Several supervised learning algorithms were tested, including Logistic Regression, Decision Tree, Random Forest and K-Nearest Neighbors. Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices. Random Forest consistently outperformed the other models, offering a better balance between predictive power and interpretability.

Feature importance analysis revealed that the most influential predictors of churn were: number of customer service calls, international plan subscription and total day charges. These insights highlight behavioral and service-related drivers of churn.

5. Recommendations

Based on the analysis, the following strategic recommendations are proposed:

• Implement Random Forest to flag high risk customers based on churn predictors.

• Focus retention efforts on international plan subscribers and customers making frequent service calls.

• Investigate and resolve pain points in customer service processes to improve satisfaction.

• Offer personalized retention packages or incentives for customers with short account histories.

• Continuously monitor churn trends and retrain predictive models to maintain performance.

6. Conclusion

The project demonstrated the effectiveness of using machine learning to predict customer churn with high reliability. Random Forest emerged as the best performing models. Business insights derived from the analysis highlighted the role of service quality and subscription choices in churn behavior. By implementing the recommendations, SyriaTel can proactively reduce churn, retain valuable customers, and strengthen market competitiveness.