Project Proposal: Predicting Customer Churn in the Telecom Industry Using Data Mining

Problem Statement

One of the most important issues facing telecom firms is loss of customers. Although keeping current customers is more economical than finding new ones, many telecom companies find it difficult to spot clients who are about to depart. Excessive turnover rates hinder company growth and cause substantial revenue losses.

With the use of data mining tools, this initiative seeks to create a data-driven strategy for forecasting client retention. Our aim is to develop predictive models that will assist telecom carriers in proactively addressing and lowering customer attrition by examining customer demographics, service usage trends, and billing data from the Telco Customer Churn dataset.

Motivation

The telecom industry is very competitive, and even a small drop in customer churn can save companies a lot of money. That's why it's important not only to predict which customers are likely to leave, but also to understand the reasons behind their decision.

Our project is motivated by two main goals:

1. Business Value: Helping telecom companies keep their customers by giving them insights that can be used for special offers, better services, or loyalty programs.

2. Learning Value: Giving us the chance to apply data mining techniques to a real-world problem, including classification, analyzing important features, and comparing different models.

By working with the Telco Customer Churn dataset, we hope to show how data mining can be used both to support business decisions and to help us practice and understand important concepts from the course.

Data Type and Collection

We will use the Telco Customer Churn dataset available on Kaggle. The relational dataset consists of numerical and categorical data. The following is an explicit description of the dataset.

Dataset Description:

- Rows: 7,043 customers
- Features: 21 attributes
- Target Variable: Churn (Yes/No)

Feature Groups:
- Customer demographics: gender, SeniorCitizen, Partner, Dependents
- Services subscribed: PhoneService, InternetService, OnlineSecurity, etc.
- Contract & billing info: Contract, PaymentMethod, PaperlessBilling
- Tenure & charges: tenure, MonthlyCharges, TotalCharges

Data Mining Tasks

We plan to apply the following data mining techniques:

1. Preprocessing:
- Handle missing values (e.g., in TotalCharges).
- Encode categorical variables (Yes/No, Contract types, etc.).
- Normalize numerical features (tenure, MonthlyCharges, TotalCharges).

2. Classification Models:

We may apply a few of the following methods to extract data that will then be further used to predict and classify/identify churn. What methods we use will highly depend on how much time we have to thoroughly process the information in different ways through the course of this project.
- Potential Baseline Models and Algorithms: Logistic Regression, Decision Tree.
- Potential Advanced Models: Random Forest, Gradient Boosting (XGBoost), Support Vector Machine.
- (Optional) Neural Networks for comparison.

3. Evaluation:
- Effectiveness: Accuracy, Precision, Recall, F1-score, ROC-AUC.
- Efficiency: Training time and computational cost.

4. Interpretability:
- Feature importance analysis to identify factors influencing churn (e.g., contract type, tenure, monthly charges).

Expected Outcomes & Impact

- Predictive Insight: Accurate churn prediction model that identifies customers at risk.
- Business Insight: Understanding key drivers of churn such as short tenure, high monthly charges, and month-to-month contracts.
- Practical Impact: Telecom providers can design proactive retention strategies (e.g., discounts, loyalty rewards, improved services) targeted at high-risk customers.

Schedule

- Week 1: Data exploration, cleaning, preprocessing.
- Week 2–3: Build baseline models (Logistic Regression, Decision Tree).
- Week 4–5: Build advanced models if time allows. (Random Forest, XGBoost, SVM, Neural nets).
- Week 6: Model comparison, feature analysis and evaluations.
- Week 7: Report writing, visualization, and presentation.

Members:
Tajwar-Ul Hoque
Luis Escamilla
Isaiah Malcom

NOTE: Please reach out to the team if you have any further question.