Project Proposal

Predicting Customer churn for a Telecommunication Company

**Presented by**: Arivarasan Ramasamy

Business Objective:

One of the most important KPIs for business intelligence is churn. To increase profitability, businesses must constantly evaluate their churn rate and treat it as of great importance.

On average, telecommunication companies experience a churn of 22% over the course of each quarter. This is a concerningly high number. Every business tries to expand its customer base but, retaining its existing customer is equally significant. After all, as a telecommunication company, we are reliant on subscriptions for income.

US telecom customers are more willing to switch

Customers were less likely to change providers during the early months of the pandemic, but since then, the percentage of those likely to switch has steadily risen

Percentage of customers who say they are likely to switch carriers
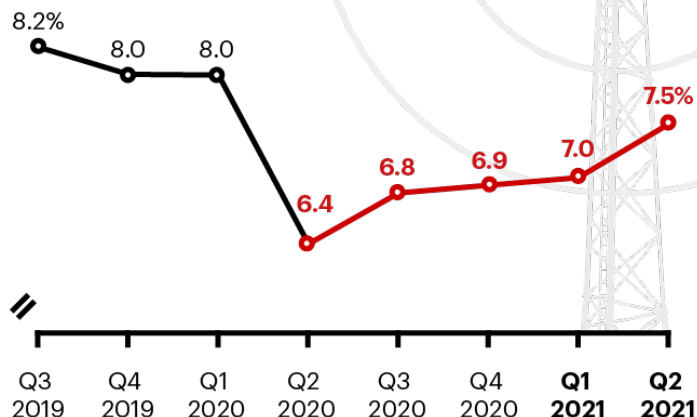
Notes: Aggregated results for Verizon, AT&T, T-Mobile/Sprint
Source: NPS Prism™ Survey Q3 2019 to Q2 2021

BAIN & COMPANY

Customers churn is an everyday thing with telecom companies, although the percentage rate increased slightly following the pandemic.

To address this business problem, we are trying to create a 'machine learning model that can predict customer behavior'. We are planning to train multiple ML models with customer data from IBM Business Analytics Community. These data sets have all been tested with IBM Cognos Analytics.

By comparing the best metrics from the model and coming up with a solution and recommendations for the Telecom companies to handle this issue.

Some of the use cases for the best-predicted model can be used to:

1. Analyze who might discontinue our service and proactively reach out to them to identify and resolve any issues they might face. This way we can also showcase to our customers that we care about their needs.
2. Speculate on the reasons why consumers are departing by using the churned customer data. To implement preventative actions, utilize data from an analysis of how and when churn happens during a customer's relationship with the business.
3. Identifying who our best and most loyal customers are. Instead of just rewarding customers who might be considering discontinuing our service by providing additional benefits, we would like to reward our most loyal customers. This will incentivize customers to stay. This will also give us a better idea of which demographics of customers need to be targeted to gain new customers.

Source:

This data set contains information about a Telecommunication company that provided home phone and internet services to 7043 customers in San Deigo, California in Q3. This data was used as an analytics simulation by IBM to better understand what are the factors that might lead to a customer switching or discontinuing a service. It indicates which customers have left, stayed, or signed up for their service. This dataset can be downloaded from [here](#).

Data Description:

It contains the following information:

- Information about customer's demographic - Gender, Age, Marital/Relationship status, and if they have dependents.
- Services opted for by the customer - Phone, Multiple Lines, Internet, Online Security, Online Backup, Device Protection, Tech Support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.
- Customers who left – "Churn". (Target variable)

Our target variable is Churn and we are using different classification models to predict the Churn of a Telecom customer.

The course of Action:

- By implementing various machine learning algorithms, we hope to come up with a model which accurately predicts if a customer will discontinue a telecommunication service.
- Following that, we will attempt to identify each demographic and customer type and then expand on who our best and worst customers are.

- Finally, we can come up with optimal marketing as well as advertising strategies to target the leaving customers.

Metrics:

In this context,

True Positives – Number of customers that the model correctly predicted will churn. ***

False Positives - Number of customers the model predicted will churn but they didn't. *

True Negatives - Number of customers whom the model correctly predicted will not churn*

False Negatives - Number of customers that the model predicted will not churn but they did. ***

For identifying people who will churn:

The ideal case would be for the total customer count will be split into True Positive and False Negative

The cost of True Positives and False Negatives is very important in this business problem because any miscalculation in these values will result in a customer loss.

	Predicted No	Predicted Yes
Actual No	TN	FP
Actual Yes	FN	TP

Recall:

$$TP / (TP+FN)$$

- Ability of a classification model to identify all relevant instances
- Also referred to as Sensitivity, Probability of Detection, True Positive Rate
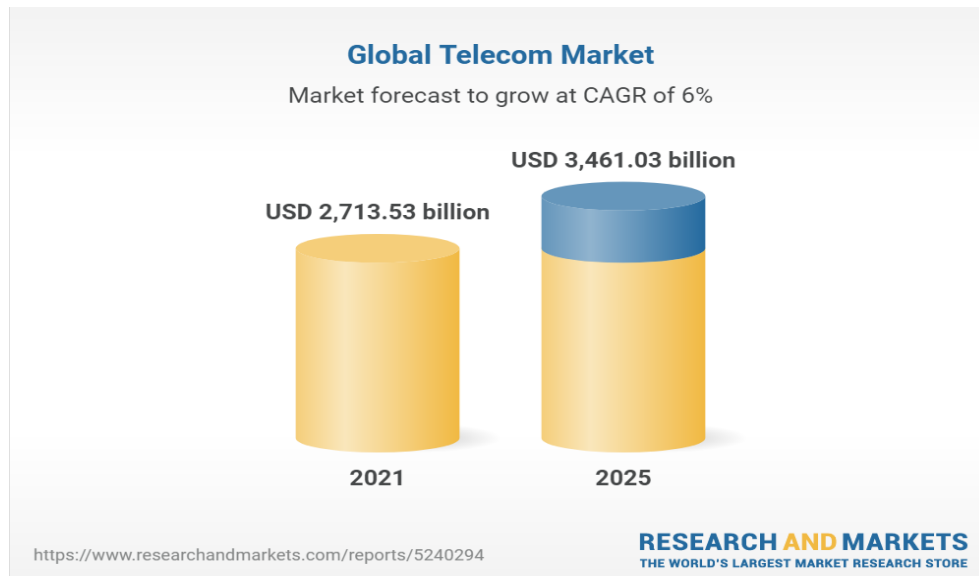
We will try to calculate the best Recall score for this problem with different models.

The cost of False Positive and True Negative will have less impact as we retain customers who will not leave, any offers or incentives to them will act as a better strategy to retain the customer.

We will also try to get a better model for the recall score using hyperparameter tuning and measuring the best threshold value for the model using ROC AUC and PR AUC plots.

We will set the F-beta value of f2 which Favor Recall, but also consider Precision.

Impact of the problem and the solution:

Global Telecom Market

Market forecast to grow at CAGR of 6%

USD 3,461.03 billion

USD 2,713.53 billion

2021   2025

https://www.researchandmarkets.com/reports/5240294

RESEARCH AND MARKETS
THE WORLD'S LARGEST MARKET RESEARCH STORE

Because the telecom industry is multibillion-dollar, even addressing a minor issue and improving the whole process by a small proportion will have a significant impact on the amount of revenue generated. As a result, this problem is worth resolving numerous times if we enhance our performance even slightly.