TELECOM CHURN CASE STUDY

Name : Adhya A. Kamat

Data Science

OVERVIEW OF THE BUSINESS CHALLENGE

In the telecom sector, customers have the option to select from a variety of service providers and actively switch between them. Within this intensely competitive market, the telecommunications industry witnesses an annual churn rate averaging between 15-25%. It's crucial to recognize that acquiring a new customer typically costs 5-10 times more than retaining an existing one. Consequently, prioritizing customer retention has become paramount in this industry.

For numerous established operators, the primary business objective is preserving high-profit customers.

To combat customer churn, telecommunications companies must proactively identify those customers who are most likely to churn.

This project involves the analysis of customer-level data from a leading telecom company, the construction of predictive models aimed at pinpointing customers at a heightened risk of churning, and the identification of the primary churn indicators.

Understanding and Defining Churn in the Telecom Industry:

In the telecom sector, two primary payment models exist: postpaid (where customers settle bills monthly/annually after service usage) and prepaid (where customers prepay or recharge in advance and then utilize services).

Within the postpaid model, when customers wish to switch to another provider, they typically inform their current operator, making it straightforward to identify these instances as churn.

Conversely, within the prepaid model, customers who want to switch to a different network can simply cease using the services without prior notice. Determining whether someone has genuinely churned or is temporarily not using the services (e.g., due to international travel for a month or two with the intention to resume service later) becomes challenging.

As such, predicting churn is notably more crucial (and intricate) for prepaid customers, necessitating a meticulous definition of the term 'churn.' Additionally, it's worth noting that the prepaid model predominates in India and Southeast Asia, whereas the postpaid model is more prevalent in Europe and North America.

This project specifically focuses on the Indian and Southeast Asian telecom markets.

Churn can be defined in various ways, including:

Revenue-Based Churn: This refers to customers who have not utilized any revenue-generating services such as mobile internet, outgoing calls, or SMS over a specific timeframe. Alternatively, it can involve aggregate metrics like identifying customers who have generated less than a specified amount, such as INR 4 per month, in total, average, or median revenue. However, it's important to note that this definition may have limitations, as it may categorize customers who only receive calls or SMS from others, especially in rural areas, where they don't generate revenue themselves.

Usage-Based Churn: This definition identifies customers who have not engaged in any usage, whether incoming or outgoing, such as calls or internet activity, during a particular period. One potential drawback of this definition is that by the time a customer stops using services for a while, it may be too late to take corrective actions to retain them. For example, if churn is defined as "two months of zero usage," predicting churn may not be effective since the customer might have already switched to another provider.

In this project, the definition of churn is based on the usage-based definition.

High-Value Churn:

In the Indian and Southeast Asian markets, a significant portion, approximately 80%, of revenue is generated by the top 20% of customers known as "high-value customers." Reducing churn among these high-value customers is critical to preventing substantial revenue loss.

In this project, high-value customers will be defined based on specific metrics (which will be detailed later), and the focus will be on predicting churn exclusively among these high-value customers.

Understanding the Business Objective and Data:

The dataset encompasses customer-level information spanning four consecutive months, represented as June (6), July (7), August (8), and September (9). The primary business objective is to forecast churn during the last month, which is September (9), utilizing data (features) gathered from the initial three months. To effectively achieve this goal, it's crucial to comprehend typical customer behavior during the churn process.

Understanding Customer Behavior During Churn:

Churn decisions by customers usually evolve over a period rather than happening instantly, especially in the case of high-value customers. In churn prediction, we assume that there are three distinct phases within the customer lifecycle:

The 'Good' Phase: During this phase, customers are content with the service and exhibit regular behavior.

The 'Action' Phase: Customer satisfaction starts to decline in this phase. Factors like receiving enticing offers from competitors, encountering unjust charges, or experiencing a decrease in service quality may trigger this phase. Customers tend to display different behavior compared to the 'good' months during this stage. Identifying high-churn-risk customers during this phase is critical because corrective actions, such as matching competitor offers or improving service quality, can be taken at this point.

The 'Churn' Phase: This phase signifies that the customer has officially churned, and we define churn based on this phase. Importantly, during the time of prediction (i.e., the 'action' months), data pertaining to this phase is not available for prediction purposes. Therefore, after tagging churn as 1 (churned) or 0 (not churned) based on this phase, all data corresponding to this phase is discarded.

In this specific case, given the four-month window, the first two months represent the 'good' phase, the third month represents the 'action' phase, and the fourth month signifies the 'churn' phase.

Let's take a closer look at our model's performance:

Training Set Metrics
Accuracy: 0.84
Sensitivity: 0.81
Specificity: 0.83

Test Set Metrics
Accuracy: 0.78
Sensitivity: 0.82
Specificity: 0.78

In examining the model's performance on the test set, we can conclude that it has effectively applied what it learned during training.
Now, when we consider the model without Principal Component Analysis (PCA), we find:The logistic regression model, without PCA, exhibits commendable sensitivity and accuracy. These performance metrics are on par with the models utilizing PCA. Therefore, we can opt for the simpler logistic regression model without PCA. This choice offers the advantage of explaining the significance of each predictor variable, aiding in the identification of crucial factors for decision-making regarding customer churn. Consequently, this model provides valuable insights for our business operations.In summary, the model without PCA is a relevant and effective tool for explaining and addressing customer churn in our business.

Inferences

Target customers whose usage of local incoming calls and outgoing ISD calls decreases during the action phase (typically in August).

Focus on customers whose outgoing charges for other services in July and incoming charges for other services in August are lower.

It's notable that most of the top variables have negative coefficients, indicating an inverse correlation with the probability of churn.

Customers experiencing an increase in value-based costs during the action phase are more likely to churn, making them a suitable target for offers or incentives.

Customers with a higher 3G recharge in August are more likely to churn.

Customers showing a decrease in STD incoming minutes of usage for calls from operators T to fixed lines in August are at a higher risk of churning.

Customers reducing their monthly 2G usage in August are also more probable to churn.

Pay attention to customers who reduce their incoming minutes of usage for calls from operators T to fixed lines in August.