**Case Study Telecom Churn\_Sanjeev**

**Problem Statement**

**Business Problem**

Customers in the telecom sector have access to a variety of service providers and can actively switch from one operator to another. The telecoms business has an average annual churn rate of 15 to 25 percent in this fiercely competitive market. Customer retention has now surpassed customer acquisition in importance due to the fact that it is 5–10 times more expensive to gain new customers than to keep existing ones.

Retaining highly profitable consumers is the top business objective for many established operators.

Telecom businesses must identify the consumers who are most likely to leave in order to reduce customer turnover.

In this study, we will examine customer-level data from a top telecom company, create prediction models to find customers who are most likely to leave, and pinpoint the key signs of churn.

**Definition and Analysis of Churn**

In the telecom sector, there are two primary payment options: postpaid (where consumers pay a monthly or annual fee after using the services) and prepaid (when users pay/recharge with a set amount in advance and then use the services).

In the postpaid, customers who want to switch operators often contact the current operator to cancel the services, which allows us to recognise this as a churn situation right away.

But, in the prepaid model, customers who want to switch to another network can just stop using the services without giving any prior notice, and it can be difficult to tell whether someone has actually churned or is merely not using them at the moment (for example, someone may be travelling abroad for a few months with the intention of returning to using the services later).Thus, churn prediction is usually more critical (and non-trivial) for prepaid customers, and the term ‘churn’ should be defined carefully.

Also, prepaid is the most common model in India and southeast Asia, while postpaid is more common in Europe in North America.

This study is based on the Indian and Southeast Asian market.

**Definitions of Churn**

Customers who have not used any revenue-generating services during a certain length of time, such as mobile internet, outbound calls, SMS, etc. The use of aggregate metrics is another option. For example, "customers who have generated less than INR 4 per month in total/average/median revenue."

The biggest flaw in this definition is that some consumers just use the services to receive calls or SMSs from their wage-earning counterparts; in other words, they don't produce any income but nevertheless use the services. For instance, many people in rural locations only get calls from their siblings who work in cities.

Customers that have not made any calls, used the internet, or made any other outgoing or incoming usage over a period of time (usage-based churn).

This definition may have the drawback that it may be too late to take corrective measures to retain the client once they have stopped utilising the services for a period. For instance, if we use a "two-months zero usage" definition of churn, projecting churn may be meaningless because the client will have already gone to a different operator by that point.

In this project, churn will be defined using the usage-based definition.

**High-value Churn**

Approximately 80% of revenue in the Indian and Southeast Asian markets comes from the top 20% of clients (referred to as high-value consumers). Therefore, if we can lower the high-value client churn, we can lower large revenue leakage.

In this project, we will exclusively forecast churn on high-value customers and identify high-value customers based on a certain indicator (discussed below).

**Understanding the Business Objective and the Data**

The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

**Customer Behaviour During Churn**

clients typically take some time to decide whether to switch to a rival business (this is especially true for high-value clients). In churn prediction, we presum that the client lifetime comprises three stages:

1. **The ‘good’ phase:** In this stage, the customer is happy with the service and behaves as usual.
2. **The ‘action’ phase:** In this stage, the customer's experience starts to suffer; for instance, he or she receives an alluring offer from a rival, pays unfair fees, is dissatisfied with the level of service, etc. The client typically behaves differently during this time than during the "good" months. Additionally, as some remedial actions (such matching the competitor's offer/improving the service quality etc.) can be implemented at this stage, it is vital to identify high-churn-risk clients during this period.
3. **The ‘churn’ phase:** The client is said to have churned during this stage. This phase is the basis for how we define churn. It's also crucial to remember that we cannot foresee using this data at the moment (the action months) because it is not yet available. In light of this, we remove all data associated with this phase after marking churn as 1/0 based on this phase.

The first two months in this scenario are the "good" phase, the third month is the "action" phase, and the fourth month is the "churn" period because we are working over a four-month window.

**Data Prep**

The following data preparation steps are important for this problem:

1. Create new features The ability to distinguish between excellent and bad models using good features makes this one of the most crucial steps in the data preparation process. We will leverage our knowledge of business to derive characteristics that could serve as key churn indicators.
2. Eliminate high-value clients We simply need to forecast turnover for high-value customers, as was already mentioned. As an example of a high-value client, consider: Those who have recharged with a sum more than or equal to X, where X represents the 70th percentile of the typical recharge sum over the first two months (the favourable phase).
3. Tag churners and take away churn phase traits. Now, depending on the fourth month, tag the customers who churned (churn=1, otherwise 0) as follows: In the churn phase, those who have not placed any calls (incoming or outgoing) OR accessed mobile internet even once. We need to identify churners using the following attributes:

* total\_ic\_mou\_9
* total\_og\_mou\_9
* vol\_2g\_mb\_9
* vol\_3g\_mb\_9

All the attributes pertaining to the churn phase (all attributes with "\_9," "\_10," etc.) must be removed after labelling churners.

**Modelling**

Create models to foresee churn. The predictive model we'll create will have two functions:

1. It will be used to forecast the churn phase, or whether a high-value customer would leave in the near future. Knowing this allows the business to take appropriate action, such as offering customised programmes or discounts on recharges.
2. It will be utilised to pinpoint crucial elements that are reliable churn predictors. These factors might also reveal the factors influencing customers' decisions to transfer networks.

In some circumstances, a single machine learning model can accomplish both of the aforementioned objectives. But because there are so many features in this case, we should try a dimensionality reduction method like PCA before creating a predictive model. We can apply any classification model after PCA. We will try to use ways to tackle class imbalance because the rate of churn is normally modest (between 5 and 10%; this is known as class-imbalance).

To create the model, we might consider the following steps:

1. Perform data preprocessing (convert columns to the proper formats, deal with missing values, etc.).
2. Perform appropriate exploratory analysis to glean valuable insights (whether immediately applicable to business or ultimately applicable for modeling/feature engineering).
3. Derive new features.
4. Use PCA to lessen the number of variables.
5. Train several models, fine-tune model hyperparameters, etc. (adequate procedures should be used to handle class imbalance).
6. Use the right evaluation measures to assess the models. Remember that effectively identifying churners is more crucial than precisely identifying non-churners; use an evaluation metric that matches this business objective.
7. Finally, pick a model based on an assessment metric.

Only one of the two objectives will be accomplished by the aforementioned model—predicting clients who will leave. The aforementioned methodology cannot be used to pinpoint the critical churn factors. Because PCA typically produces components that are difficult to read, this is the case.

As a result, we will create a new model with the primary goal of identifying key predictor features that aid in the understanding of churn indicators by the business. A model from the tree family or one based on logistic regression is a solid option for determining important variables. When using logistic regression, we shall take multi-collinearity into account.

Use plots, summary tables, or other visual representations to visually represent the most significant predictors after finding them.

Finally, based on our observations, make suggestions for measures to control customer attrition.