
upGrad
Raho Ambitious

#LifeKoKaroLift

Telecom Churn Case Study – Pre Assignment Session

Image created by Generative AI

What's for Today?

1. Introduction to the Problem Statement
2. About the Data
3. Modeling Approach (Guidance)
4. Key Goals of the Case Study
5. Final Submissions
6. Guidelines on the Evaluation Metrics

Introduction to the Problem Statement

Overview

upGrad

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another.

In this highly competitive market, the telecommunications industry experiences an average of **15-25%** annual churn rate.

Given the fact that it costs **5-10 times more to acquire a new customer** than to retain an existing one, **customer retention** has now become even more important than customer acquisition.

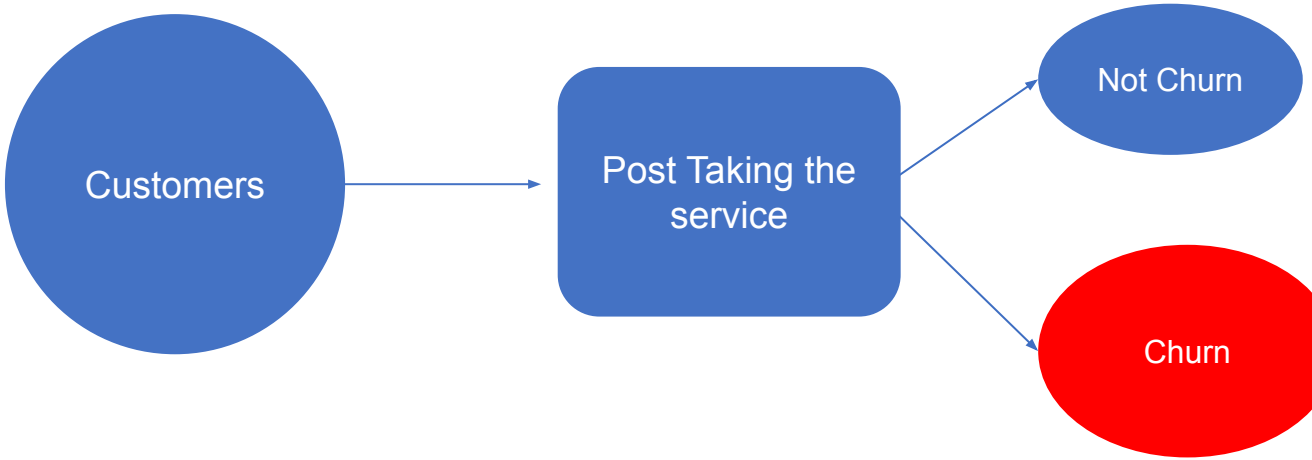
Customer Retention > Customer Acquisition

Objective

upGrad

Predicting Customer Churn

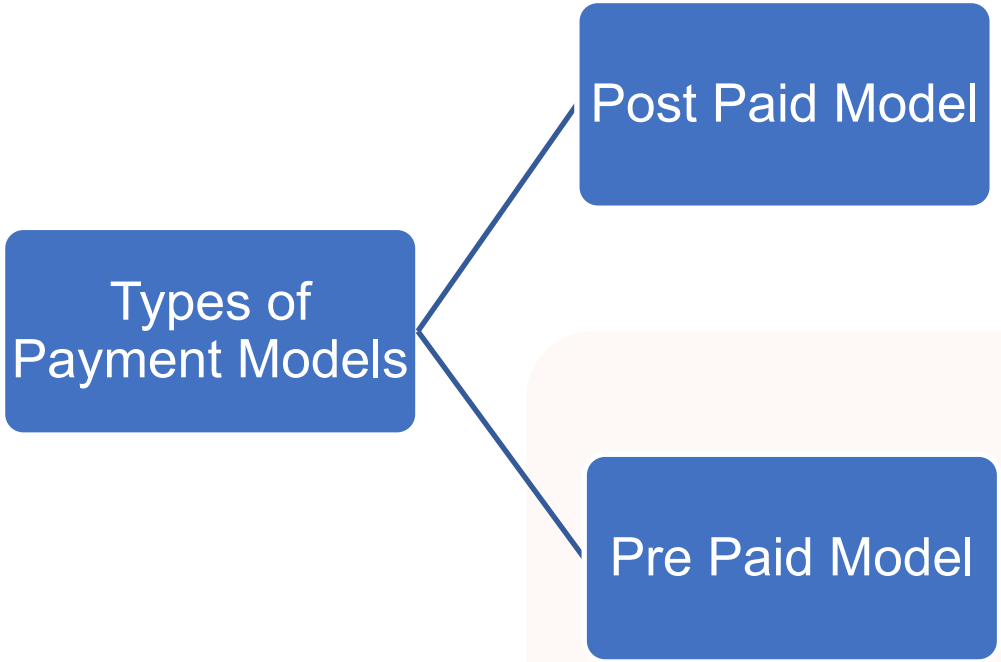
To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**



Understand what factors contribute to Customer's Churn

Types of Payment Model

upGrad



In the postpaid model, when customers want to switch to another operator, they usually inform the existing operator to terminate the services, and you directly know that this is an instance of churn.

More Critical

In the prepaid model, customers who want to switch to another network can simply stop using the services without any notice, **and it is hard to know whether someone has actually churned or is simply not using the services temporarily**

Definitions of Churn

upGrad

Churn Types	Definition
Revenue-based churn	Customers who have not utilised any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as ‘customers who have generated less than INR 4 per month in total/average/median revenue’.
Usage-based churn	Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.

A potential shortcoming of this definition is that when the customer has stopped using the services for a while, it may be too late to take any corrective actions to retain them. For e.g., if you define churn based on a ‘two-months zero usage’ period, predicting churn could be useless since by that time the customer would have already switched to another operator.

High Value Churn

upGrad

In the Indian and the Southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers). Thus, if we can reduce churn of the high-value customers, we will be able to reduce significant revenue leakage.

In this project, you will define high-value customers based on a certain metric (mentioned later below) and predict churn only on high-value customers.

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.

Understanding customer behaviour during churn

upGrad

In this phase, the customer is happy with the service and behaves as usual.

Good Phase

The customer experience starts to sore in this phase, for e.g. he/she gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behaviour than the 'good' months.

Action Phase

the customer is said to have churned. You **define churn based on this phase**. Also, it is important to note that at the time of prediction (i.e. the action months), this data is not available to you for prediction. Thus, after tagging churn as 1/0 based on this phase, you discard all data corresponding to this phase.

Churn Phase

About the Data

Telecom Churn Case Study

upGrad

Data Overview

Some common attributes:

- loc (local),
- IC (incoming),
- OG (outgoing),
- T2T (telecom operator to telecom operator),
- T2O (telecom operator to another operator),
- RECH (recharge) etc.

The attributes containing 6, 7, 8, 9 as suffixes imply that those correspond to the months 6, 7, 8, 9 respectively.

Dependent Variable	Churn or Not Churn (to be created)
Total features	226
Total Rows	99999
Granularity	Customer Level

Data Strategies

upGrad

1. **Derive new features**
This is one of the most important parts of data preparation since good features are often the differentiators between good and bad models. Use your business understanding to derive features you think could be important indicators of churn.

2. **Filter high-value customers**
As mentioned above, you need to predict churn only for the high-value customers. Define high-value customers as follows: Those who have recharged with an amount more than or equal to X, where X is the **70th percentile** of the average recharge amount in the first two months (the good phase).

After filtering the high-value customers, you should get about 29.9k rows.

3. **Tag churners and remove attributes of the churn phase**
Now tag the churned customers (churn=1, else 0) based on the fourth month as follows: Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase. The attributes you need to use to tag churners are:

- total_ic_mou_9
- total_og_mou_9
- vol_2g_mb_9
- vol_3g_mb_9

After tagging churners, **remove all the attributes corresponding to the churn phase** (all attributes having ' _9', etc. in their names).

Modeling Approach (Guidance)

Modeling Gudinance Case Study

upGrad

Modeling Guidelines (Indicative not exhaustive) (1/2)

Modeling Steps	Remarks
1. Data Sanity	• Check the head/tail, shape, info, column names, describe, initial no. of missing value
2. Data Cleaning & Preparation	• Drop columns where missing values is greater than 30-40% • Drop any features which are not relevant for the analysis like IDs, locations etc • Drop columns which have high proportions of 'Select' since they are same as null • Create dummy features where ever applicable for the categorical features • Also, check the variance rule of the dummy features, to categorize the smaller ones into 1 category
3. Create feature & dependent data set	• Create the X & Y data frame containing the features & target respectively
4. Create the train & test data	• 80/20 or 70/30 recommended

Telecom Churn Case Study

upGrad

Modeling Guidelines (Indicative not exhaustive) (2/2)

Modeling Steps	Remarks
1. EDA & Correlations	• Visualize the distributions & relationship among the variables using correlations
2. Feature Selection	• Can use PCA to select important features to iterate in the model
4. Model Building	• Use Logistic Regression/Tree Based to iterate & select the best features
5. Model Evaluation	• P-values, VIF, Accuracy score, F1, Recall, Precision, ROC, AUC, Sensitivity, Specificity
6. Model Prediction	• Prediction of Telecom Churn in train, test, recheck Model Accuracy in test • Also coefficients / factors contributing to the churn
7. Model Summary	• Final Important variables, accuracy summary

Submission Guidelines

Telecom Churn Case Study

upGrad

After identifying important predictors, display them visually - you can use plots, summary tables etc. - whatever you think best conveys the importance of features.

Finally, **recommend strategies to manage customer churn** based on your observations.

Note: Everything has to be submitted in one Jupyter notebook.

Evaluation Criteria

Lead Scoring Case Study

upGrad

Stage	Meets expectations	Does not meet expectations
Data understanding, preparation, and feature engineering (35%)	All important data quality checks are performed and inconsistent/missing data is handled appropriately. Relevant EDA is done using plots and summaries. The insights from EDA are clearly derived and explained. Filtering high-value customers and tagging churned customers is done correctly. Feature engineering is conducted rigorously and correctly. An appropriate set of features is used to build the model.	Data quality checks are not performed/missing data is not handled correctly. Exploratory analysis is not conducted/useful observations are either not extracted or mentioned clearly. Filtering high-value customers or tagging is done incorrectly. Feature engineering is not conducted or is conducted on an inappropriate set of features.
Modelling (churn prediction) (35%)	Dimensionality reduction is conducted correctly, including data preparation required for it. Class imbalance is handled using at least one of the techniques. Model hyperparameters are tuned using correct principles and the approach is explained clearly. A reasonable number and variety of different models are attempted and the best one is chosen based on key performance metrics. Model evaluation is conducted using an appropriate metric. Model evaluation results are at par with the best possible models on this data set.	Dimensionality reduction is not conducted correctly/data is not preprocessed. Class imbalance is not handled. Model hyperparameters are not tuned correctly or the approach is not explained clearly. Few models are experimented with resulting in suboptimal results. Model evaluation is not conducted using an appropriate metric. The results are suboptimal compared to what is possible on this dataset.
Identifying important churn indicators and business recommendation (20%)	Important churn indicators are identified correctly. Clear actionable recommendations are provided based on supporting evidence.	Important indicators are not identified correctly. Recommendations are unclear, unactionable or not backed with supporting evidence.

Thank You

upGrad

Questions/ Doubts?