



# Telecom Churn Case Study

SUBMITTED BY:

PRACHI GARG

# Problem Statement

## *Business problem overview*

- ▶ 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.
- ▶ To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn**. In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn

# *High-value churn*

- ▶ 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, we will define high-value customers based on a certain metric (mentioned later below) and predict churn only on high-value customers.

# *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.
- ▶ In this case, since you are working over a four-month window, the first two months are the 'good' phase, the third month is the action phase, while the fourth month is the churn phase.

# Summarizing findings in the table below

Modeling and Analysis done:

- Principle component Analysis and Regression
- Logistic Regression with RFE and VIF
- Decision Tree
- ADA Boosting with Decision Tree
- Random Forest

	Test- Precision	Test- Recall
<i>PCA with regression</i>	40.6	71.3
<i>Logistic Regression</i>	40.7	89.6
<i>Decision Tree</i>	72.1	48.6
<i>ADA Boosting with DT</i>	69.1	52.3
<i>Random Forests</i>	73.5	50.0

- We see that almost on all models the values are coming very similar to each other and more often than not there is a trade off between precision and recall
- Since both the metrics are important, we feel that going ahead with Random forests

# Recommendation

The following 5 factors affect the churn rate considerably -

	Churn Customers Average	Non Churn Customers Average
<i>Incoming for July</i>	182.91	306.84
<i>Incoming for August</i>	63.44	299.1
<i>2G Sachet for July</i>	0.27	1.12
<i>2G Monthly for August</i>	0.02	0.17
<i>Roaming Total Usage</i>	70.88	28.96

# Recommendations

- ▶ Standard Outgoing Calls and Revenue Per Customer are strong indicators of Churn.
- ▶ Local Incoming and Outgoing Calls for 8th month and average revenue in 8th month are important columns to predict Churn.
- ▶ Customers with tenure less than 4 yrs are more likely to churn.
- ▶ Max Recharge Amount is a strong feature to predict churn.
- ▶ Random Forest produced the best prediction results followed by SVM.
- ▶ Telecom company needs to pay attention to the roaming rates. They need to provide good offers to the customers who are using services from a roaming zone.
- ▶ The company needs to focus on the STD and ISD rates. Perhaps, the rates are too high.
- ▶ To look into both of the issues stated above, it is desired that the telecom company collects customer query and complaint data and work on their services according to the needs of customers.