Part C: Final Project Proposal

South Asian Telecom Operator (SATO) data set is a real life data collected from a major wireless telecom operator in South Asia. Most of the attributes in the data sets are associated with call detail records (CDR), billing and personal information. It contains 2000 subscribers. All of these subscribers were not contract based and had a monthly based subscription. The subscriber data was extracted from the time interval of two months i.e. August and September 2015. The data set contains information of 2000 customers.

The business question for the data scientist to predict if a customer will churn (leave the network) or not. The data set contain both customers that have churned and active customers. The dependent column is class in the data set. I will be analyzing the data using supervised methods, i.e.by applying the k nearest neigbor (knn), Naïve Bayes, and decision tree.

The following fields were provided with the data:

network\_age – The time passed since the subscriber started using the services of the carrier.

Aggregate\_Total\_Rev – this is the total aggregated revenue for the customer

Aggregate\_SMS\_Rev – this is the sms revenue for the customer

Aggregate\_Data\_Rev – this is the data revenue for the customer

Aggregate\_Data\_Vol – this is the data volume used by the customer

Aggregate\_Calls – these are the calls that was made by the customer

Aggregate\_ONNET\_REV – The revenue earned by the calls etc. made to the on-network (on the same network as the subscriber) customers by the carrier’s present subscriber.

Aggregate\_OFFNET\_REV – The revenue earned by the calls etc. made to the off-network (not the same network as the subscriber) customers by the carrier’s present subscriber.

Aggregate\_complaint\_count - this the number of complaint by customer

aug\_user\_type – Is the user a 2G or a 3G user in the month of August

sep\_user\_type – Is the user a 2G or a 3G user in the month of September

aug\_fav\_a – what is the customers favorite other network in the month of August

sep\_fav\_a - what is the customers favorite other network in the month of September

Class – has the customer churn or still active.