Case study

TELECOM CHURN –CASE STUDY

BUSINESS PROBLEM

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another.
- Telecommunication Industry is very Competitive and experience about 15% to 25% annual churn rate.
- Here it costs 5-10 times more to acquire new customer than to retain existing customer.
- Hence Costumer retention is more important than customer acquisition.

DEFINITION OF CHURN IN TELECOMMUNICATION

There are two main models of payment in the telecom industry:-

(1) POSTPAID-customers pay a monthly/annual bill after using the services.

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

(2) PREPAID-customers pay/recharge with a certain amount in advance and then use the services.

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.

So Churn in Prepaid is very critical and should be defined carefully.

CHURN DEFINITION USING FOR THE CASE STUDY

- **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.

BUSINESS OBJECTIVE

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.

Steps involved in the prediction

- Reading and Understanding the Data
- Exploratory Data Analysis
- Modeling
- Evaluating model on test set
- Business Insights

1) Reading and Understanding the Data

- This is large dataset with 226 columns and 99,999 number of rows.
- Out of 226 columns, 179 are float variables, 35 integer and 12 object variables.

2) Exploratory Data Analysis

- Following steps are performed as part of EDA:
- Handling null values
- Dropping columns (Date column, column with only 1 unique value, mobile number, circle id and columns with null values > 35%)
- Getting high value customers

- Tagging churn customers (Class imbalance check)
- Handling outliers for numerical columns
- Extracting useful insights using EDA
- Checking correlation matric for all data month-wise
- Tenure analysis

3) Modeling

- Data is pre-processed with train-test split (80-20%) and SMOTE(Synthetic Minority Oversampling Technique) is used to handle data imbalancing.
- Feature has been scaled using standard scaler.
- 3 basic models are created:
- 1) Basic logistic regression with RFE (30 features)
- 2) Decision Trees (With max depth = 4)
- 3) Basic random forest model (With max depth = 4)

- Optimized random forest model is tried with grid search cv with different values of max depth, n estimators, max features and min samples leaf
- On training set, accuracy value is close to 97% and there is also high value for all the other evaluation measures like specificity, sensitivity, precision, recall and F1 score with all values in range (96 – 97%)

4) Evaluating model on the test set

- Final model has overall accuracy around 92% on the test dataset and it also has high specificity value around 94%. Sensitivity is also comparatively good with value around 64%.
- Precision value is close to 38% whereas recall value is close to 64%. And F1 score is close to 48%.

5) Business Insights

- **Few high-value customers are churning, but there have been no new high-value customers onboarded in the last 6 months, which is concerning.**
- **Customers with less than 4 years of tenure are more likely to churn, so the company should focus on this segment and introduce new schemes.**
- **Average revenue per user is a crucial factor in predicting churn.**

- **Strong indicators of churn behavior include:**
- **Incoming and outgoing calls on roaming in the 8th month.**
- **Local outgoing calls made to landline, fixed line, mobile, and call center.**
- **Better 2G/3G area coverage, especially in areas with poor 2G/3G services.**