DATA 5000 PROJECT PROPOSAL

CUSTOMER CHURN ANALYSIS IN TELECOM INDUSTRY: A CLOSE LOOK ON IBM SAMPLE DATA

Saad Hasan & Chanpreet Singh

ABSTRACT

This project proposal discusses about the churn prediction of telecom company customers and relationship between factors contributing the customer to leave the company

KEYWORDS

Churn analysis, IBM data, prediction, supervised learning, and classification

1. INTRODUCTION

Telecom churn has mainly two types includes voluntary and involuntary. Involuntary customers are those who are forcefully remove by the telecom company due to overcharges, fraud. Voluntary customers are those who left the company based on many factors that could be poor performance, cost and could be the bad behavior of company operators or customer personal reasons. Voluntary churn has two types as well which include incidental and deliberate churn. incidental churn is happening when customer either moved to different location or due to change in financial conditions. While deliberate churn mainly due to quality, advancement in technology and economic reasons. Most of the telecom service providers try to focus more on deliberate churn rather than other types. [1]

Customer is more important than everything for the company because it is big loss for the company to lose an older customer even get a new customer. For telecom industry where almost, every company is on the race to compete their rivals in terms of technology, increasing businesses, providing better services to the customer to earn maximum profit.

2. MOTIVATION

In Canada lots of telecom service providers are keep trying to make more profit by increasing numbers of customers. Giant companies include Rogers communication, BCE Inc., Tellus Corporation and others. By applying different data mining techniques on this IBM sample dataset will be helpful to understand the churn prediction in Canadian Telecom industry for the large dataset.

3. PROBLEM STATEMENT

As the problem is to predict the telco customer churn. There are many classification algorithms could be implemented to classify whether the customer leave or not. Famous methods which includes Logistic Regression, Random forest and decision trees.

Furthermore, Artificial Neural Network with forward feed could also be implemented to check the accuracy of the churn prediction.

4. OBJECTIVES

As the problem discussed in the paper, the main objectives of the project are the following,

- First the question arises which factors to consider and are important for the telecom company for churn analysis. In other words, the features contributing the customers to churn.

- In predicting the churn, a comparison of different algorithms to compare based on accuracy, specificity, recall and precision.

- Furthermore, unsupervised learning, Clustering could also be done in this dataset to cluster the telco customers based on the importance for Telecom company like Highly valuable customer, moderate customers and low valuable customers.

5. DATASET

For the project, the dataset containing 21 features and 7044 rows. The features include customer ID, Gender, contract time, monthly charges, total charges etc. which could take as input variables. For the output variable churn to check the accuracy, recall and precision of the techniques used in the project.

6.REFERENCES

[1] Manpreet Kaur and Dr. Prerna Mahajan, "Churn Prediction in Telecom Industry Using R," International Journal of Engineering and Technical Research (IJETR), vol. 3, Issue.5, May. 2015

[2] IBM data analysis on customer support retrieved from
https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets