Laxmi Prasanna Chennadi

DATA 606

Deliverable-2 Report

**Analysis of Telecom Churn Data**

**Literature Survey/Preliminary Work:**

Churn analysis is the evaluation usually done by the organization on the customer’s loss rate. It is very helpful for many organizations as it will help them in analyzing the reason for the churn and come up with precautionary methods as well as solutions for solving the issues related to churn in an organization. Frequent analysis of churn data is very important. The organizations are very interested in analyzing these trends and use it to reduce churn rate as the price for acquiring a new customer is usually higher than retaining the old one. It is important to deal with churn for every organization in the consumer market and enterprise sectors as it has an impact on the revenue and policy decisions of the organization. According to the authors of “Leading on the Edge of Chaos”, a 2% increase in customer retention (or decreasing churn) is equivalent to 10% reduction in costs. So, it is no wonder that companies that care about customers pay a lot of attention to Churn Analysis.

**Exploratory Data Analysis:**

Initially, we perform the basic steps of data loading and analyzing to understand the characteristics of the dataset. We then start the data cleaning process where we check for missing values as well as other inconsistencies. In this dataset, we can observe that there are no missing values. We perform conversion of data types to make the data more reliable and relevant for analysis. In order to do the exploratory analysis, heat map and correlation methodology is used to understand the correlation between the columns. From this, we can observe that there is correlation between churn and few other columns like total day charge, customer service calls and international plan. They are dependent on each other and by analyzing this using visualization techniques, we can find out the dependency (direct or inverse). We can find out the impact of these columns on the churn rate. We also try to identify outliers in these columns and try to clip them and by doing so, better analysis results can be obtained.

**References:**

* <https://www.researchgate.net/publication/321723500_Review_of_Customer_Churn_Analysis_Studies_in_Telecommunications_Industry>
* <https://towardsdatascience.com/machine-learning-powered-churn-analysis-for-modern-day-business-leaders-ad2177e1cb0d>
* <http://www.computerscijournal.org/vol10no1/churn-analysis-in-telecommunication-using-logistic-regression/>
* [https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6 )