Telecom Churn Analysis

TEAM MEMBERS

1) CHANDAN KUMAR RAXIT
2) DEEPAK KUMAR JENA

Abstract:

The telecoms industry is a highly competitive sector which is constantly challenged by customer churn or attrition. This study implemented exploratory data analysis and feature engineering in a public domain Telecoms dataset and this study discussed how these results are essential in reducing customer churn and improving customer service.

For a telecom industry it is very hard to acquire new customers so they should focus on how to retain existing customers. The principal contribution of our project is to develop a customer churn analysis using exploratory data analysis which helps telecom operators to analysis Telecom's dataset and how these results are essential in reducing customer churn and improving customer service.

In given data we are looking for churn analysis for given data we first performer EDA

INTRODUCTION:

Customer churn is one of the biggest fears of any industry. From various studies in the past, we know that the cost of acquiring a new customer has been far greater than retaining one. Churn or churn rate is defined as the percentage of customers who stop subscribing to a service or percentage of employees leave a job. Banks, insurance and telecommunication companies can be Customer's loss is a major problem for companies which are likely to lose their customers easily. Banks, insurance and telecommunication companies can be given as examples. For companies, the cost of acquiring new customers is increasing day by day. Therefore, retaining customer is much more important than anything.

BUSINESS UNDERSTANDING; This initial phase of data analysis focuses on understanding the objectives of the project and requirements from a business point of view, and then converting this knowledge into a data analysis problem definition. Customer retention consists of “Identifying which customers are likely to Churn, determining which customers should retain and developing strategies to retain profitable customers”. The main thing in retention process is identifying Churn ratio which is a very meaningful and vital determination for many companies. Determination of Churn ratio indicators is also very important. By using those indicators, firms can make prediction on future behavior of new customers and can develop new strategies much before customers start to think about churn. Thus, it is vital to build a very successful and accurate Churn model during the retention studies.

CUSTOMER CHURN:

If a customer terminates a membership with one company and become a customer of another company, this customer is called as Churn customer. Today’s economic trend dictates that price cuts are not the only way to build customer loyalty. Accordingly, adding new value-added services to the products has become an industry norm to have loyal customer. The main goal of Customer’s loss is a major problem for companies which are likely to lose their

Churn has affected industries such as banking, insurance, internet streaming and telecommunications to just name a few. Although there are many reasons for customer churn, some of the major reasons are service dissatisfaction, costly subscription, and better alternatives. Hence, in this paper the problem of churning is addressed and data factors affecting the churn are analyzed for their effect on the rate

REASONS FOR CUSTOMER CHURNING:

- Price: comparatively high Pricing leads the customers to flee from one carrier to a competitor.
- Service quality: Lack of network coverage may make a customer go to another company with good network coverage.
- Lack of customer service: Slow or no response to customer complaints makes a customer more likely to churn.
- Billing disputes
- New competitors entering the market.
- Competitors introducing new products or technology

PROBLEM STATEMENT:

Maximize: Company's profit by retaining customer

Minimize: Customer churn by identifying the key cause of the problem

Finding factors and cause those influence customers to churn. Retain churn customers by taking appropriate steps providing offers based on affecting factors. Using the data provided, this paper aims to

DATA DESCRIPTION:

The data description phase starts with an initial data collection and proceeds with activities in order to get familiar with the data. Identifying data quality problems, discovering first insights into the data and detecting interesting subsets to form hypotheses from hidden information are activities of this step. Data which is collected from a telecommunication company to get analysed, involves usage details of customers from. The data was taken from Orange Telecom Company. It had 3333 rows and 20 columns. Most columns related to subscriber personal. Other column was indicative of service usage by the subscriber. Based on the business understanding of the data 18 columns was chosen to analyse the data

DATASET PREPARATION:

The customer churn dataset from orange telecom company contains 20 features and 3333 observations. The feature 'Churn' shows customer churn or non-churn based on existing conditions. Approximately 14.5% are churn and 84.5% are no churn. Below Table shows the data features.

Data-set description

Feature Name	Type
State	object
Account length	Int64
Area code	Int64
International plan	object
Voice mail plan	object
Number vmail messages	Int64
Total day minutes	Float64
Total day calls	Int64
Total day charge	Float64
Total eve minutes	Float64
Total eve calls	Int64
Total eve charge	Float64
Total night minutes	Float64
Total night calls	Int64
Total night charge	Float64
Total intl minutes	Float64
Total intl calls	Int64
Total intl charge	Float64
Customer service calls	Int64
Churn	bool

FEATURE BREAKDOWN:

STATE: 51 Unique States

Account Length: Duration of length customer use their The Account

Area Code: There are 3 unique area code present 415, 408, and 510

International Plan: Yes, Indicate International Plan is Present and No Indicates no subscription for International Plan

Voice Mail Plan: Yes, Indicates Voice Mail Plan is Present and No indicates no subscription for Voice Mail Plan

Number vmail messages: Number of Voice Mail Messages ranging from 0 to 50

Total day minutes Total Number of Minutes Spent by Customers in Morning

Total Day calls: Total Number of Calls made by Customer in Morning.

Total day charge: Total Charge to the Customers in Morning.

Total eve minutes: Total Number of Minutes Spent by Customers in Evening

Total eve calls: Total Number of Calls made by Customer in Evening.
Total eve charge: Total Charge to the Customers in Morning.
Total night minutes: Total Number of Minutes Spent by Customers in the Night.
Total night calls: Total Number of Calls made by Customer in Night.
Total night charge: Total Charge to the Customers in Night.
Churn: churning status of the customer

EXPLORATORY DATA ANALYSIS:

If we want to explain EDA in simple terms, it means trying to understand the given data much better, so that we can make some sense out of it. we using univariate frequency analysis was conducted to describe key characteristics of each feature including, minimum and maximum value, average, standard deviation and others. It was also used to produce a value distribution and identify missing values, and outliers. EDA is a process of examining the available dataset to discover patterns, spot anomalies, test hypotheses, and check assumptions using statistical measures. In this chapter, we are going to discuss the steps involved in performing topnotch exploratory data analysis summarizing the data, finding the hidden correlation and relationships among the data, developing predictive models, evaluating the models, and calculating the accuracies. Some of the techniques used for data summarization are summary tables, graphs, descriptive statistics, inferential statistics, correlation statistics, searching, grouping, and mathematical models.

- **MISSING VALUES:**

There is a representation of each service and product for each customer. Missing values may occur because not all customers have the same subscription. Some of them may have a number of service and others may have something different. In addition, there are some columns related to system configurations and these columns may have null values but in our orange telecom data set there are no null values present

If there are missing values in the Dataset before doing any statistical analysis, we need to handle those missing values.

- **a) Numeric -Numeric Analysis:**

Analyzing the two numeric variables from a dataset is known as numeric-numeric analysis. We can analyze it in three different ways.

  - Scatter Plot
  - Pair Plot
  - Correlation Matrix

- **b) Numeric - Categorical Analysis:**

Analyzing the one numeric variable and one categorical variable from a dataset is known as numeric-categorical analysis. We analyze those mainly using mean, median, and box plots.

- **CORRELATION AMONG VARIABLES**:

In words, the statistical technique that examines the relationship and explains whether, and how strongly, pairs of variables are related to one another is known as correlation. Correlation

answers questions such as how one variable changes with respect to another. If it does change, then to what degree or strength? Additionally, if the relation between those variables is strong enough, then we can make predictions for future behavior

GRAPHICAL REPRESENTATION OF THE RESULTS:

This step involves presenting the dataset to the target audience in the form of graphs, summary tables, maps, and diagrams. This is also an essential step as the result analyzed from the dataset should be interpretable by the business stakeholders, which is one of the major goals of EDA. Most of the graphical analysis techniques include Line chart, Bar chart, Scatter plot, Area plot, and stacked plot Pie chart, Table chart, Polar chart, Histogram, Lollipop chart etc.

Conclusion:

There were two main objectives of the analysis. First, to segment the customers based on an unsupervised learning method. Second, was to predict churn in the respective segment and assess if segmentation helped improve our prediction, which was demonstrated by a considerable improvement on our baseline. The metric for model performances was misclassification rate, which decreased significantly from our baseline of 41% to approximately 28% for both the segments. Companies can use this process to segment and benchmark processes to determine who is at risk for leaving or discontinuing services. Benchmarking improvement in your models is important to ensure that companies are meeting the needs of all their clients uniquely.

REFERENCES:

- Hands-On Exploratory Data Analysis with Python Perform EDA techniques to understand, summarize, and investigate your data by Suresh Kumar Mukhiya, Usman Ahmed (z-lib.org)
- https://bunker2.zlibcdn.com/dtoken/01c5fc197a94283bfb0c0943bd5b2d0c