International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

TELECOM CHURN PREDICTION USING MACHINE LEARNING ALGORITHMS

N.GURUMOORTHI

ABSTRACT

Customers are the base for any business success and that is why firms become aware of the significance of acquiring satisfaction of customers. Customer churn is an essential issue and it is regarded as one of the most essential concerns among firms because of increasing rivalry among firms, increased significance of marketing strategies and customers conscious behavior in present years. Organizations must develop different strategies to resolve the churn issues relying on the services they offer. Customer churn practice is essential in competitive and rapidly developing in telecom sector. The process of migrating from one service provider to another telecom service provider occurs due to good services or rates or due to various advantages which the rivalry firm provides customers when signing up. Due to the greater cost related with acquiring new customers the prediction of customer churn has developed as an indispensable part of planning process and strategic decision making in telecom sector. The main aim of the study is to explore the customer churn prediction in telecom using in machine learning algorithms.

In this project Machine learning techniques have been used for estimating the telecom customer probability to churn. This study makes use of Logistic Regression, SVM, Random Forest, ADA Boost and XG boost with big data for predicting consumer churn in the telecom sector. Logistic regression has been used widely to estimate the probability of churn as a function of variables set or features of customers. Support Vector Machine (SVM) has been successfully used in many applications such as image recognition, medical diagnosis, and text analytics. *Random forest* is a flexible, even without hyper-parameter tuning, a great result most of the time. ADA boost works on the principle of learners growing sequentially. Similarly, XG boost implementation of gradient boosted decision trees designed for speed and performance. This study

uses Kaggle's website dataset for telecom churn to predicting and analyzing churn. The results of the study show that the accuracy rate of prediction in consumer churn is found to be 0.829 percent for a XG Boosting method.

OBJECTIVE:

Telecom Customer churn is important because it's more expensive to acquire a new customer than to sell more to an existing client. In fact, this is one metric that can make or break a business: If you do a better job keeping telecom customers around, you should see your average telecom customer lifetime value increase, making every future sale even more valuable and ultimately improving your unit margins. In fact, the best use of company resources is often to gain an increase in recurring subscription revenue or reliable repeat business versus investing more in new telecom customer acquisition. Keep loyal telecom customers for many years and you'll have a much easier time growing and weathering financial rough patches versus having to spend to bring in new customers to replace those who left.

Telecom Churn rates need to be assessed in line with the industry average, or it may be much better or much worse. If yours is a telecom industry where churn data is available publicly or through an industry trade organization, it's well worth benchmarking your results against those of competitors. High-performing telecom companies often set a below-industry-average target telecom churn level; track their metrics over time, perhaps using a KPI scorecard; and act promptly to address blips before they become unfavorable trends.

Predicting **customer churn** is critical for **telecommunication companies** to be able to effectively retain customers. It is more costly to acquire new customers than to retain existing ones. For this reason, large telecommunications corporations are seeking to develop models to predict which customers are more likely to change and take actions accordingly. The objective is to obtain a data-driven solution that will allow us to reduce churn rates and, as a consequence, to increase customer satisfaction and corporation revenue using different machine learning algorithms.

SOFTWARE REQUIREMENTS

- Language : Python 3.8
- Operating System : Windows/ Linux / macOS
- Tools : Anaconda Navigator / Google Colab
- Dataset : Kaggle dataset

SOFTWARE REQUIREMENT SPECIFICATION

A Software requirements specification (SRS), a requirements specification for a software system, is a description of the behaviour of a system to be developed and may include a set of use cases that describe interactions the users will have with the software.

PYTHON

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is a must for students and working professionals to become a great Software Engineer specially when they are working in Web Development and analytics Domain. I will list down some of the key advantages of learning Python:

- **Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository, install them in an environment, run the packages and update them. It is available for Windows, macOS and Linux.

The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- QtConsole
- Spyder
- Glue
- Orange
- RStudio
- Visual Studio Code

KAGGLE DATASET

Kaggle is a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Kaggle has run hundreds of machine learning competitions since the company was founded.

Competitions have ranged from improving gesture recognition for Microsoft Kinect to making a football AI for Manchester City to improving the search for the Higgs boson at CERN.

Competitions have resulted in many successful projects including furthering the state of the art in HIV research, chess ratings and traffic forecasting. Most famously, Geoffrey Hinton and George Dahl used

deep neural networks to win a competition hosted by Merck. And Vlad Mnih (one of Hinton's students) used deep neural networks to win a competition hosted by Adzuna. This helped show the power of deep neural networks and resulted in the technique being taken up by others in the Kaggle community. Tianqi Chen from the University of Washington also used Kaggle to show the power of XGBoost, which has since taken over from Random Forest as one of the main methods used to win Kaggle competitions.

Several academic papers have been published on the basis of findings made in Kaggle competitions. A key to this is the effect of the live leaderboard, which encourages participants to continue innovating beyond existing best practice. The winning methods are frequently written up on the Kaggle blog.