International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Telecom Churn Prediction using Machine Learning Algorithms

Mohammed Zakir Nooraj

Student, Rajam, Vizianagaram, 532127, India.

ABSTRACT

The field of study referred to as data science works with enormous amounts of data using reducing tools and methods to uncover hidden patterns, glean valuable information, and make business decisions. To create prediction models, data scientists use sophisticated computer algorithms. The data used for analysis can be presented in a variety of ways and come from a wide range of sources. Banking and finance were the first businesses to fully grasp the power of data science. Data science is frequently used for risk analysis and risk management. Businesses are looking for data scientists for portfolio management to use business intelligence tools to assess financial patterns through data. In addition to these sectors, data science is used in sectors that use data in healthcare, automotive, telecommunications, etc. In many situations, including product improvement, network security, fraud detection, predictive analysis, and reducing customer turnover, data science is crucial. Future client attrition can be extremely damaging to a digital company. This is frequently the case, particularly as a business and its products develop and it is harder to find new clients. This paper mentions machine learning techniques used in the telecom industry, including Artificial Neural Networks (ANN), and K-Nearest Neighbor (KNN)

Keywords: *Churn, Machine algorithms, Software tools, Data models.*

1. Introduction

Customer loyalty is the key to profitability in the telecom industry because telecommunications companies are typically not the most well-liked businesses with consumers. People frequently complain about the performance of service providers, whether it be too convoluted billing, unsolicited emails with marketing messages, challenging customer support, slow internet, poor connectivity, or expensive plans. The fact that telecommunications businesses have a high customer churn rate is not surprising as a result. Customer turnover (attrition) is an issue in this sector since telecom companies operate sizable fixed infrastructures that must be compensated by revenue.

The rate at which consumers cease doing business with a company is known as the churn rate, sometimes referred to as the rate of attrition or customer churn. Most frequently, it is stated as the percentage of service subscribers who cancel their memberships within a specified time frame. In developing areas, the turnover rate ranges from 20% to 70%. More than 90% of all mobile customers use prepaid services in several of these markets. In underdeveloped areas, some operators lose the entirety of their subscriber base through churn in a single year. The churn rate formula is:

100*times(Lost Customers/Total Customers at the Period's Beginning)

The churn can be determined using this formula. Divide the number of clients you had last quarter by the number you had at the beginning of the previous quarter. It displays the percentage loss.

The classification and Prediction algorithms of artificial intelligence (AI) provide strong support for churn prediction techniques. High dimensionality and unbalanced datasets present challenges for these classification algorithms, making reliable churn prediction difficult. The amount of data available for analysis, along with the dataset's balance or unbalancedness, has a significant impact on the effectiveness of some data mining techniques. The category imbalance problem, which occurs when the churn class (minority class or class of interest) churns samples more than the non-churn class, is shown by churn datasets (majority class). Because of this, it is more difficult for some machine learning techniques to identify the minority class. Because of this, ensemble approaches for churn prediction show good results when addressing imbalanced datasets.

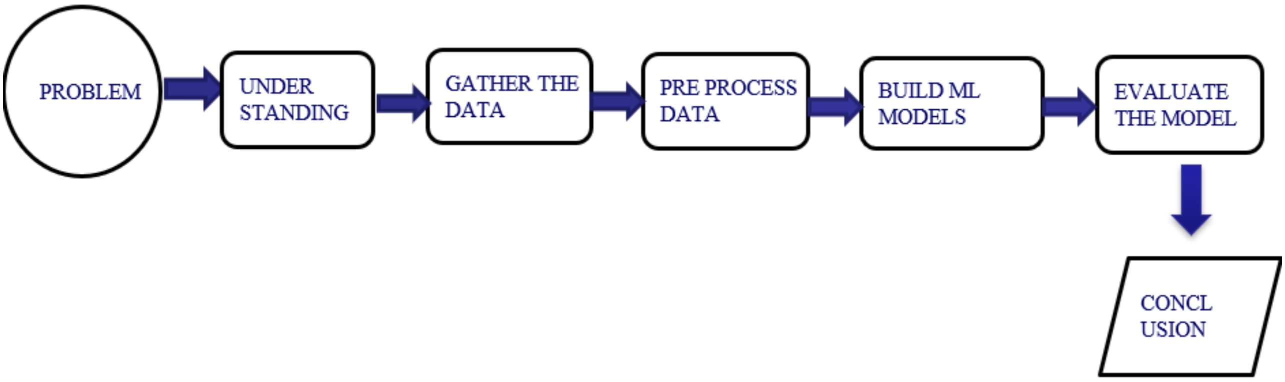


Fig-1: -Stepsto Solving Problem inTheData Science

2. Literature Survey

In paper [1]. A Systematic Review of Methods Used for Predicting Telecom Churn The authors of the report employ a variety of algorithms to forecast churn in the telecom sector. The major goal of this trio is to create a churn prediction model that aids telecom providers in identifying customers who are likely to leave. They discussed several algorithms and their advantages with regard to customer churn. They also employ a hybrid algorithm (ANN+KNN) for predicting customer turnover in the telecom sector. After comparing all of the data, they determined that the hybrid algorithm model performed more accurately than the others.

In paper [2]. Use of Machine Learning Technique for Predictive Analysis and Modeling of Customer Churn in Telecom. suggested a model combining multiple machine learning and data processing techniques, such as Logistic Regression and Decision Trees, and compared the effectiveness of the two algorithms to examine consumer churn prediction in telecom firms. The results of the work have been erased using R programming, and the confusion matrix is used to illustrate the findings. The training and test phases have explicitly included data cleaning. The shine package in R was used to create an internet interface that displayed results graphically. The statistical survival analysis tool has been used to do this in order to forecast churn and contrast two algorithms.

In paper [3]. proposed a compound prediction model for churn modeling within the cellular telecom organization. The machine learning program WEKA was used to conduct the study. Their main objective is to evaluate the model's application and develop a new hybrid model that is more accurate than the existing single method. As a result, utilizing the Voted Perceptron in conjunction with the clustering mixed logistic regression approach, we were able to pinpoint the reason for churn as well as the delay between the moment of deactivation and the decision to churn. Each attribute will be explained from the data and chosen based on its ability to forecast, and then

For starting the feature selection approach and creating a model using logistic regression, they used various statistical tests. It has been found that the logistic regression model has produced improved churn prediction outcomes. It will produce a better outcome by raising the edge values and choosing the appropriate features in various combinations.

In paper [4]. Using a Random Forest Classifier, the telecommunications industry can predict customer churn. Divide and conquer is the strategy used by Random Forest. The precision rate of the Random Forest is 0.893, which is excellent when compared to other algorithms. A creates the numerical sort of decision tree, and every Decision Tree is trained by selecting any random attribute from the entire predictive set of attributes.

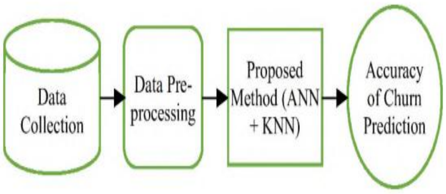
Carefully make a decision. The company may create a model from the customer's reaction that will be more useful to decision-makers because it can identify comparable groups of customers based on their responses and estimate the chance of churn.

Through RF, they successfully classified 88.63% of the data. However, because RF creates a complicated forest that is challenging to perceive and perform rule inference on, it is inappropriate for rule generation for factor identification. As a result, a comparable classifier, such as Attribute Selected Classifier, is employed in this study's rule creation for factor identification. We will also achieve "one-to-one marketing," which prioritises marketing tospecific customers, based on outcomes and customer habits.

3. Methodology

Using a hybrid algorithm, predict churn

The primary goal of the study is to look at a comprehensive analysis of the algorithms used for telecom chum prediction. This study suggests using a hybrid technique (ANN + KNN) to anticipate telecom chum. A brand-new KLMM (K local maximum margin) feature extraction algorithm is suggested by the research. The field's specific fundamental structure is produced through perusing on diversity of miles of subspace partition. The scalability dimension demonstrates the inherent relationship between categorization results and data properties based on the source of the data. The chum prediction dimension in the telecom dataset can be reduced by the extracted features. The auto-selection sigma factor is used in the K-Local Maximum Margin technique to reflect the anisotropy of the feature. To evaluate the attribute weights and forecast the essential function, use the essential function.



3.2 Description of Hybrid Algorithm

Together, several simplistic algorithms support and improve one another. Together, they can find solutions to issues that they were not intended to handle separately. There are several different kinds of HML techniques that work with the data in various ways.

3.3 Working On Hybrid Algorithm
:

- Data Collection
- Data Pre-processing

The stage when we clean up the data is called processing. As well, data accuracy is checked.
Cleaning and formatting

- Modelling the data
- Accuracy of Churn Prediction

3.4 K-Local Maximum and Minimum

The points of the functions that define the maximum and minimum range are local maximum and minimum. The derivative of the function can be used to calculate the local maximum and minimum. The two key techniques for locating the local maximum and minimum are the first derivative test and the second derivative test.

Let's find out more about finding the local maximum and minimum, the techniques for doing so, and some examples of local maximum and minimum.

7. Results and Discussion

An Ann+Knn hybrid algorithm for churn prediction has a 90% accuracy rate. We may infer from the Spectrogram that the accuracy of the hybrid algorithm used for churn prediction is between 85% and 90%.

Table -1

S.NO	Algorithms used
1	ANN
2	KNN

8. Conclusion

The main goal of any business nowadays, regardless of sector, is to draw clients. Because of the advancement of technology, people have all become highly intelligent and are constantly looking for new advantages. Let's give an illustration: Telenor and Docomo Sims were accessible a few years ago, but after Jio was founded, everyone stopped using these networks because Jio was providing more telecom services and offers. Churns are causing networks like Telenor and Docomo to discontinue operating. The investigations come to the conclusion that the hybrid algorithm used for churn prediction has provided better accuracy of 85 to 90% .

REFERENCES

I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019, DOI: 10.1109/ACCESS.2019.2914999.

A M. B. A. Joolfoo, R. A. Jugumauth and K. M. B. A. Joolfoo, "A Systematic Review of Algorithms applied for Telecom Churn Prediction," 2020 3rd In on Emerging Trends in Electrical, Electronic, and Communications Engineering (ELECOM), 2020, pp. 136-140, DOI: 10.1109/ELECOM49001.2020.9296999.

V. Geetha, A. Punitha, A. Nandhini, T. Nandhini, S. Shakila and R. Sushmitha, "Customer Churn Prediction In Telecommunication Industry Using Random Forest Classifier," 2020, pp. 1-5, DOI: 10.1109/ICSCAN49426.2020.9262288.

A. Chowdhury, S. Kaisar, M. M. Rashid, S. S. Shafin and J. Kamruzzaman, "Churn Prediction in Telecom Industry using Machine Learning Ensembles with Class Balancing," 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2021, pp. 1-6, DOI: 10.1109/CSDE53843.2021.9718498.

A. Siddika, A. Faruque and A. K. M. Masum, "Comparative Analysis of Churn Predictive Models and Factor Identification in Telecom Industry, 2021, pp. 1-6, DOI: 10.1109/ICCIT54785.2021.9689881.

P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016, pp. 1-4, doi: 10.1109/CDAN.2016.7570883.

U. Yabas, H. C. Cankaya and T. Ince, "Customer Churn Prediction for Telecom Services," 2012 IEEE 36th Annual Computer Software and Applications Conference, 2012, pp. 358-359, doi: 10.1109/COMPSAC.2012.54.

S. Wu, W. -C. Yau, T. -S. Ong and S. -C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," in IEEE Access, vol. 9, pp. 62118-62136, 2021, doi: 10.1109/ACCESS.2021.3073776.

Wai-Ho Au, K. C. C. Chan and Xin Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," in IEEE Transactions on Evolutionary Computation, vol. 7, no. 6, pp. 532-545, Dec. 2003, doi: 10.1109/TEVC.2003.819264.

C. Phadke, H. Uzunalioglu, V. B. Mendiratta, D. Kushnir and D. Doran, "Prediction of subscriber churn using social network analysis," in Bell Labs Technical Journal, vol. 17, no. 4, pp. 63-76, March 2013, doi: 10.1002/bltj.21575.

T. Lee, J. Nam, D. Han, S. Kim and H. Peter In, "Developer Micro Interaction Metrics for Software Defect Prediction," in IEEE Transactions on Software Engineering, vol. 42, no. 11, pp. 1015-1035, 1 Nov. 2016, doi: 10.1109/TSE.2016.2550458.

M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," in IEEE Transactions on Neural Networks, vol. 11, no. 3, pp. 690-696, May 2000, doi: 10.1109/72.846740.

E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang and H. K. Kim, "Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games," in IEEE Transactions on Games, vol. 12, no. 1, pp. 41-53, March 2020, doi: 10.1109/TG.2018.2871215.

A. Idris and A. Khan, "Churn Prediction System for Telecom using Filter–Wrapper and Ensemble Classification," in The Computer Journal, vol. 60, no. 3, pp. 410-430, March 2017, DOI: 10.1093/comjnl/bxv123.

W. Bi, M. Cai, M. Liu and G. Li, "A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn," in IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1270-1281, June 2016, DOI: 10.1109/TII.2016.2547584.