© January 2024| IJIRT | Volume 10 Issue 8 | ISSN: 2349-6002

Customer Churn Prediction in Telecom Using AI-ML

T Prithvi Charan[1], Tanush T [1], Vinyas S R Gowda[1], Mrs. Keerthi Mohan[2]

[1] Student, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India

[2] Professor, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India

Abstract –The telecom industry generates enormous amounts of data every day from a large number of clients. The phenomenon known as "churn" refers to customers switching companies within a predetermined amount of time. Getting new clients is more expensive than maintaining your current clientele. Managers and analysts in the telecom industry need to understand why customers terminate their subscriptions and what patterns of behavior they consistently see. This technology identifies the customers who are most likely to leave the telecom industry by compiling the factors that affect their decisions and applying classification algorithms to that data. In order to give customers retention strategies and plans, this system's primary goal is to compare various machine learning algorithms that may be used to create customer churn prediction models and identify the reasons behind churn. This system uses machine learning techniques like KNN and decision tree classifiers, along with classification algorithms like Random Forest (RF) to analyze customer turnover data. It provides a successful business model that can precisely forecast customer attrition and assist management in taking appropriate action during the attrition period to avoid attrition and lost revenue.In order to give customers retention strategies and plans, this system's primary goal is to compare various machine learning algorithms that may be used to create customer churn prediction models and identify the reasons behind churn.

1. INTRODUCTION

The amount of clients that leave a business each year is known as the churn rate. This could present a problem for any company. Therefore, one of the most important tasks for any business is to predict which customers might want to leave. Numerous Customer Relationship Management (CRM) systems have been integrated as data analytical techniques due to recent advancements in data analytics, and this has led to numerous studies and practices focusing on these methods. These analytical techniques prioritize customer-centric strategies over product-centric strategies. As a result, interactions between the company and its customers have changed to the point where many new advertising opportunities have arisen. Reducing churn and retaining current clients are the most profitable marketing strategies that maximize shareholder value.

In the telecom sector, customers are still the most valuable asset for a business to have in order to stay in business. For businesses, a decrease in their customer base is an unforeseen circumstance. Because of this, companies need to look at customer profiles in order to segment their markets and make better choices. One of the biggest concerns in today's fiercely competitive industries, like the telecom sector, is customer attrition.

The telecom industry has focused on retaining current customers due to the high cost of acquiring new ones. Retaining current customers generates more sales and cheaper marketing expenses than acquiring new ones. As a result, forecasting customer attrition has emerged as a crucial component of the strategic executive and planning process in the telecom industry.

A report claims that the telecom sector experiences an annual churn rate of between 20% and 40% and that keeping existing customers is five to ten times less expensive than bringing in new ones. It costs 16 times less to predict customer turnover than it does to bring in new business. When the churn rate is lowered by 5%, the profit rises from 25% to 85%. This highlights the significance of anticipating customer attrition in the telecom sector. CRM is essential for telecom businesses to keep their current clientele and lower attrition. If analyzers fail to predict customer attrition accurately, no advertising can be run. Recent developments in data science have led to improvements in data mining and machine learning technologies, which offer solutions for decreasing customer attrition.

Businesses require personalized marketing strategies—which can be implemented through social media marketing—in order to strengthen their customer relationships and attract more loyal clients. In order to achieve this, they must provide training so that their sales and customer service representatives can effectively communicate with each and every one of their clients, as well as the capability to gather data about the company's users. Such a task could be easily completed in the era of big data using AI techniques, sparing the customer support and sales teams a great deal of effort. To successfully engage customers and gain their trust, AI must be incorporated into all business operations, including sales, CRM, social media marketing, and marketing.

Understanding how to use, adapt, and implement social media and other electronic marketing platforms effectively is crucial, especially given their significant role in today's business world. By enabling more individualized and focused marketing initiatives, deep learning and customer behaviour analysis can have a substantial impact on a company's social media marketing as well as other marketing initiatives. Businesses can learn what is likely to resonate with their audience by analyzing customer data. Such data can be used by a company to develop social media and marketing campaigns that are more successful. Higher customer engagement and conversion rates will follow from this. The current research on churn prediction has been conducted on a variety of industry customers, such as employee churn, as will be covered in the next section. However, since the telecom industry suffers the most from customer attrition, the bulk of these studies have focused on telecom data. While some researchers have focused on improving accuracy in churn prediction, others have been more concerned with optimizing profit while attempting to predict the churn rate. The target should be increasing prediction accuracy without sacrificing profit, which calls for simpler strategies that can be employed without requiring a significant financial commitment. This is the primary, understudied issue.

This paper investigates different kinds of machine learning strategies to develop a churn prediction model in order to tackle this problem. The model was constructed using traditional classifiers, ensemble learning, and deep learning, and it was tested on two publicly available datasets. The telecom markets in India and Southeast Asia provide the first dataset, while the US telecom market provides the second.

- When compared to non-churn customers, the proposed model is able to overcome the challenge of a lower quantity of churn in the datasets.
- Two public datasets have been used to thoroughly evaluate the suggested model using a variety of performance metrics. They are AUC- (Area Under the ROC Curve) ROC (Receiver Operating Characteristic), recall, accuracy, precision, and F1 score.

In addition, our model outperforms all other related works on churn prediction rate that were compared to the model. Additionally, the suggested models are statistically significant when compared to alternative models, according to the statistical analysis performed using Wilcoxon and ANOVA. Apart from the previously mentioned segment, the document is organized as follows. The related work is discussed in Section 2. The research's methodology, dataset, and instruments are briefly explained.

2. RELATED WORK

Tong and Chang's paper presented a customer churn prediction model that takes customer value into account. An Empirical Research of Telecom Industry in China. This project was put into place to continue serving customers and lower the cost of customer retention while anticipating churned customers. It greatly depends on how customers feel about a specific product and its quality. This model's advantages include KNN, instance-based architecture, and lack of a training phase, which enable rapid adaptation to shifting customer behavior patterns. One of the model's drawbacks is that KNN is memory-intensive and might not be practical for large datasets because it needs to store the entire dataset in memory in order to make predictions. It makes use of technologies like support vector machines (SVM), k-nearest neighbors (KNN), naive bayes neural networks (NBL), and linear regression for task analysis and classification.

A machine learning algorithm-based model for customer churn prediction analysis in a telecom company was proposed by Ravi Kumar Gupta and Geeta Rani. The goal of this project was
to identify customers who are most likely to leave and take proactive steps to keep them by using customer churn prediction.

One of this model's advantages is that Random Forests, which aggregate predictions from several decision trees, typically offer high predictive accuracy. The model's drawback is that Random Forests can require a lot of resources, particularly when handling a big number of trees. It makes use of technology like CCP technique. Four data mining algorithms—Naive Bayes, ELM schemes, RandomForest, and RandomTree—have been assessed for CCP analysis.

Using machine learning methods, B Prabhadevi and Shalini devised a model for customer churning analysis. The purpose of this research was to address It uses a pro- and reactive strategy and waits for the client to ask for the business to end their service connection. One of this model's benefits is that neural networks are excellent at identifying complex patterns and correlations in data, which makes them appropriate for scenarios with complicated customer attrition. This model's downside is that it requires a lot of computing and time to train neural networks, particularly deep architectures. It makes use of technologies including decision trees, KNN, linear regression, naive bayes, and neural networks.

A model for a survey on customer churn prediction using machine learning techniques was suggested by Saran Kumar A and Chandrakala D. The purpose of this initiative was to address the reactive approach. With this strategy, the business waits for the client to seek cancelation before providing incentives to keep them around. Proactive Methodology This strategy entails identifying potential churn clients ahead of time and offering targeted incentives to keep them from leaving. One of this model's benefits is that control charts work well for tracking and preserving steady processes, which guarantees that fluctuations in customer attrition stay within reasonable bounds. This model's limitation is its inability to handle complicated relationships well; control charts might not be able to adequately depict complex interactions. It makes use of technological tools including control charts, generalized additive models (GAMs), and data mining.

Utilizing machine learning techniques, Kiran Dahiya and Surbhi Bhatia suggested a model for customer churn analysis in the telecom industry. In the publication, a methodology for data mining-based churn prediction in the telecom sector is proposed. The model's benefits are In exploratory data analysis, randomized procedures might be helpful in revealing hidden patterns. This model's downside is the efficiency may be impacted by several randomization approaches that elevate computing complexity. Technology like decision trees, logistic regression, and data mining tools are used in it.

The Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study utilizing Machine Learning Techniques model was introduced by Hussain and A. Amin. In order to rebalance the class distribution, this model uses two techniques: internal-level approaches that resample the data using current classification algorithms, and external-level approaches that preprocess the data to lessen the impact of the uneven class distribution. The model's benefits are Naive Bayes is suited for huge datasets because of its computationally efficient training approach. The model's shortcoming is that Naive Bayes works best when there is an adequate volume of training data. It makes use of technology like MTDF and Naive Bayes.

The article "Intelligent Data Analysis Approaches to Churn as a Business Problem: a Survey" emphasizes the value of predictive models in tackling customer desertion while providing a thorough survey of the literature on churn analysis across a range of business industries. The study examines conventional and computational intelligence techniques, such as decision trees, regression, artificial neural networks, support vector machines, Bayesian networks, and ensemble learning, with a focus on banking and telecommunications. It commends the study for its comprehensive overview of the literature and its particular model-building recommendations, but it also criticizes the lack of comparison studies between traditional and computational intelligence approaches, as well as the absence of defined guidelines for using the latter. Notwithstanding these drawbacks, the report offers insightful information on business analytics for attrition and suggests future research avenues.

Using a systematic literature review (SLR) in accordance with the recommendations of Tranfield, Denyer, and Smart, the paper provides a thorough examination of customer attrition in the telecom industry. Thematic content analysis makes relationships more organized by pointing out separate factors that affect churn. Benefits of the SLR include its methodical approach, perceptive theme analysis, and proactive assessment of its shortcomings and potential future study areas. However, there are several drawbacks, such as the study's exclusive focus on English, reliance on scholarly publications, and preponderance of quantitative research, which may exclude other points of view. It could be necessary to narrow the wide range of suggested future study directions. Essentially, the paper offers a methodical and comprehensive analysis of telecom customer attrition, providing insightful analysis and suggestions for more study.

K-means clustering and survival analysis are two data mining techniques used by the Integrated Data Mining and Survival Analysis Model for Customer Segmentation to improve customer churn prediction. Through the use of survival analysis to estimate survival functions and the grouping of consumers based on behaviour factors, the model helps businesses identify high-risk customers and comprehend the distinct churn rates within each cluster. This makes focused retention tactics easier to implement.

The model presents difficulties because of its intricate procedure, which calls for sophisticated statistical and data mining knowledge, even though it excels in consumer segmentation and predictive analysis. Furthermore, there is a challenge in estimating the ideal number of clusters (K-value), which affects the precision of consumer segmentation. In summary, the model provides insightful information but necessitates careful examination of its complexity.

A combination of feature selection, customer profile, classification algorithms, and customized retention methods are used in the churn prediction and customer retention model. It determines important traits by using correlation ranking and information acquisition. Random Forest is one of the classification methods that achieves an accuracy rate of 88.63% in classification. K-means clustering makes it possible to profile customers and implement tailored methods. The model is quite good at identifying client segmentation, retention strategy recommendations, and churn factor identification.

Although it depends on the quality of the dataset, it offers useful retention suggestions, boosts productivity, and makes promotion recommendations based on behaviour patterns. The model emphasizes accuracy and actionable information in a concise way, providing a complete and individualized approach to telecom client retention using machine learning and behaviour analysis, despite computing costs and algorithmic variances.

A strong churn prediction system designed specifically for the telecom industry is provided by the publication "Churn Prediction Model Using Random Forest". It combines feature selection, clustering approaches, and classification algorithms by utilizing machine learning techniques. For precise consumer categorization, the Random Forest method is used, and k-means clustering enables detailed customer profile.

Data preparation is improved by feature selection utilizing Information Gain and Correlation Attributes Ranking Filter. Using the WEKA toolbox, the model is assessed using metrics such as ROC area, F-measure, accuracy, precision, and recall. The ability to precisely classify churn customers, identify factors, and implement tailored retention initiatives are noteworthy advantages. The reliance on top-notch datasets and the difficulty of implementation in actual telecom systems, however, are possible drawbacks.

In order to build a prediction model, the study takes a thorough approach and assesses a variety of methodologies, including deep learning, ensemble learning, and conventional classification approaches. Prominent advantages include improved customer relationship management and targeted retention efforts made possible by the model's accuracy in anticipating attrition. On the other hand, significant drawbacks include the need for substantial data analysis and computer power for model training and validation.

Using public datasets from the Southeast Asian and American telecom markets, the study evaluates the efficacy of a variety of technologies, including ensemble learning (Adaboost, random forest, etc.), traditional classification methods (logistic regression, decision tree, etc.), and deep learning (artificial neural network, convolutional neural network). Essentially, the study emphasizes churn reduction and client retention by integrating machine learning techniques into the telecoms industry in a comprehensive way.

CONCLUSION

Our system effectively addresses the primary objective of analyzing diverse machine learning algorithms essential for developing predictive models related to customer churn in the telecommunications sector. The key focus is on identifying churn reasons and providing retention strategies. Customer churn in this industry is a complex issue with both visible and hidden causes. While it is challenging to control, our system emphasizes the inevitability of customer loss and the need for proactive responses from telecom operators. Our approach employs machine learning techniques, specifically a decision tree classifier and a random forest algorithm, within a customer churn system. The models undergo evaluation before and after employing techniques such as up-sampling and KNN (K Nearest Neighbors). Initial results from the decision tree classifier on an imbalanced dataset were suboptimal, emphasizing the importance of ultimate accuracy. Subsequently, the random forest classifier outperformed the decision tree classifier, achieving an impressive overall accuracy of 99%, with precision, recall, and accuracy metrics all exceeding 99%.

REFERENCE

[1]Tong, Chang, and Zhao, "A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China”,2021.

[2]Gupta and Rani, "Customer Churn Prediction Analysis in a Telecommunication Company with Machine Learning Algorithms”,2021.

[3]Prabhadevi and Shalini, "Customer Churning Analysis Using Machine Learning Algorithms “,2020.

[4]Kumar and Chandrakala, "A Survey on Customer Churn Prediction using Machine Learning Techniques" ,2018.

[5]Hugo Ribeiro, "Determinants of Churn in Telecommunication Services: A Systematic Literature Review”;2020.

[6] Irfan Ullah, Basit Raza" A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector”,2019

[7]Dahiya and Bhatia, “Customer Churn Analysis in Telecom Industry”,2018.

[8]Yash Singh, Vedika Avhad “ Prediction of Customer Churn Using Machine Learning”,2016

[9]Vidushi Mishra, Shubhangi Jain, “Survey on Prediction of Customer Churn Analysis in a Telecommunications Industry”,2016

[10]Hussain, "Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study”,2016.

[11]David L.Garcia, Angela Nebot Alfredo,“Intelligent Data Analysis Approaches to Churn as a Business Problem: a Survey”,2016