IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Predicting Customer Churn On OTT Platforms: Customers With Subscription Of Multiple Service Providers

Bharathi L and Suganthi P

School of Management
SASTRA Deemed University
SASTRA,Thanjavur-613 401

ABSTRACT

In today's competitive landscape, where customers have a plethora of choices across various industries, understanding and predicting customer churn has become paramount for businesses. This study focuses on discerning the factors influencing churn in Over-The-Top (OTT) platforms and predicting customer behavior. Data from 336 respondents, who subscribe to multiple OTT platforms, was collected via questionnaires, encompassing demographic information, platform usage, and user satisfaction metrics. Through a combination of Recursive Feature Elimination (RFE) and feature ranking techniques, factors influencing churn were identified. Hierarchical Logistic Regression was employed to gauge the impact of new variables such as 'Switching Frequency' and 'Primary Platform' on churn prediction. Finally, customer churn prediction was executed using Decision Tree, Random Forest, and AdaBoost methods, with Random Forest yielding superior results.

Keywords: Customer Churn Prediction, Over-The-Top (OTT), Multiple Subscription, Machine Learning Classifiers, Decision Tree, Random Forest, AdaBoost

INTRODUCTON

Customers are the lifeblood of any organization. In today's cutthroat market, customer satisfaction holds more significance than ever. The way a customer perceives not only the product itself but also the brand as a whole is a critical factor in determining a company's success in the modern business landscape. Organizations can either attract new customers or keep existing ones to expand their customer base. Research shows that acquiring new customers costs five times more than retaining current ones, making customer retention a more profitable strategy. Additionally, customer retention has positive social implications that can give a competitive advantage in today's market. Stakeholders naturally lean towards customer retention to boost overall profits. Research provides insights into the customer lifecycle, the processes involved in acquiring and retaining customers. Acquiring new customers requires more time and resources compared to retaining existing ones. The dynamics of the Over-The-Top (OTT) platforms industry, which was once monopolistic, have evolved due to the entry of major players from various sectors. This increased competition has led to a focus on retaining customers by understanding their emotions and reasons for churning. Leveraging the vast amount of data generated by OTT platforms, Machine Learning (ML) emerges as a sophisticated tool to gain insights and drive strategic decisions.

OTT giants are leveraging various ML techniques like Classification models, predictive models, Clustering Algorithms, Neural Networks, and more to differentiate themselves in the market. The correct implementation of these methods enables organizations to intervene at the right time and take action before customers abandon their platform. Apart from customer retention, churn prediction also plays a crucial role in revenue prediction and enhancing customer service. The paper is divided into seven sections. While existing literature on churn prediction models primarily focuses on Telecommunication, Finance, and Retail and E-commerce sectors, there is a lack of extensive and reliable work done specifically for OTT platforms. Furthermore, the trend of customers subscribing to multiple service providers has become more prevalent due to the rise in the number of service providers. This factor has not been adequately addressed in previous literature when developing churn models. Our research takes into account this factor, among others, when constructing predictive models. Additionally, most of the churn prediction work is based on secondary data, which is a reactive approach and provides limited time for customer retention efforts. To adopt a proactive approach, we utilize primary data in our study. This paper identifies the key features that significantly impact customer churn in relation to OTT platforms. Furthermore, we compare the performance of baseline models with various ensemble binary classification models.

In the past, our options for OTT platforms were limited, but in recent years, the situation has changed significantly. The emergence of the covid-19 pandemic acted as a catalyst for major players to enter the market. As the number of service providers increased, so did the churn rate. According to a study by Statista, 77% of people in the USA had a Netflix subscription, while 56% had Amazon accounts, both being giants in the OTT market. These numbers clearly indicate that people now enjoy multiple subscriptions. The research objective is to analyze the data of OTT platform users in order to understand customer preferences, factors influencing customer loyalty, and factors contributing to customer churn for their primary OTT platform.

OBJECTIVES

1) To investigate the factors contributing to customer churn within OTT platforms

2) To develop a robust classification model to effectively predict customers likely to discontinue their subscription on OTT platforms in the near future.

3) To offer suggestions based on the results obtained from the study.

LITERATURE REVIEW

This section discusses the literature available around customer churn prediction. Most of the prediction work is related to the Telecommunication, Finance, Retail, and E commerce sector.

T. Chih-Fong et.al., Many different approaches are applied across various sectors to improve the accuracy of the models. Authors have suggested adding new factors such as social aspects. They have put forward improvised Machine Learning and Deep Learning models to improve the prediction task to help companies with customer churn. and widely used method for churn prediction is classification - a Machine Learning algorithm to classify the customers into different classes basis different factors.

Various Machine Learning and Data Mining classification models like Logistic Regression, Decision Trees, and SVM facilitate customer churn prediction. Generally, studies revolve around optimizing the model performance by augmenting data or improvising algorithms. Some paper talks about time window optimization for improving the performance of Logistic Regression and Classification Trees algorithms.

RESEARCH METHODOLOGY

The study focuses on the population utilizing paid OTT platforms for streaming video content on any device. To conduct the research, a questionnaire was distributed to individuals across all demographics to gather data. The collected data underwent several pre-processing steps to prepare it for machine learning models. The questionnaire comprised 22 questions designed to explore the demographic profile of OTT users and their satisfaction levels regarding different factors influencing churn. All demographic-related questions were multichotomous. Responses to questions concerning churn factors were rated on a 5-point Likert Scale, where one denoted the lowest satisfaction level and five denoted the highest satisfaction level. Among the 336 respondents, 76.02% had subscriptions to multiple OTT platforms. The top three OTT platforms, based on user numbers, were Netflix, Amazon Prime, and Disney Hotstar, with 46.69%, 24.61%, and 14.83% of users, respectively. We will aggregate feature scores from various methods to obtain a more dependable ranking of churn-affecting factors for feature ranking. Hierarchical Logistic Regression in Python will be utilized to determine the impact of having active subscriptions to multiple OTT platforms. Finally, we will employ different classification models in Python and compare their performance.

INPUT DATASET

The dataset comprises 22 variables, specifically 21 independent variables and one dependent variable. The dependent variable, Churn, is binary with two possible values, making our research a binary classification study. The remaining ten predictors are ordinal variables that gauge satisfaction levels regarding factors influencing churn using a 5-point Likert Scale. A score of one signifies the lowest satisfaction level, while a score of five indicates the highest satisfaction level for the corresponding factor. In order to assess the target variable 'Churn,' we converted the 5-point Likert Scale into a binary variable. Scores of one, two, and three on the 5-point Likert Scale represent customers likely to churn, whereas scores of four and five denote customers who are unlikely to leave the platform.

Feature ranking

It is essential to invest time and effort in understanding the factors that impact the outcome variable. Identifying the relevant features not only helps in reducing the number of predictors but also aids in cutting down computational costs and enhancing model performance. To ensure a more dependable and comprehensive factor score, we have evaluated the feature score through four different methods. The final feature score is calculated as the average of scores obtained from all methods. One of the methods employed is Recursive Feature Elimination (RFE), which is an iterative procedure that selects the best or worst performing feature and eliminates it from the feature set. This iterative process continues until all features in the set have been considered. Typically, RFE utilizes models such as SVM to carry out this process.

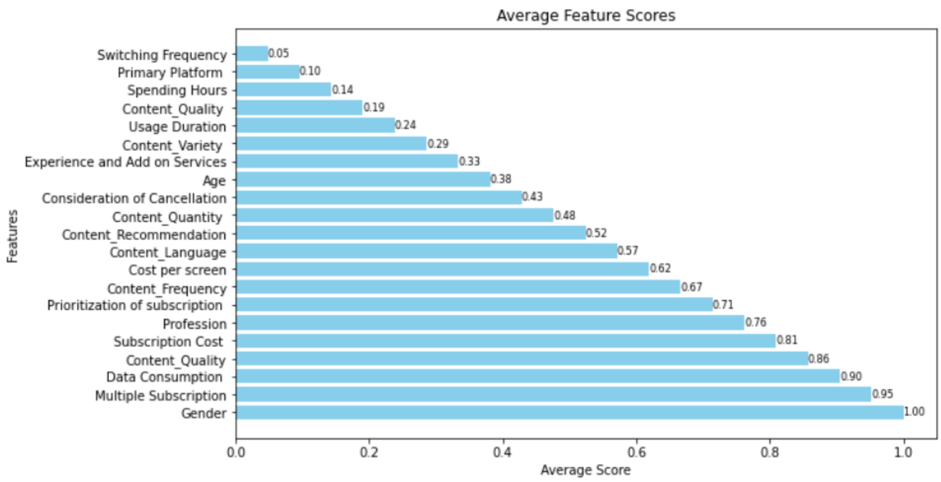

Fig 1 Factor Score

Model implementation

In our study, four distinct models were utilized. The Decision Tree classifier was employed to establish a baseline accuracy, being one of the most commonly utilized models. The remaining three models are ensembles - Random Forest and Ada Boost. Following data preprocessing in our study, the data was divided into two sets for training and testing. 80% of the data was utilized for model training, while 20% was reserved for testing model performance. All churn prediction models developed are binary classification models aimed at predicting customer churn for OTT platforms. Sklearn, a Python library, was utilized for model development.

Decision Tree classifier

The Decision Tree classifier is a popular supervised model that is extensively utilized for classification tasks. It represents all potential solutions to a decision based on specific conditions through a graphical structure. This structure consists of nodes and leaves. At each node, the decision tree poses questions regarding the attributes of the test record. These questions continue based on the answers to the previous questions until the tree determines the class label of the record at the terminal node. By employing a decision tree classifier, the model attained an accuracy rate of 61%. The confusion matrix of the decision-tree prediction model can be found in Table 5.

N=81	Predicted: Churn	Predicted: Not Churn
Actual: Churn	24	22
Actual: Not Churn	12	23

Random Forest classifier

Ensemble methods are machine-learning techniques that integrate various weak algorithms, either of the same kind or different, to create a strong algorithm. This merging results in a model with enhanced performance compared to individual stand-alone models. Random Forest Classifier is an ensemble of decision trees. It randomly selects subsets of the training dataset to train the models individually. Then it performs voting on the results of the individual decision tree to reach the optimal prediction output. By using a random forest classifier, the model achieved an accuracy of 76%. Table 6 provides the confusion matrix of the random forest prediction model.

N=81	Predicted: Churn	Predicted: Not Churn
Actual: Churn	26	20
Actual: Not Churn	8	27

AdaBoost classifier

The AdaBoost classifier is a type of ensemble model that merges several weak algorithms to create a powerful algorithm. It employs a random selection process to choose training samples and iteratively trains the model. The selection of the training set is based on the model's predictions from previous training iterations. Finally, the algorithm assigns weights to these predictions and determines the optimal prediction through voting. By utilizing the AdaBoost classifier, the model attained an accuracy rate of 73%. Table 7 provides the confusion matrix for the AdaBoost prediction model.

N=81	Predicted: Churn	Predicted: Not Churn
Actual: Churn	26	20
Actual: Not Churn	20	23

FINDINGS

In this segment, we shall delve into the outcomes derived from the aforementioned prediction models. The visualization showcases a comparison of the accuracy achieved by the constructed models. Accuracy serves as a metric to comprehend the model's ability to accurately predict both negative and positive classes.

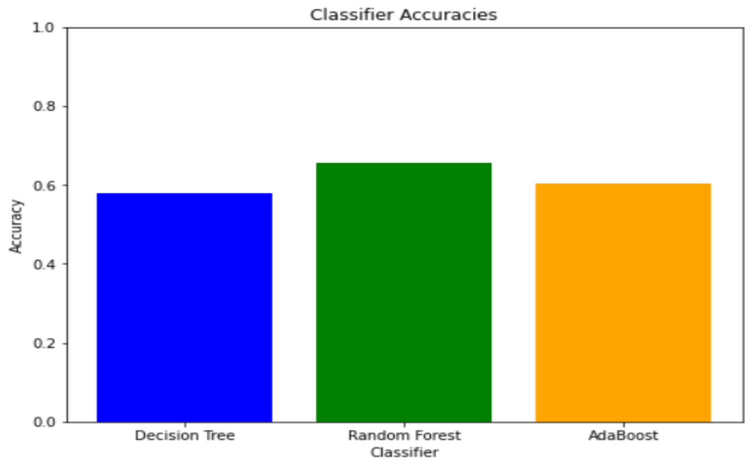

Fig 2 Model Accuracy

It is evident that ensemble models surpass general machine learning models in terms of accuracy, as anticipated. Additionally, Random Forest models outperform other models when considering accuracy scores. However, relying solely on accuracy cannot provide a comprehensive understanding of a model's overall performance.

In the realm of churn prediction, both False Positive and False Negative outcomes have implications for business decisions. In either case, the company may lose a customer if the model fails to identify them as a churn prospect, or allocate resources towards retaining a customer who is unlikely to churn. However, as previously discussed, since the cost of acquiring new customers outweighs that of retaining existing ones, False Negatives have a more significant long-term impact on the business. In the context of churn prediction in OTT platforms, Random Forest and Gradient Boost classifiers exhibit similar accuracy performance. Nevertheless, when considering overall performance metrics, Random Forest proves to be a superior churn predictor.

SUGGESTION

The research work could be further extended into two directions. Firstly, towards improving the model performance by adding social media analysis or adding customer complaints as a new factor.

In the same line, we can use deep learning models for customer churn prediction in OTT platforms. The second would be gauging the impact on customer churns prediction models by using 'Multiple Subscription' as a factor in other domains such as E-Commerce.

Enhance the overall user experience by addressing user satisfaction metrics identified in the study. Offer personalized content recommendations and services based on user preferences and behavior to increase engagement and satisfaction.

Analyze and optimize the primary platform's performance and offerings based on user preferences and feedback. Provide incentives or exclusive content to encourage users to remain loyal to the primary platform.

Implement targeted marketing strategies based on demographic information and usage patterns to retain at-risk customers. Utilize data-driven insights from the study to tailor promotional campaigns and communications aimed at reducing churn.

By implementing these recommendations based on the study's findings and predictive insights, OTT platforms can effectively reduce customer churn, enhance user retention, and ultimately drive sustainable growth in today's competitive market

Conclusion

As previously discussed, customer churn has a significant impact on the company's expenses in maintaining a loyal customer base. Moreover, it also affects the company's reputation within society. By understanding the various factors that contribute to customer churn and accurately predicting it, business owners can proactively make decisions to prevent churn and address the factors that greatly impact customer satisfaction. Our research has successfully identified the key factors that influence customer churn in OTT platforms and effectively predicted which customers are likely to churn based on these factors. To gain a more reliable ranking of these influential features, we utilized four different methods to calculate feature scores and then aggregated them using the mean. Additionally, we have determined that the most critical factors influencing churn are customers frequently switching between multiple OTT platforms and having multiple subscriptions. Furthermore, reducing the cost per screen and enhancing the availability of content in multiple languages are key areas where OTT companies can directly focus on to mitigate churn.