10.31341/jios.46.2.10 UDC 004.85:621.397:004.738.5-047.37
Open Access Original Scientific Paper

Predicting Customer Churn on OTT Platforms: Customers with Subscription of Multiple Service Providers

Manish Mohan *pv.manish8@gmail.com*
Symbiosis Centre for Information Technology, Pune, India

Anil Jadhav *anil@scit.edu*
Symbiosis Centre for Information Technology, Pune, India

Abstract

No industry can thrive without customers and with customers comes the chances of customer churn. Since customer churn have direct-impact on the revenue, all the industries are focusing in understanding the factors influencing churn and are developing methods to predict the customer churn effectively. Today, never as before, customers have wide variety of options to choose between any service or product. In addition, nowadays customers enjoy multiple subscriptions of service providers across sectors. In this study we aim to identify: i) Factors influencing customer churn on OTT platform, and ii) Predict customer churn on OTT platform. The data for this study is collected from 317 respondents, using questionnaire method, who have multiple OTT platform subscription. The questionnaire data contains 19 items which includes demographic features, usage of OTG platform, and user contentment factors about OTT service. We have identified factors influencing customer churn in Over-The-Top (OTT) platform by combining Recursive Feature Elimination (RFE), Linear Regression, and Ridge Regression feature ranking methods. We have used Hierarchical Logistic Regression, to understand impact of two newly introduced factors namely 'Multiple Subscription' and 'Switching Frequency' on the overall performance of the customer churn prediction. Finally, customer churn prediction is done using Decision Tree, Random Forest, AdaBoost, and Gradient boosting techniques. We found that random forest method gives better prediction results.

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

1. Introduction

Customers are the heart and soul of any organization. In today's competitive market, customer satisfaction carries more weight than ever before. How a customer feels, not only while merely using the product but being part of the brand itself, is one of the most crucial factors determining how a company will thrive in today's business world.

There are two ways an organization can increase or maintain its customer base, either acquire new customers or retain existing ones. Empirical studies have shown that cost of acquiring new customers is five times that of retaining a customer. The research makes the latter a better solution for increasing the overall profit. Apart from profit, retention has positive social effects that give an edge in today's competitive market. Because of this, customer retention becomes an obvious choice of stakeholders to increase the overall profit.

The research done [1] gives us a clear picture of the customer's life cycle, the steps involved in acquiring a new customer, and retaining an existing customer. It depicts that the stages of acquiring a new customer are more, implying investing a more significant amount of time and resources.

Industry dynamics of Over-The-Top (OTT) platforms, which initially had a monopolistic market, have changed in recent years. The change is mainly because of moguls of different sectors diversifying in the OTT market. The increase in the competition gave rise to a fight to retain the customer base, where a better understanding of customer emotions and factors inducing churning ensures winning.

Considering the kind of data generated by OTT platforms, Machine Learning (ML) stands as a sophisticated way to get insights and facilitate business decisions. OTT giants are taking the help of different ML techniques such as Classification models, predictive models, Clustering Algorithms, Neural Networks, and others to stand out in the market. Correctly implementing these methods helps the organization intervene at the appropriate time and act before the customer leaves their platform. In addition to customer retention, churn prediction is helpful from other aspects, like revenue prediction and improving customer service.

The following paper consists of seven sections. The first section will discuss the existing literature, followed by the research objective. The third part will talk about the research methodology used in the paper. The fourth and fifth parts will cover the implementation of the research objective, followed by result discussions and a conclusion.

In the existing literature of churn prediction model, the work revolves around Telecommunication, Finance, and Retail and E-commerce sectors. We did not find any extensive, robust, and reliable work done concerning OTT platforms. In addition, customers taking a subscription of multiple service providers is a factor that came into dominance, like never before, because of the increase in the number of service providers. The previous literature has not considered this factor for building churn models. These research works consider this factor, along with others, while building the predictive models.

Moreover, most of the work done around the churn prediction is basis the secondary data. Research based on secondary data has been a reactive approach, giving less time for any action to retain the customer. To make the approach proactive, we use primary data in our study.

This paper identifies the features that strongly influence customer churn concerning OTT platforms. In addition, we compare the performance of baseline models with multiple ensemble binary classification models.

While a few years back, we had limited options when it came to OTT platforms, the situation changed drastically in the last few years. The covid-19 pandemic gave the final thrust required for all the big players to jump into the market.

With the increase in service providers, the churn rate also increased. A study of Statista shows that 77% of people had Netflix's subscription in the USA, and 56% have Amazon accounts, two of the giants in the OTT market. The numbers make it evident that people are enjoying multiple subscriptions nowadays.

The research objective is to analyze the data of OTT (Over-The-Top) platform users to understand the customer preferences, factors affecting customer loyalty, and factors promoting customer churn for their primary OTT platform.

The objective of this study is:

1. To study the factors relevant to customer churn for OTT platforms.
    a) Feature ranking of the factors influencing customer churn for OTT platforms.
    b) Gauge the impact of having multiple subscriptions of OTT platforms on customer churn prediction.
2. Accurately predicting the customers who might leave the OTT platforms shortly, using a classification model.

2. 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. 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. [2], [3], [4] 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. [5] talk about optimizing the model by answering the question - 'How long is long enough?' This paper talks about time window optimization for improving the performance of Logistic Regression and Classification Trees algorithms. [6] Compares the performance of Fisher's discriminant equations and logistic regression and concludes that logistic regression performs better with an accuracy of 93.94% in the churn prediction model for telecom companies.

To improve the model performance and reliability, researchers have tried various ensembles and hybrid ML models that work on the concept of information fusion. [7] Propose and evaluate different ensemble models by combining clustering and classification techniques. Of various ensembles, the combination of k-med clustering and Gradient boosting, Decision Tree, and Deep Learning classifier ensemble gives the best prediction on two telecommunication datasets. [8] Studies various supervised learning algorithms with similar evaluation setup and same validation technique, k-fold cross-validation. The comparison revealed that random forest outperforms decision trees, k-nearest neighbors, elastic net, logistic regression, and support vector machines. Moreover, Random Forest performs better than the ensemble of the above classifiers. [9] [10] Random Forest and Boosting algorithms are examples of ensembles used in the same lines. Studies [11] also discuss optimizing ensembles methods and explore a one-step dynamic classifier model that fuses a preprocessing step of dealing with missing value with multiclass ensembles. Later, the author concludes with the outperformance of the one-step model over the traditional two-step classification models. [12], [13] have discussed the implementation of hybrid models. On the one hand, the former talks about the improved top decile lift by implementing hybrid-clustering models; the latter builds a hybrid classification model with 20 features that could achieve accuracy greater than 85%. Implementing hybrid models to improve prediction does not confine to general ML classification and clustering algorithms. [14] proposes a hybrid model made by Feedforward Neural Network and Particle Swarm Optimization. In the proposed model, Particle Swarm Optimization tunes the weight and improves the structure of the neural network simultaneously, resulting in improved prediction scores. Along with predicting customer churn, using classification and clustering techniques, [15] recognize the reason for customer churn. The author implements information gain, fuzzy particle swarm optimization, and divergence kernel-based support vector machine for classification. The model gives 94.11% and 95.41% accuracy for two different data sets.

Researchers have also presented rule-based algorithms that identify the relationship between different variables as an effective method of predicting customer churn. [16] Researchers have studied to generate different rules generation algorithms on different datasets. [17] take it a step further by defining customer behavior attributes for the prediction model.

Various authors [18], [19] have depicted the implementation of Deep Neural Networks for customer churn prediction. [19] Comparison of performance Deep Q Neural Network and other data mining techniques shows that Deep Q Neural Network surpasses general machine learning models performance. [20] Implements the Deep-BP-ANN model and achieved 88.12% and 79.38% accuracy for two different data sets. The author used two feature selection methods; Variance Thresholding and Lasso Regression. Moreover, to counter overfitting, early stopping criteria were used. The model performance across metrics were better than other ML techniques implemented; XG_Boost, Logistic_Regression, Naïve_Bayes, and KNN. [21] Set the side-by-side effects of various monotonic activation functions, batch sizes, and optimizers on the performance of the neural network model. The author found that applying the Relu function in a neural network's hidden layer gives better performance. However, performance dropped as the batch size reached closer to the test data set. RemsProp optimizer outperforms the stochastic gradient descent Adadelta algorithm, the Adam algorithm, the AdaGrad algorithm, and the AdaMax algorithm. [22] Compares Artificial Neural Network with Machine Learning algorithms - Support Vector Machine, Gaussian Naïve Bayes, Decision Tree, and K-Nearest Neighbor; over accuracy and F-score and recommends artificial neural network and Gaussian Naïve Bayes as the most appropriate algorithm to predict customer churn in the telecom industry.

[23], [24] Models based on Negative Correlation Learning (NCO) for improving the performance of churn prediction models is another effective way to predict customer churn. [23] Train an ensemble of Multilayered Perceptron using NCO and depict the model's outperformance compared to common data mining and ML models. In the same lines, [24] incorporates NCO ensemble models and concludes that customer retention rate is higher in Atom Search Optimization and Particle Swarm Optimization approach.

Apart from improvising algorithms and introducing new factors, a way to improve the model performance is improvising data preprocessing techniques. Imbalance Data is always a challenge for any Data Mining or prediction model. [25], [26], [27] Research extensively comparing various methods of dealing with data imbalance with in-depth exploration is available in the literature. [28] have effectively compared six different sampling techniques; majority weighed minority-oversampling technique, couples top-N reverse k-nearest neighbor, adaptive synthetic sampling approach, synthetic minority oversampling technique, immune centroid oversampling technique, and mega-trend diffusion function. The author implemented these six data balancing techniques on four different data sets and built four rule generation algorithms. The author ceases the discussion with the conclusion that the mega-trend diffusion function and rules generation based on genetic algorithms surpass all other models' performance.

Another preprocessing step that helps in improving the model performance is Feature Engineering. Feature engineering is a method used to determine the factors that represent the entire data set better and then give those features input to the model instead of the entire raw data. Many authors have [9], [29] performed feature engineering before feeding the data to the predictive models. By doing so, they improved the model performance by a significant margin. [29] depicted an improved accuracy, precision, and recall of XGBoost to 99.41%, 99.44%, and 99.94%, respectively, by combining feature engineering. In the same lines, authors [25] identified 18 relevant predictor variables among 75 predictors and provided them to the deep neural network model for efficient customer churn prediction. Researchers, to refine the model, combine ensemble models with feature engineering. [30] Predicts customer churn in banking domain by implementing Meta classifier algorithm with an adaptive genetic algorithm for feature selection. Feature selection is done using DragonFly and Firefly algorithms, and then the XGBOOST classifier is implemented. Along the same lines [31] use stacking and soft voting models to predict customer churn. Firstly, a stoking model is built using Xgboost, Logistic regression, Decision tree, and Naïve Bayes algorithms. Further, the outputs of the second level are given for soft voting. With this technique, the author can get high accuracy of 96.12% and 98.09% for the original and new churn datasets.

Although optimizing algorithms and improvising preprocessing helps improve the model performance, researchers have worked on different ways of adding new features influencing churn to yield the desired performance. [32] discussed that customer churns are not a mere statistical phenomenon but occurrences whereby social factors play roles. The author successfully builds a model with social factors with accuracy as high as 91.44%. In the same lines, authors [33] refines adding social aspects in the churn prediction model by using the 'The- group first social network' approach. They build models for predicting the social groups at high risk of churning, even though none of the members in the social group has churned until time.

Similarly, research has identified [34] the impact of yet another factor – geographical factors, on customer churn of an Insurance company. The authors demonstrate that the probability of customer churn is associated with the proximity of the customers with respect to the branch office. The churning probability of customers closer to the branch office is lower than customers away from the office. Similarly, the customers in closer proximity to their competitor's office branches are more likely to be churned.

In the era of social media, the ability to perform analysis on social media content gives an edge to companies over competitors. Authors [35] used user-generated content (UGC) to build the customer churn model and have made performance comparisons with general ML models and Deep Learning models. The UGC model considers comments, posts, messages, and product reviews and segregates them into positive and negative text polarity using sentiment analysis.

In consonance with the early research done about exploring new features to make the customer churn prediction model more effective and robust, the effectiveness of lower and upper sample distance [36] was still unexplored. The investigation shows that lower distance test data sets achieve better performance in multiple performance measures – accuracy, f-score, precision, and recall.

In addition, even in an era where data is abundant, there are situations when a particular company does not have sufficient data to predict the customer churn in the organization. The cross-company churn prediction model comes in handy to tackle this problem statement [36]. The research extensively compares multiple digital transformation techniques on the cross-company churn prediction model.

Customer retention, improved customer satisfaction, and an improved social stand of a company are some of the benefits of bringing in a customer churn prediction model. However, the sole business motive is always profit maximization. Though most models help achieve the goal, it is usually more inclined towards model performance. In concurrence to this, many researchers have extensively discussed the implementation of data mining techniques keeping profit maximization as the prime objective. While most of the research assumes the same customer lifetime value for all the customers, various models [37] take variability in customer-life time value into consideration with the goal of profit maximization. This research brings the prediction model closer to situations that resemble real-world situations. In the same direction, other researchers [38] aligned their research towards the core business requirement of profit maximization. The authors consider the misclassification cost and present a new classifier that integrates the expected maximum profit measure for customer churn with classifier model construction. This model, named 'ProfTree,' achieves significant improvement in profit as compared to accuracy-driven tree-based methods.

Analogous to the above researchers [39], instead of traditional error-based classification algorithms, the author focuses on improving the classifier's accuracy over cost sensitization. AdaBoostWithCost a cost-sensitive boosting algorithm, is proposed to reduce the churn cost. AdaBoost with cost applies the misclassification cost more specifically to the costly high-risk errors instead of directly applying a constant cost to all misclassification errors in each iteration of boosting. This algorithm, by reducing false-negative errors, outperforms the discrete AdaBoost algorithm. The model successfully consistently decreases the total misclassification error, false-negative error count, and training and testing error rates by 10, 20, and 40, respectively, for each set of boosting rounds.

This paper contributes to the literature of predicting customers by bringing in new unexplored factors in the industry that is still a newbie compared to other traditional industries that have existed in the market for decades.

3. Research methodology

The study focuses on the population using paid OTT platforms to stream video content on any device. For the research, considering people across all the demographics, the questionnaire was distributed to collect the data, applied various pre-processing steps on the data received to make it viable for machine learning models.

The questionnaire consisted of 19 questions formulated to understand the demographic profile of the OTT users and their contentment level concerning different factors affecting churn. All the demographic-related questions were multichotomous. The response to questions related to factors affecting churn was on 5- point Likert Scale, where one indicated the lowest level of contentment and five indicated the highest level of contentment.

Out of the 317 respondents, 76.02% have multiple OTT platform subscriptions. The top three OTT platforms, with respect to the number of users, were Netflix, Amazon Prime, and Disney Hotstar, with 46.69%, 24.61%, and 14.83% users, respectively.

We will be combing feature scores of various methods to get a more reliable ranking of the factors affecting churn for feature ranking. We are implementing Hierarchical Logistic Regression in SPSS to identify the impact of having an active subscription of multiple OTT platforms.

Lastly, we will be implementing various classification models on Python and comparing their performance.

4. Input data set

The data collected consist of 19 variables, i.e., 18 independent and one dependent variable. The dependent variable - Churn, takes two values, implying that our study is a binary classification study. Table.1 gives us the details of all the variables that are in the study:

Serial No.	Attribute Details		
	Attribute	***Data Type***	***Description***
1	Name	Categorical	Name of the respondent
2	Gender	Categorical	Gender of the respondent
3	Age	Categorical	Age of the respondent (In Years)
4	Profession	Categorical	Profession of the respondent
5	Usage Duration	Categorical	How long the respondent have been using OTT platforms
6	Multiple Subscription	Categorical	Does the respondent have subscription of multiple OTT Platforms?
7	Switching Frequency	Categorical	If Yes, how frequently does the respondent switch between the Platforms?
8	Primary Platform	Categorical	Primary OTT Platform of the respondent
9	Subscription Cost	Ordinal	Cost Of Subscription of primary platform
10	Cost per screen	Ordinal	Cost per screen in primary platform
11	Data Consumption	Ordinal	Average data consumption in primary platform
12	Content_Varity	Ordinal	Varity of Content Available in primary platform (Availability of content of various Genre)
13	Content_Language	Ordinal	Availability of content in different languages in primary platform (International, National and Regional)
14	Content_Quantity	Ordinal	Quantity of content available in primary platform
15	Content_Quality	Ordinal	Quality of content available in primary platform
16	Content_Frequency	Ordinal	Frequency of release of new content on primary platform
17	Experience and Add -on Services	Ordinal	Platform Experience and Add -on Services of primary platform
18	Content_Recommendation	Ordinal	Closeness of recommended content on primary platform
19	Churn	Ordinal	Plan of changing the primary OTT platform

Table 1. Data Set Attributes

4.1. Data pre-processing

Out of the 19 variables, excluded name variable as it does not add value to the analysis. Seven out of the 17 predictors, Gender, Age, Profession, Usage Duration, Multiple Subscription, Switching Frequency, and Primary Platform, are categorical variables.

The remaining ten predictors are ordinal variables that measure the level of contentment for factors affecting churn on the 5- point Likert Scale. One indicates the lowest level of contentment, and five indicates the highest level of contentment for the respective factor.

To measure the target variable 'Churn,' converted the five-point Likert Scale to a binary variable. One, Two, and Three values of 5- point Likert Scale indicate the customers who will churn, and values four and five are classified as customers who will not leave the platform. We have excluded Twenty-three responses out of 317 from the analysis because of high noise.

We have plotted a correlation matrix to understand how a variable responds to changes in other corresponding variables. The correlation matrix also helps understand features with strong and weak dependencies. Fig. 1 shows the correlation matrix. Dark blue color represents strong correlation, and light color shows weak correlation. We will consider any factors with a correlation coefficient greater than positive 0.7 or less than negative 0.7 as extreme correlation and define further steps to deal with it.

In the factors we have considered, the highest positive correlation is 0.63 between 'Multiple Subscription' and 'Switching Frequency,' whereas 'Age' shows the highest negative correlation, -0.11, with both 'Content Frequency' and 'Content Recommendation.'

5. Understanding churn factors

This section of the paper will discuss our first objective. Firstly, we will discuss the ranking of the factors that influence churn in OTT platforms, followed by a discussion on the impact of users having multiple subscriptions on the customer churn prediction.

6. Feature ranking

Understanding the features influencing the outcome variable is indeed a task worth investing time and energy in. Understanding the relevant features will help us reduce the number of predictors but also helps in reducing the computational cost and improving the model performance.

In order to get a more reliable and generalized factor score, we have measured the feature score using four methods. The final feature score is the average of the scores of all the methods.

The first method is Recursive Feature Elimination (RFE). RFE is an iterative process that selects the best or worst performing feature them excludes it from the feature set. The iterative process continues until all the features from the set are exhausted. Generally, RFE uses models like SVM to perform the process.

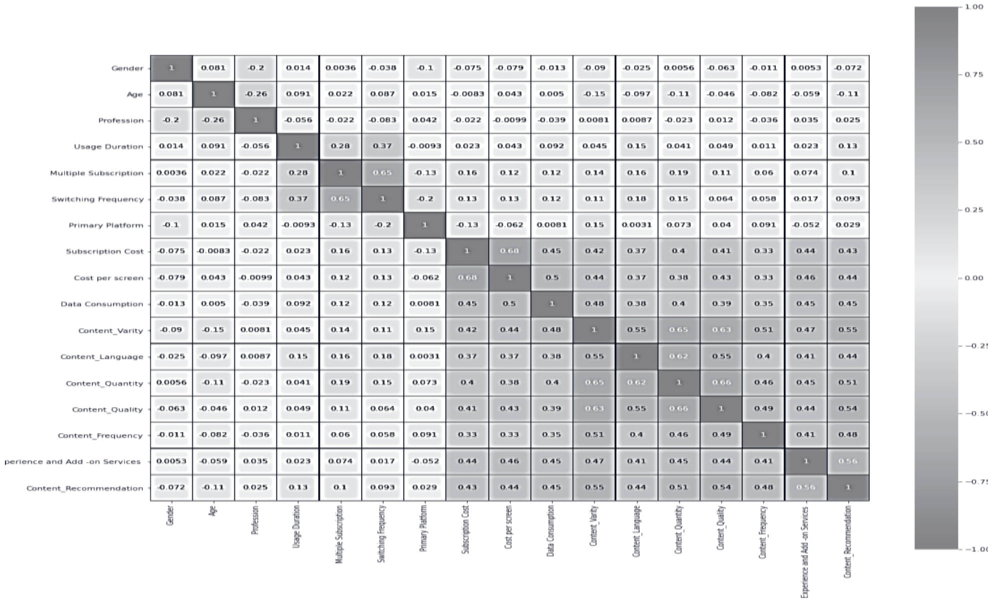

Figure 1. Correlation Heat Map

In the second and third methods, we used linear models - Linear Regression and Ridge Regression. Via these methods, we collected the coefficients for each feature to select and prioritize the features. In the final method, we used the inbuilt feature ranking function of Sklearn's Random Forest model known as 'feature importance.' In Fig. 2, we have visualized the all the features as per their rank using bar chat.

As we can see in the bar graph, the most relevant feature for predicting churn in OTT platforms are 'Switching Frequency' and 'Multiple Subscription.' Whereas 'Experience and Add-on Services' and 'Content Quality' have the most negligible impact on the model. As discussed earlier, both the features, 'Switching Frequency' and 'Multiple Subscription,' are newly introduced factors that came into dominance because of the recent changes in industry dynamics.

7. Impact of multiple subscription

This section of the paper discusses the influence of two newly introduced factors, 'Multiple Subscription' and 'Switching Frequency,' on the overall performance of the customer churn prediction models. Since the task is a binary classification problem, we have used' Hierarchical Logistic Regression to gauge the impact of these two variables.'

The principle that governs logistic regression is the natural logarithm of odds ratio given as:

$$logit(p) = \log\left(\frac{p}{1-p}\right)$$

Where p is the probability and $\frac{p}{1-p}$ denotes the corresponding odds.

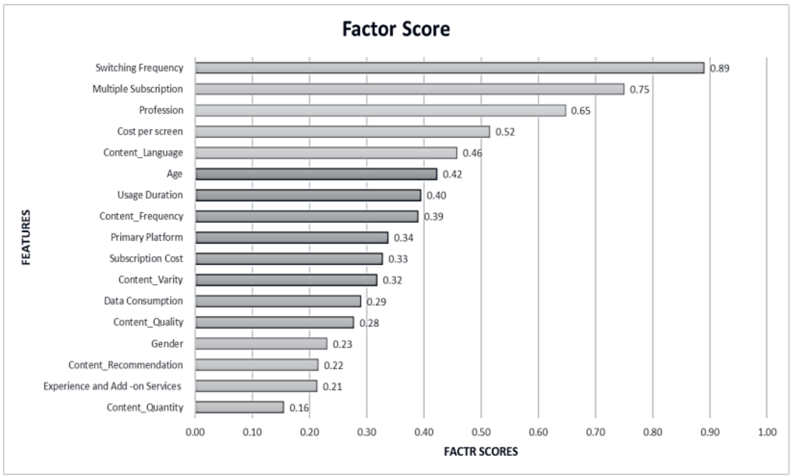

Figure 2. Feature Ranking

We have used SPSS for performing hierarchical regression analysis. In this research paper, we have fitted a two-block logistic model to the data. With the churn variable in the dependent variable section, the first block measures the performance of logistic regression classification using all the predictor variables except 'Multiple Subscription' and 'Switching Frequency.'

We added 'Multiple Subscription' and 'Switching Frequency' in block two to estimate the improvement in model classification. To gauge the classification task's improvement and understand the significance and reliability of the model, we will discuss the classification table along with the Omnibus Test of Model Coefficient and Hosmer and Lemeshow Test to check the goodness of fit.

BLOCK 1					**BLOCK 2**				
		Predicted Churn					Predicted Churn		
		Churned	Not Churned	Percentage Correct			Churned	Not Churned	Percentage Correct
Actual Churn	Churned	188	11	94.5	Actual Churn	Churned	181	19	90.5
	Not Churned	82	13	13.7		Not Churned	64	30	31.9
Overall Percentage				68.4	Overall Percentage				71.8

Table 2. Classification Table

Tab. 2 compared the classification performance of the two models. We can observe that by adding 'Multiple Subscription' and 'Switching Frequency,' we improved the model performance by 3.4%.

Omnibus tests of model coefficients help us in defining the significance of the model built. It uses the chi-square test to check the improvement in the model performance over the baseline model. Tab. 3 shows the Omnibus tests of model coefficients for our model. It shows that the model is significant at $\chi^2 = 34.485$ with df = 16 (p-value = 0.005).

Chi-square		df	Sig.
Step	12.624	1	0.000
Block	12.624	1	0.000
Model	34.485	16	0.005

Table 3. Omnibus Tests of Model Coefficients

In order to understand the goodness of fit of the model, we have considered Hosmer and Lemeshow test. The test returns the chi-square value and p-value, which helps in understanding the model fit. Here, a small p-value indicates a poor fit model. Tab 4 depicts the output of the Hosmer and Lemeshow test for our model. For the model built, it is significant at $\chi^2 = 9.012$ (df = 8, p-value 0.341). The high p-value indicated that our model good fit.

Step	Chi-square	df	Sig.
1	9.012	8	0.341

Table 4. Hosmer and Lemeshow Test

8. Model implementation

In this research, we have implemented four different models. We used the Decision Tree classifier to get a baseline accuracy, one of the most widely used models. The rest three models are ensembles - Random Forest, Ada Boost, and Gradient Boost.

In our research work, after preprocessing, we split the data into two sets for training and testing purposes. We have used 80% of the data to train our model and 20% to test the model performance.

All our churn prediction models are binary classification models predicting customer churn for OTT platforms. To build the models, Sklearn, a Python library, is used.

8.1. Decision Tree classifier

Decision Tree classifier, a type of supervised model, is one of the most widely used classification algorithms. The decision tree is a graphical representation of all the possible solutions to a decision based on certain conditions. The tree has nodes and leaves. At every node, the decision tree carefully formulates questions on the attributes of the test record. Questions follow the answer to the previous question until the tree concludes the class label of the record on the terminal node.

Using a decision tree classifier, the model achieved an accuracy of 61%. Tab. 5 gives us the confusion matrix of the decision-tree prediction model.

n = 59	Predicted: Churn	Predicted: Not Churn
Actual: Churn	24	12
Actual: Not Churn	11	12

Table 5. Decision Tree Confusion Matrix

8.2. Random Forest classifier

Ensemble methods are machine-learning techniques that combine various weak algorithms, either of the same kind or different, to form a strong algorithm. This combination results in a model with enhanced performance as 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.

Using a random forest classifier, the model achieved an accuracy of 76%. Tab. 6 gives us the confusion matrix of the random forest prediction model.

n = 59	Predicted: Churn	Predicted: Not Churn
Actual: Churn	35	1
Actual: Not Churn	13	10

Table 6. Random Forest Confusion Matrix

8.3. AdaBoost classifier

AdaBoost is also an ensemble model that combines multiple weak algorithms to come up with a strong algorithm. AdaBoost randomly selects training samples and iteratively trains the model. Adaboost selects the training set based on model predictions of previous training. Lastly, the algorithm assigns weights to the predictions and outputs the optimal prediction through voting.

Using the AdaBoost classifier, the model achieved an accuracy of 73%. Tab. 7 gives us the confusion matrix of the AdaBoost prediction model.

n = 59	Predicted: Churn	Predicted: Not Churn
Actual: Churn	30	6
Actual: Not Churn	10	13

Table 7. Adaboost Confusion Matrix

8.4. Gradient Boost classifier

Gradient Boost is yet another ensemble-boosting model that works sequentially. In the first step, Gradient Boost builds a weak model. Then it uses the exponential loss function to calculate the loss function for the weak model previously made. The goal of the algorithm is to reduce the loss function in order to increase the accuracy. Until the model reaches a certain threshold, it repeats the steps.

Using the Gradient Boosting classifier, the model achieved an accuracy of 76%. Tab. 8 gives us the confusion matrix of the Gradient Boosting prediction model.

n = 59	Predicted: Churn	Predicted: Not Churn
Actual: Churn	33	3
Actual: Not Churn	11	12

Table 8. Gradient Boosting Confusion Matrix

9. Results and discussion

In this section, we will discuss the results obtained from the prediction models modeled above. Fig. 3 visualizes the comparison of accuracy for the models built. Accuracy helps us understand how accurately the model can predict the actual negative and positive classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

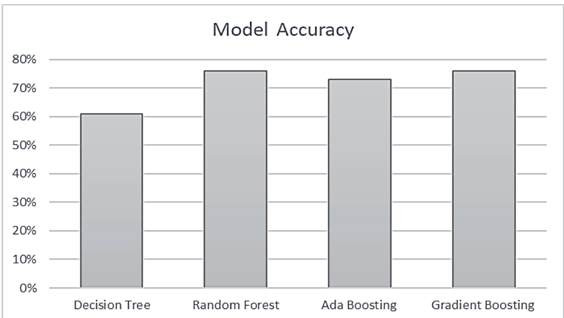

Figure 3. Model Accuracy

It is evident that, as expected, ensemble models accuracy is better than the general machine learning model. In addition, Random Forest and Gradient Boosting come out to better performing models considering the accuracy scores.

Accuracy, though it gives us a bird' eye view of the model's performance, alone cannot tell us about the overall performance. In order to understand the overall performance of the models, metrics that would be discussed are:

- Precision: This metric helps us in determining the reliability of the model. With respect to churn prediction, it tells us how many customers whom the model predicted as churned belong to the churn class.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: Also known as true positive rate or sensitivity. Recall talks about the numbers of actual churned cases that our model correctly classified.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score: By the nature of the formula, if we try to improve the precision, recall reduces. Since both the metrics give an idea of the model performance, F1-Score gives us a combined idea about both the metrics. F1-Score is the Harmonic mean of both these matrices.

$$F1 - Score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

Tab.9 summarises all these matrices for all our models. Fig. 4 helps us in visual comparison of Precision, Recall and F1-Score.

Model	Accuracy	Precision	Recall	F1
Decision Tree	61.02%	68.57%	66.67%	67.61%
Random Forest	76.27%	72.92%	97.22%	83.33%
AdaBoost	72.88%	75.00%	83.33%	78.95%
Gradient Boost	76.27%	75.00%	91.67%	82.50%

Table 9. Overall Performance

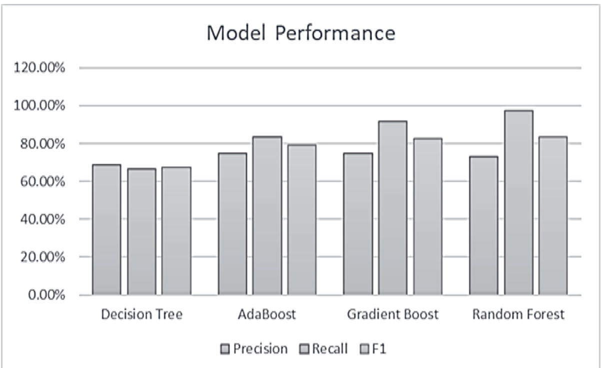

Figure 4. Overall Performance

In churn prediction, both False Positive and False Negative have their share of impact on the business decision. In both cases, the company either would lose a customer, as the model never predicted him as a churn prospect, or would end up spending on customer retention of a customer who is not a churn prospect. However, as discussed earlier, since the cost attached to customer acquisition is always more significant than the cost of customer retention, False Negatives will have a more significant business impact in the long run.

For churn-prediction in OTT platforms, though, Random Forest and Gradient Boost classifiers perform equally well in accuracy scale, considering overall performance matrices makes Random Forest a better churn predictor.

10. Conclusion

As discussed, customer churn increases the cost to the company considering keeping the customer base intact. In addition, it affects organizations' societal stand. Understanding the factors influencing customer churn and predicting customer churn helps the business owners make the business decision beforehand that would resist churn and work on the factors that are having a maximum influence on customer satisfaction.

Our research has identified the critical factors influencing customer churn in OTT platforms and accurately predicted the customers who might get churned basis these factors. For understanding the essential features influencing customer churn in the OTT platform and get to a more reliable feature ranking score, we calculated feature scores using four different methods and aggregated the scores using mean. We also concluded that the most critical factors influencing churn are customers frequently switching between multiple OTT platforms and having multiple subscriptions. Apart from this, the factors that highly influence churn and OTT companies could directly work upon is reducing cost per screen and improving the availability of contents of multiple languages.

As discussed earlier, with the increase in the number of service providers, a new factor that is users taking multiple subscriptions came into the picture. Adding factor related to this as a feature helps us improve the model performance of predictive classifiers by 3.4%.

To achieve our second and final objective of accurately predicting customer churn, we modeled four predictive classifiers. Since accuracy cannot solely judge overall models' performance, we looked at other performance matrices. We inferred, in the end, that Random Forest – an ensemble classifier would be more efficient than Decision Tree, Gradient Boosting, and AdaBoost classifiers for predicting customer churn on OTT platforms.

10.1. Future scope

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.

References

[1] O. Sigurdur, L. Xiaonan and W. Shuning, "Operations research and data mining," European Journal of Operational Research, 2006.

[2] T. Chih-Fong and L. Yu-Hsin, "Data Mining Techniques in Customer Churn Prediction," Recent Patents on Computer Science, pp. 28-32, 2009.

[3] S. Hergovind and V. S. Harsh, "A Business Intelligence Perspective for Churn Management," Procedia Social And Behavioral Sciences, p. 51 – 56, 2014.

[4] H. Benlan, S. Yong, W. Qian and Z. Xi, "Prediction of customer attrition of commercial banks based on SVM," Procedia Computer Science, p. 423 – 430, 2014.

[5] B. Michel and d. P. DirkVan, "Customer event history for churn prediction: How long is long enough?," Expert Systems with Applications, pp. 13517-13522, 2012.

[6] Z. Tianyuan, M. Sérgio and F. R. Ricardo, "A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation," Future Internet , 2022.

[7] F. B. Syed, A. A. Abdulwahab, B. Saba, H. K. Farhan and A. A. Abdulaleem, "An ensemble based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry," PeerJ Computer Science, 2022.

[8] A. d. L. L. Renato, C. S. Thiago and M. T. Benjamin, "Propension to customer churn in a financial institution: a machine learning approach," Neural Computing and Applications, 2022.

[9] K. A. Abdelrahim, J. Assef and A. Kadan, "Customer churn prediction in telecom using machine learning in big data platform," Journel of Big Data, pp. 1-24, 2019.

[10] B. R. J. and C. P. S., "An Optimal Ensemble Classification for Predicting Churn in Telecommunication," Journel Of Engineering Science and Technology Review, pp. 44 - 49, 2020.

[11] X. Jin, Z. Bing, T. Geer, H. Changzheng and L. Dunhu, "One-Step Dynamic Classifier Ensemble Model for," Mathematical Problems in Engineering, 2014.

[12] B. Indranil and C. Xi, "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn," in Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, 2009.

[13] L. Xueling and L. Zhen, "Hybrid Prediction Model for E-Commerce Customer Churn Based on Logistic Regression and Extreme Gradient Boosting Algorithm," Ingénierie des Systèmes d'Information, pp. 525-530, 2019.

[14] F. Hossam, "A Hybrid Swarm Intelligent Neural Network Model," Information, 2018.

[15] P. C. K. and S. B. L., "Fuzzy particle swarm optimization (FPSO) based feature selection and hybrid kernel distance based possibilistic fuzzy local information C means (HKD PFLICM) clustering for churn prediction in telecom industry," SN Applied Sciences, 2021.

[16] A. Adnan, A. SAJID, A. AWAIS, N. MUHAMMAD, H. NEWTON, Q. JUNAID, H. AHMAD and H. AMIR, "Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study," IEEE Acces, pp. 7940-7957, 2016.

[17] M. Ibrahim and I. B. E. M. B. Ahmed, "Customer churn prediction model using data mining techniques," in 13th International Computer Engineering Conference, IEEE, 2017.

[18] U. V. and I. K., "Automated Feature Selection and Churn Prediction using Deep Learning Models," International Research Journal of Engineering and Technology (IRJET), pp. 1846-1854, 2017.

[19] P. M and L. Y, "Applying Reinforcement Learning for Customer Churn Prediction," in 13th International Conference on Computer and Electrical Engineering, Beijing, 2020.

[20] W. F. Samah, S. Suresh and A. K. Moaiad, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," Information Sciences Letters, pp. 185-198, 2022.

[21] D. Anouar, "Impact of Hyperparameters on Deep Learning Model for Customer Churn Prediction in Telecommunication Sector," Hindawi, vol. 2022, 2022.

[22] M. Moh, B. Arif, J. A. Hasan, A. Sami and A. Ryan, "Classification methods comparison for customer churn prediction in the telecommunication industry," International Journal of Advanced and Applied Sciences, pp. 1-8, 2021.

[23] R. Ali, F. Ayham, F. Hossam, A. Jamal and A.-K. Omar, "Negative Correlation Learning for Customer Churn Prediction:," The Scientific World Journal, 2015.

[24] M. R., S. R. and S. S., "An Effective Architectural Model for Early Churn Prediction – NELCO," International Journal of Engineering and Advanced Technology (IJEAT), pp. 4667-4672, 2019.

[25] B. J. and V. d. P. D., "Handling class imbalance in customer churn prediction," Expert Systems with Applications, p. 4626–4636, 2009.

[26] Z. Bing, B. Bart and K. v. B. Seppe, "An empirical comparison of techniques for the class imbalance problem in churn prediction," Information Sciences, pp. 84-99, 2017.

[27] B. Zhu, B. Baesens, A. Backiel, B. vanden and K. L. M. Seppe, "Benchmarking sampling techniques for imbalance learning in churn prediction," Journal of the Operational Research Society, 2018.

[28] A. Adnan, A.-O. Feras, S. Babar, A. Awais, L. Jonathan and A. Sajid, "Customer churn prediction in telecommunication industry under uncertain situation," Journal of Business Research, pp. 290-301, 2019.

[29] S. P and R. B. Dayananda, "Improvised_XgBoost Machine learning Algorithm for Customer Churn Prediction," EAI Endorsed Transactions on Energy Web, 2020.

[30] S. B., S. L. V. P. S. Gutta, I. D. N. V. S. L. S., R. K. S. R. and S. Khasim, "Adaptive XGBOOST Hyper Tuned Meta Classifier for Prediction of Churn Customers," Tech Science Press, vol. 33, pp. 22-34, 2022.

[31] T. Xu, Y. Ma and K. Kim, "Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping," Applied Sciences, 2021.

[32] M. Jelena and G. Jamil, "Customer Churn Prediction in Mobile Operator Using Combined Model," in ICEIS 2014 - 16th International Conference on Enterprise Information System, 2014.

[33] R. Yossi, Y.-T. Elad and S. Noam, "Predicting Customer Churn in Mobile Networks through Analysis of Social," in Proceedings of the 2010 SIAM International Conference on Data Mining (SDM), Colombus, 2010.

[34] M. Á. De la Llave, F. A. López and A. Angulo, "The impact of geographical factors on churn prediction: An application to an insurance company in Madrid’s urban area," Scandinavian Actuarial Journal, p. 2017, 188-203.

[35] e. K. Essam Abou, M. A. Alaa, A. H. Shereen and K. A. Fahad, "Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content," International Journal of Advanced Computer Science and Applications, pp. 522-531, 2020.

[36] A. Adnan, S. Babar, M. K. Asad, J. L. M. Fernando, A. Gohar, R. Alvaro and A. Sajid, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," International Journal of Information Management, pp. 304-319, 2019.

[37] Ó. María, B. Baesens and V. Jan, "Profit-Based Model Selection for Customer Retention Using Individual Customer Lifetime Values," Big Data, pp. 53-65, 2018.

[38] H. Sebastiaan, S. Eugen, B. Bart, v. B. Seppe and V. Tim, "Profit Driven Decision Trees for Churn Prediction," European Journal of Operational Research, 2017.

K. T. Hiren, D. Ankit, G. Subrata, S. Priyanka and S. Gajendra, "Clairvoyant: AdaBoost with Cost-Enabled Cost-Sensitive Classifier for Customer Churn Prediction," Hindawi Computational Intelligence and Neuroscience, vol. 2022, 2022.