Predicting Customer Churn in the Internet Service Provider Industry of Developing Nations: A Single, Explanatory Case Study of Trinidad and Tobago

Chapter · March 2021

DOI: 10.1007/978-3-030-69717-4_77

2 authors:

Lackeshwar Bachan
University of Liverpool
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Tarek Gaber
University of Salford
98 PUBLICATIONS **1,705** CITATIONS

SEE PROFILE

PREDICTING CUSTOMER CHURN IN THE INTERNET SERVICE PROVIDER INDUSTRY OF DEVELOPING NATIONS: A SINGLE, EXPLANATORY CASE STUDY OF TRINIDAD AND TOBAGO

Lackeshwar Bachan

University of Liverpool, Liverpool L69 3BX, United Kingdom
lackeshwar.bachan@online.liverpool.ac.uk

Tarek Gaber

Faculty of Computers & Informatics
Suez Canal University, Ismailia, Egypt.
Member at SRGE Research Group (www.egyptscience.net)
tmgaber@gmail.com

Abstract. Customer churn in the telecommunications industry is a problem of great concern to companies. The high cost involved in acquiring new customers has shifted the telecoms sector's focus on retaining new customers. With an average churn rate of 1.9%, the ability to predict accurately when a customer might churn has become an asset to telecommunications companies. This research aims to find a subtle method of predicting customer churn by applying machine learning algorithms - Decision Tree, Logistic Regression, and Support Vector Machine – to a dataset and evaluate how they stack up against each other in determining customer churn. The dataset provided for this research comes from Amplia Communications Limited, a major ISP in Trinidad and Tobago. Python is used to develop the IT artifact in the research. The dataset is preprocess using label-encoding, one-hot encoding, and Pearson correlation is used for feature selection. The processed dataset is then normalized and split into 70% / 30% for training and testing, respectively. Using the evaluation metrics of Accuracy, AUC, and F1 score, the Decision Tree model with values of 89.5%, 81.9%, and 86.9% respectively outperformed the other two machine learning models. In conclusion, the decision tree model would be the best suited for predicting customer churn from the results.

Keywords: Predictive analytics · Customer Churn · Machine learning · Support vector machine (SVM) · Decision tree (DT) · Logistic Regression (LR)

1 Introduction

Customer churn is an always present factor in any service-related industry, with each sector having its unique methods and techniques to deal with this event [1]. In the telecoms sector, customer churn is defined as the event when a customer is no longer interested in a company's services and wishes to terminate said service. Customer churn is a rapidly growing issue in the telecoms section. ISP has a more reactive approach to customer churn, which, in most cases, is too late to save the customer [2]. The cost of acquiring new customers has shifted the telecoms sector's focus to ensure that their subscriber base retention of new customers is optimal.

There is a gap in the ISPs of Trinidad and Tobago between understanding their customer base and the various issues they face [3]. The internet service providers (ISP) of the country, however, are faced with unique customer churn issues. ISP traditional approach to customer churn, is reactive only when a customer has informed the ISP of their decision to terminate the service would the retention team try to identify the customer issues and prevent churning [2]. The machine learning model will address these limitations by using more localized data and broader metrics to give a greater model accuracy would allow the ISP to take a proactive approach to address the customer issue, may it be specific to that customer or a more significant problem.

This paper aims to demonstrate the use of Decision Tree, Logistic Regression and Support Vector Machine algorithms applied to the dataset, and compare that which classifiers have better performance using various measures, like accuracy, AUC, and F1-score, confusion matrix in the prediction customer churn. Moreover, the research aims to create a model using machine learning that would apply specifically to the Trinidad and Tobago ISP industry to predict at-risk customers and their issues. This model would allow the ISP industry to improve its customer experience and overall service level significantly. Using more localized data and broader metrics to give a greater model accuracy would allow the ISP to take a proactive approach to address the customer issue, may it be specific to that customer or a more significant problem [4].

The remaining paper is structured as follows: Sect. 2 discusses related work in this field. The experimental methodology is presented in Sect. 3. Section 4 comparative analysis of the machine learning classifiers. At last, Sect. 5 concludes the paper.

2 Literature Review

Research conducted by [5] has demonstrated that mobile telecommunication has become the most dominant mode of communication in the last two decades. This is not only being witnessed in developed countries but also this mode has revolutionized communication in developing countries. In developing countries, telecommunication entities are competing to acquire and retain clients. Whereas in developed countries, the markets have reached a saturation point where new clients or lost clients are worn by other competitors. In order to protect consumer rights, various regulations have been enacted and implemented by various jurisdictions. Regulations have been accompanied by standardizing mobile telecommunications, which has enabled customers to shift

from one service provider to the other. This has led to the emergence of a very fluid telecommunication market.

According to [6], it is quite expensive to acquire new clients; therefore, what most internet service providers have been implementing is shifting attention to customer retention instead of customer acquisition. This has accelerated the adoption of churn prediction models, which has become a prevalent tool for Business intelligence applications whose goal is to identify customer churn rates and the correlation between churn features and shift to competitors.

[7] used a cluster analysis technique to investigate mobile users churn with Large scale telecommunication companies in Jordan. [7] used a dataset obtained from 700 surveys. By employing the k-means cluster technique, he was able to showcase the correlation between customer satisfaction and customer churn. The research leveraged on the "Elbow rule" to get insights into cluster k numbers, whereas the ward Method adapted as the algorithm to minimize within-cluster differences. This methodology was able to generate a total of 3 clusters. In the analysis, gender, education, mobile network service providers, age, tenure, and residence were used as the demographic factors. The insight derived from the research demonstrated the varied results between the independent variables. The k-means cluster technique demonstrated that customers were not willing to churn their current service provider because of any inconvenience caused. For this paper, K-means will not be adapted. However, this literature blends well with the objective of Decision trees and Logistic regressions, which feed on binary data in their implementation [7].

For this particular literal work, [8] focused on customer churn prediction with little emphasis on a lot of historical data due to many factors. They used a Just in time across the company model to complement their research, which served as a practical alternative. This was achieved by obtaining publicly listed datasets fed to Support Vector Machines as the base classifier tool. Just in Time model works by using data from mature companies as the training data and data from newly established companies as the validation test. Results from the model showed that the Just In Time model had a 55.33% accuracy. This would be ideal for the Trinidad and Tobago research as the test subject has little historical data to do with the features that affect customer churn.

[9] applied a deep learning stacked autoencoder network model along with Logistic regression for their churn prediction. They used data from 400,000 client's behavioral records. 80% of the data formed the training dataset as the rest was used as the test data set. Their model used a Stacked Autoencoder Network popularly known as SAE, renowned as a classical unsupervised machine learning model. This model, SAE, is preferred if the research has large datasets, reduces data dimensions, and when the data has few labels. From the models' implementation, the output is a logistic regression classification, which can classify whether a customer has churned or not churned. The output from the model is then evaluated by using a confusion matrix to get results precision-recall. An F1 score was also used to measure precision and recall balance. Further evaluation demonstrated that the SAE model had an accuracy of 73.5%, which was greater than the Logistic Regression accuracy of 71.8%. The researchers identified the model's challenge and recommended that it be optimized by continuously adjusting

the parameters. This dissertation will apply both the SVM and Logistic Regression models [9].

Research done by [10] focus on the use of logistic regression model to solve the problem of customer churn in telecommunications companies. Churn as described in the paper is when an individual of the telecommunications company decides to opt out on their service with said ISP. Is churn lost is a cost to the telecommunications company from monthly rental cost to cost of collection of equipment. To address is problem the researchers acquired a dataset which contain various customer information such as demographic data, account information, customer service information etc. This dataset if firstly pre-processed to address any missing data, attributes having NA's, incorrect data types or data format. Once this pre-processing is done feature selection is started. Selecting the right features leads to better accuracy and model execution. The researchers' approach to this is using exploratory data analysis using visualization and statistical tests. They used the chi-square test for categorical data and z-test for continuous data. Each feature is selected based on a level of confidence. Once completed the dataset is passed through the LR model and evaluated based on the Receiver Operating Characteristic (ROC). The LR had an Area under the curve value of 0.83 for a threshold of 0.5.

In summary there is a lot of work done using different models to predict customer churn. Some like the work done by [8] and [9] did not produce favorable results that would allow for a model to be used in a real-world environment. Some limitations were the quality and quantity of the dataset. Other work like that of [10] produced excellent results by its focus was limited to a singular machine learning model.

This study builds on the research done already by ensuring the dataset is preprocessed to remove all anomalies and utilized techniques such as one-hot encoding and converting categorical features to numerical. Pearson Correlation would be used for feature selection before the finalized dataset is applied to the three machine learning models. Area under the curve in conjunction with F1 score will be used as an evaluation metric as it allows for the better accuracy.

3 Methodology

In this section, we briefly explain the steps of the churn prediction methodologies.
Generally, in applying machine learning on the ISP dataset, there are four steps need to be followed:

1) Data Preparation – In this stage the raw data is cleaned by correcting any formatting issues, corrections to the data and combining of the data to enrich its value.
2) Data Pre-processing – In this stage the data in the dataset is modify so that it meets the requirements to be applied to the models. The dataset is also split into training and testing 70/30 ratio.
3) Model Construction – At this stage the three models' various parameters are selected, and the dataset applied.

4) Prediction Accuracy – In this final stage of the process the models are evaluated and compared against each other to see who has the better predictive power.

Fig 1 shows the flow chart of churn prediction model for this study.

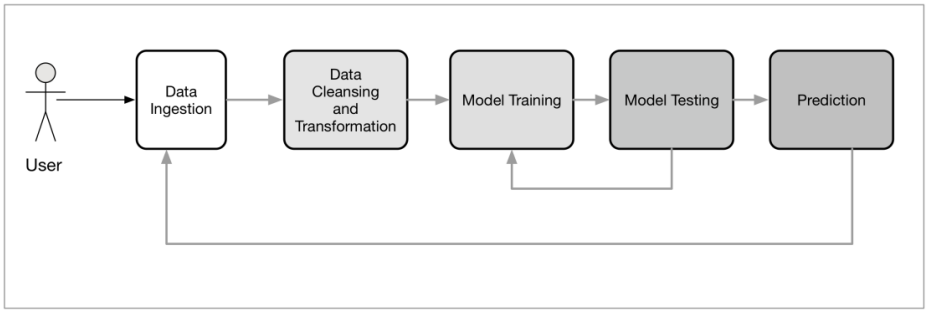

Fig. 1. Methodology Pipeline

3.1 Decision Tree

This model uses tree structures analogy to generate data classification rules. Decision trees can perform both classification and regression analysis. In classification, decision trees are able to generate nodes that act as IF statements, which are able to continuously divide the data and adding nodes to already divided data. This is done iteratively until each branch's accuracy does not improve, the results in a leaf which ends up giving us the output of the remaining dataset in the branch. This algorithm is analogous to a tree's physical structure, and it can classify a dataset based on those physical attributes. The advantages of Decision trees are that they can do both classification and regression analysis. In addition to this, they are robust to the extent that they can handle both categorical and numerical data. This algorithm can classify features and their effects to churn by continuously adding nodes on the classified data. If the accuracy is no longer improving, then the algorithm will give results of the leaves as the outputs of the datasets remaining in the branch [11].

3.2 Logistic Regression

According to [12], Logistic regression belongs to a modelling class called the generalized linear model. The main objective of these generalized linear models for binary dependent variables is to estimate regression equations that relate the output target of the dependent variable known as the y variable to predicator variables globally represented by the x variable [13]. The most common approach in regression is to model the independent variable as a function of independent variables. The drawback of linear regression is that it is impossible to predict outputs outside the 1-0 permissible range. To have a robust, we also need to loop in an algorithm that can give nonlinear predictions, mostly when we expect binary outputs. This project has the objective of

predicting customer churn rate, either churn (1), Not churn (0). This is where logistic regression comes in handy.

3.3 Support Vector Machine (SVM)

This technique classifies data by computing an optimized hyperplane that classifies different classes. A 2-Dimensional space shows the hyperplane as a simple line, which separates different labels such that when a new dataset is fed into the model. Various features can fall on either side of the line. Support Vector Machine algorithm can classify data by computing an optimal hyperplane, which is used to categorize different classes. The algorithm can separate features that contribute to churn rates and those that do not promote churning [14]. This research will utilize the Radial basis function, also called the Gaussian kernel.

3.4 Pearson Correlation

This is a method that shows a linear relationship between various attributes. In Machine learning, this linear relationship can be visualized as a matrix that can show the relationship between an attribute and itself or/and relationships between various attributes. From Pearson correlation coefficients, we can know which attributes have strong correlations [15]. In this case, before any modeling is done, we either add them together or drop one of the attributes as both of them will in the dataset not affect the outcome of the results. Pearson correlation is a unique algorithm that can produce coefficients, range from -1 to 1. If the correlation between the target output and the variable is closer to 0, then the conclusion is that the target output does not correlate with the attribute. A positive value that leans towards 1 shows a strong positive correlation [16].

3.5 Performance Measures

The performance of the three models was evaluated using Accuracy, F1 score which is seen in the classification report and Area under the Curve.

Accuracy as defined by the equation below is measured by the ratio of correct predictions to the total number of cases evaluated [17]. Thus, the higher the accurate the result.

$$ACC = (TP + TN) / (TP + TN + FP + FN)$$

F1 score as defined by the equation below is the weighted harmonic mean of recall and precision, which gives 1 the best score and 0 the worst score [17].

$$F1 = 2\ TP / (2TP + FP + FN)$$

It is important to note that F1 scores will always be lower than accuracy measures because they include both recall and precision in their computation. Standard practice dictates that F1 scores should be used when comparing classifier models and not as a global accuracy criterion [18].

AUC measures the entire two-dimensional area underneath the entire ROC curve. AUC provides an aggregate measure of performance across all possible classification

thresholds. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.

4 Experiment Results and Discussion

4.1 Dataset

The Machine learning algorithms were evaluated using the Internet service provider dataset, which consisted of 16 attributes and 31728 entities with 27166 as non-churners and 4562 churners. Details of attribute are listed in Table 1.

Table 1. ISP customer churn attribute list

Attributes	Description
customerID	unique ID of a customer.
tenure	How many months the customer has/had the service.
customerGroup	The category the Fiber to the Home (FTTH) customer belongs.
internetService	If the customer has internet service.
tvService	If the customer has tv service
voiceService	If the customer has VoIP service.
secureService	If the customer has home alarm monitoring service
technicalTickets	Number of tickets raised by the customer due to a technical service issue, ranging from WIFI issues, Ethernet cable issues, fiber breaks, damage equipment etc.
nontechnicalTickets	Number of tickets raised by the customer due to a non-technical issue, ranging from billing issues, account queries, product queries, etc
truckRolls	Number of times a technician visited the customer for a repair issue.
paperlessBilling	If customer receives their bill via email.
customersBills	Number of bills customer receives
paymentsReceived	Number of payments made by customer.
monthlyCharges	Cost of customer monthly services.
outstandingBalance	Total outstanding amount owed by the customer.
Churn	If the customer churned

4.2 Data Pre-processing

There were no anomalies found in the dataset with respect to missing data and incorrect format. A few outliers were found in the box and whiskers plot of monthlycharges vs. churn. On closer examination of these data points, it was found that these were customers taking very high-end packages, so were left in the dataset. The next step was the conversion of the categorical attributes to numerical values which is all yes/no values to No = 0 and Yes = 1. One hot encoding is done on the customerGroup attribute. Lastly the selection of attributes is done using the standard correlation function method called 'Pearson Correlation Coefficient' with results shown in Fig 2. As we can see from the Pearson correlation heat map, Employee and Residential_IPTV have a strong negative correlation. As such, we can remove one of these features. In our dataset, we will drop the Employee feature.

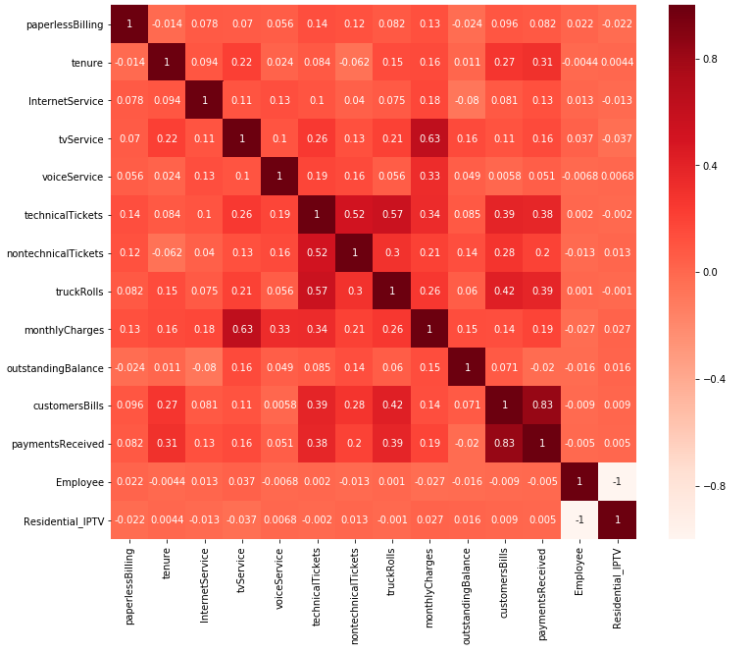

Fig. 2. Visualization of Pearson Correlation

4.3 Results and discussion

The first model tested in keeping with the objectives of this research is the Decision Tree. We first created the Decision Tree using the following features, the max_depth parameter is set to 4, and the criterion is set to entropy. We then trained the model using the training set and then applied the test set. Table 2 shows the evaluation of the decision tree after its implementation, with the decision tree algorithm, we have an accuracy of 89.5% with Area under the curve of 81.9% with an F1-score of 86.9 %.

The second model tested was the Logistic Regression model. Its aim is to evaluate the model performance so it can be compared to the other models. It was created using the C parameter, which represents the inverse regularization parameter, which is set to 0.01 and solver set to liblinear. Table 2 shows that with the Logistic Regression algorithm, we have an accuracy of 88.2%, an Area under the curve of 78.2% with a F1-score of 84.7%. The final model, evaluated as outlined in the research objectives, is the SVM. The probability parameter is set to True and the Kernel used is RBF. Table 2 shows that with the Support Vector Machine algorithm, we have an accuracy of 89.9%, An Area under the curve of 79.4% with an F1-score of 87.6%.

Table 2. Model Evaluation's comparative analysis

Model	**Accuracy**	**AUC**	**F1 Score**
Decision Tree	89.5%	81.9%	86.9%
Logistic Regression	88.2%	78.3%	84.7%
Support Vector Machine	89.9%	79.4%	87.6%

5 Conclusion

In this paper, using a dataset provided for this research comes from Amplia Communications Limited, a major ISP in Trinidad and Tobago, the customer churn in the telecommunication industry is investigated to gain insight whether machine learning technique would help to the ISPs to predict the churn rate so they can deal the problem early avoiding losing customers. Three distinct machine learning algorithms were compared to see their predictive accuracy in determining whether a customer will terminate their services with the organisation. It was observed that the Decision Tree outperformed the Logistic Regression and edged out the Support Vector Machine due to its better AUC of 81.9% but still high Accuracy of 89.5% and F1 score of 86.9%. So, it can be concluded that the decision tree model would be the best suited for predicting customer churn. A future work could be comparing the results of this study with others from different countries to see whether the customers' behavior differs from country to another.

References

1. Manďák, J. and Hančlová, J. (2019). Use of Logistic Regression for Understanding and Prediction of Customer Churn in Telecommunications.
2. Do, D., Huynh, P., Vo, P. and Vu, T. (2017). Customer churn prediction in an internet service provider. *In 2017 IEEE International Conference on Big Data (Big Data)* (pp. 3928-3933). IEEE.
3. 2013. The Digital Divide Survey Trinidad And Tobago. [ebook] Available at: <https://tatt.org.tt/DesktopModules/Bring2mind/DMX/API/Entries/Download?Command=Core_Download&EntryId=340&PortalId=0&TabId=222> [Accessed 19 October 2020].
4. de Haan, E., Verhoef, P. and Wiesel, T. (2015). The predictive ability of different customer feedback metrics for retention. *International Journal of Research in Marketing*, 32(2), pp.195-206.
5. Hadden, J., Tiwari, A., Roy, R. and Ruta, D., 2007. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), pp.2902-2917.
6. Singh, M., Singh, S., Seen, N., Kaushal, S. and Kumar, H. (2018). Comparison of learning techniques for prediction of customer churn in telecommunication. In *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 1-5). IEEE.
7. Al-Refaie, A. (2017). Cluster analysis of customer churn in telecom industry. World Academy of Science, Engineering and Technology International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 11(5).
8. Amin, A., Al-Obeidat, F., Shah, B., Al Tae, M., Khan, C., Durrani, H.U.R. and Anwar, S. (2020). Just-in-time customer churn prediction in the telecommunication sector. The Journal of Supercomputing, 76(6), pp.3924-3948
9. Cao, S., Liu, W., Chen, Y. and Zhu, X. (2019). Deep Learning Based Customer Churn Analysis. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)* (pp. 1-6). IEEE.
10. Sai, B.K. and Sasikala, T. (2019). Predictive Analysis and Modeling of Customer Churn in Telecom using Machine Learning Technique. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 6-11). IEEE.
11. TOFAN, C. (2014). Optimization Techniques of Decision Making - Decision Tree. *Advances in Social Sciences Research Journal*, 1(5), pp.142-148.
12. Jain, H., Khunteta, A. and Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. Procedia Computer Science, 167, pp.101-112.
13. Heeringa, S., West, B. and Berglund, P. (2010). Logistic Regression and Generalized Linear Models for Binary Survey Variables. *Applied Survey Data Analysis*.
14. Jung, K. (2016). Robust Algorithm for Multiclass Weighted Support Vector Machine. *The SIJ Transactions on Advances in Space Research & Earth Exploration*, 4(3), pp.1-5.
15. Hall, M.A (1999). Correlation-Based Feature Selection For Machine Learning. Ph.D. The University of Waikato.
16. Holcomb, Z. (2017). *Fundamentals Of Descriptive Statistics*. London: Routledge.
17. Umayaparvathi, V. and Iyakutti, K. (2016). Attribute selection and Customer Churn Prediction in telecom industry. In *2016 international conference on data mining and advanced computing (sapience)* (pp. 84-90). IEEE.

    Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp.2825-2830.