American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-9, Issue-2, pp-1-7
www.ajer.org
Research Paper Open Access

Deciding the Best Machine Learning Algorithm for Customer Attrition Prediction for a Telecommunication Company

Omijeh Bourdillon O. [1], Abarugo Goodness Chidinma[2]
[1](Electronic and Computer Engineering,University of Port-Harcourt, Rivers State, Nigeria.
[2]Centre for Information and Telecommunications Engineering (CITE),University of Port-Harcourt, Rivers State, Nigeria.)
Corresponding Author: Abarugo Goodness Chidinma.

ABSTRACT : *Customers are so important in business that every firm should put great effort into retaining them. To achieve that with some measure of success, the firm needs to be able to predict the behaviour of their customers with respect to churn or attrition. There are many machine learning algorithms that may be used to predict attrition, but this paper considers only four of them. Logistic regression, k-nearest neighbour, random forest and XGBoost machine learning algorithms were applied in different ways to the dataset gotten from Kaggle in order to decide the best algorithm to suggest to the company for customer attrition prediction. Results showed that the logistic regression or random forest algorithm may be adopted by the telecom company to predict which of their customers may leave in the future based on their recall and precision scores as well as the AUC values of approximately 75%. The logistic regression algorithm gave metrics of 73% accuracy and 59% f1-score while the random forest algorithm yielded 70% accuracy and a 58% f1-score. However, it was also suggested that if the choice of model was based on accuracy and f1-score, the logistic regression model would be the best to be adopted.*

KEYWORDS *– Algorithm, Attrition, Customer, Machine Learning, Prediction.*

Date of Submission: 4-02-2020 Date of acceptance: 20-02-2020

I. INTRODUCTION

Any organization that aspires to be an industry leader needs to see into the future and stay in business and stay ahead of competition even before it arises. For that to happen, the main 'asset' that the firm has to guard with jealousy is the customer. If customer actions can be predicted with a good level of confidence, organizations would do better jobs at customer management. In the past, customer relationship management was based on traditional transaction and demographic data analysis. Today, machine learning techniques are used to go beyond traditional data analysis to predicting the future. However, adopting machine learning techniques wrongly and without proper understanding may not yield any appreciable value. The choice of the best algorithm to use for prediction can best be made if the management knows exactly what it wants to achieve. The decision to adopt a particular algorithm is usually based on the proper trade-off between different performance metrics beyond accuracy. It is also based on the firm's business needs and on the features identified triggers to attrition in the firm.

Different machine learning based predictive models have been used by different researchers to predict churn with some great results and comparisons. Reference [1] used deep learning algorithms to predict customer churn in a prepaid telecommunication company with about 77.9% AUC on validation data, which they confirmed was an improved result in comparison to the value of about 73.2% gotten when they used random forests with complex feature engineering. Reference [2] used classification and Regression Trees (CART), Support Vector Machine, Additive Regression, K-Star, Multilayer Perceptron (Neural Network) and Wavelet Neural Network (WNN) for the calculation of the future value of 24 customers. They realized that the MLP gave the best accuracy amongst all others. However, they suggested that CART is still a viable algorithm for proper segmentation of customers to prediction. After comparing neural networks, regression and decision trees, [3] and [4] found that artificial neural networks exhibited better performance that the other two algorithms.

Another study carried out by [5], shows that when two-layer Back-Propagation neural network (BPN), Decision Trees, SVM and Logistic Regression were applied to predicting lifetime value from churn rate, decision trees and BPN achieved accuracy figure of 94%, SVM tagged along with 93% while Logistic Regression achieved 86% accuracy. Reference [6] analyzed the accuracy of Multi-layer perceptron (MLP) and Decision Tree (C5) machine learning models on a business dataset using the RFM and discovered that MLP beat decision Tree (C5) with accuracy of almost 96% while the later had almost 90% accuracy. To predict customer churn on a telecommunication dataset, [7] applied random forest and ADA boost, Multi-layer perceptron and Support vector machine, Decision tree, naïve Bayesian, logistic regression and Linear Discriminant Analysis (LDA) machine learning algorithms and observed that each group came out with 96%, 94%, 90%, 88% and 86.7% accuracies respectively.

It can be observed that in all reviewed cases, neural networks were used alongside other algorithms and neural networks scored highest in the accuracy performance metric. Most of the researchers also recommended it. This is expected and may be a clear sign that neural networks should not always be made to compete with other machine learning algorithms except in rare cases where they can all compete favorably. In cases where the analyst is not conversant with the functionalities of the neural networks, the analyst needs to have an idea of which other algorithm to use that is not based on accuracy alone. This paper excluded neural networks and pitted four machine learning algorithms which have different heuristics against each other in predicting churn for a firm. Accuracy was not the only metric used for evaluation. Precision, recall, F1-score and the AUC-ROC were also put into consideration. Rationalizing an investment in a facility's infrastructure can be a difficult prospect for any plant engineer or technician, often requiring extensive justification. Investments that are deemed “low-risk” by upper management and have a fast return on investment (ROI) are typically the easiest to substantiate. One such investment that will pay considerable dividends over the course of its operating life is a comprehensive power monitoring system.Even though increased energy prices have become a larger influence on the balance sheet, many facilities do not take advantage of opportunities to better manage these expenses. Those without monitoring systems likely have no understanding of their energy usage; those with them may not be using their systems to the fullest potential.

Because the quality of energy supplied can adversely affect its operation, oftentimes leading to loss or degradation of equipment, product, revenue, and reputation, plant managers must weigh the advantages of implementing a monitoring program.

The second section of this paper shows three methods for monitoring systems of solar plants. The third section discusses communication and monitoring system for wind turbines, and finally the conclusion is discussed in the fourth section.

II. METHODOLOGY

Figure 1 is a block diagram showing the methodology adopted to arrive at the results presented in this paper.

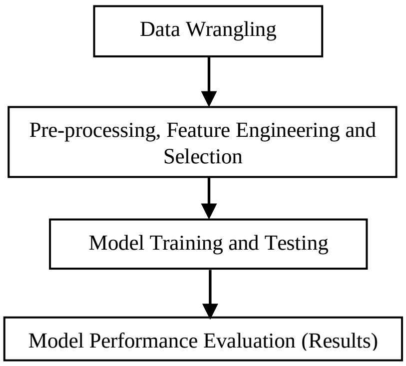

Figure 1: A block Diagram Showing Adopted Methodology

A. Data Wrangling

Data wrangling involves cleaning the data from its raw state to a more understandable format. It involves cleaning the data, taking care of missing values, assign appropriate data type to features and normalize the table. Figure 2 is a flow chart that shows the process involved in data wrangling.

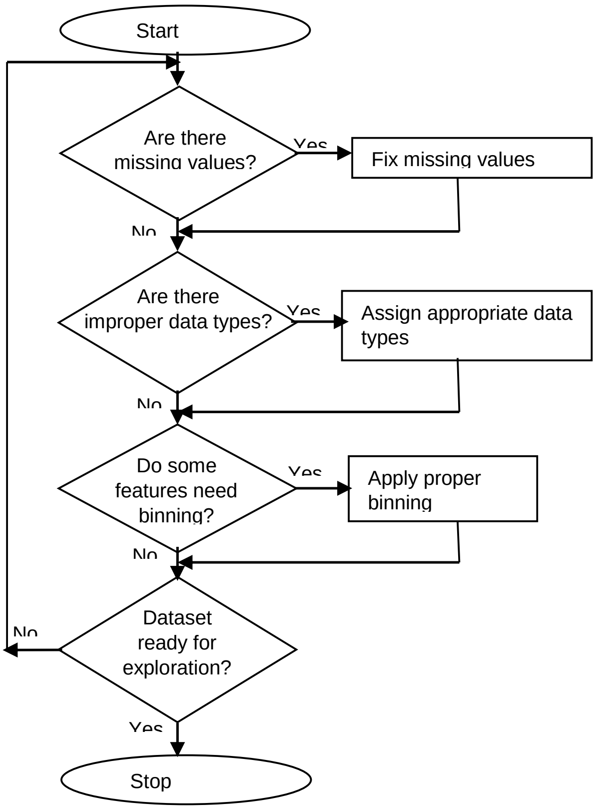

Figure 2: A Flowchart of the Data Wrangling Process

Missing values that were found in the “TotalCharges” column were taken care of by replacing them with zeros. After that, the total charges column was converted from object datatype the right datatype which is float (numeric) datatype. The “SeniorCitizen” column was defined as an integer by default but good reasoning identifies it as a categorical feature and not numeric. In order to clean it up, a python data dictionary was used to convert the values of 0’s and 1’s to “no” and “yes” respectively in order to normalize the table and ensure consistency. The tenure feature is made of a lot of different values. In order to make the feature useful, another feature was created out of it, which binned tenure into 5 different groups.

B. Pre-processing, Feature Engineering and Selection

The steps involved were carried out in the following ways:

1. Feature Encoding:The encoding was done by mapping the two unique values of features to numeric ‘1’ and ‘0’. The features that have more than two unique values were encoded using ‘get_dummies’. ‘Get dummies’ class split each multivalve feature into the number of unique values that feature has. Whatit did was to break categorical columns into as many unique values as the categorical column has by creating new features out of them and replacing the values with 0’s and 1’s (that is converting all of them to numeric columns).
2. Scaling Numeric Features: All the numeric columns were then scaled using the standard scalar function to normalize the values and place them all within the same range and to speed up process of machine learning.
3. Merging Engineered Features: All the encoded features as well as the scaled numeric features are added to the original dataset. The original features that were subjected to encoding were also dropped. This process was aimed at presenting a new pre-processed dataset, ready for machine learning.

C. Model Training and Testing

There are a lot of available algorithms for predictive modelling and machine learning. This paper applies four different algorithms: logistic regression, k-nearest neighbour, random Forest, XGBoost. These algorithms perform the same task of predicting a dependent feature from a number of independent features; however they work based on varying heuristics and have different predictive abilities based on the dataset presented. Figure 3 shows the steps applied in the training and performance analysis process.

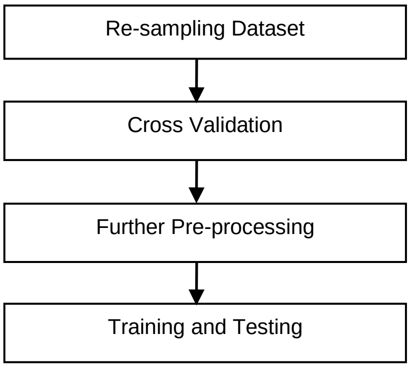

Figure 3: Steps Involved in Model Training and Testing

1. Re-sampling Datasets - The entire dataset cannot be used at once. It needs to be divided in such a way that a proportion of it remains ‘unseen’ by the machine learning algorithms until it is time to test their effectiveness in predicting outcomes. In order to split the dataset, the ‘train_test_split()’ function was used.75% of the entire record was used to train each of the machine learning algorithms while 25% was reserved for testing the ability of the algorithm to predict outcomes based on how much it has learned from the training set.
2. Cross Validation - The aim of cross validation is to evaluate how well machine learning models behave when given a new dataset using specified algorithms. This is done so that the results of the machine learning prediction can be generalizable. The k-fold method of cross validation was applied. It involved iterating through dividing the dataset into train and test and calculating the mean score of the performance metrics.
3. Further Pre-processing - The dataset was checked for imbalance because the proportion of customers who left the company was far less that the proportion of existing customers. The algorithms applied to the training dataset were SMOTE and RFE in order to optimize the performance of the learning algorithms. The designated portions of the datasets were then used for the machine learning process.
4. Training and Testing - At this point, the engineered training set was fed to the algorithms for actual training while the portion that was held-out as the testing dataset was fed to the algorithm for prediction after learning.

D. Evaluating Model performance

The performance scores from each algorithm were recorded for discussion. The metrics include the precision, recall, f1-score, accuracy and area under the curve (AUC) of the receiver operating characteristic (ROC) curve.

Accuracy means the proportion or ratio of the entre test dataset that were predicted correctly. Precision or specificity is defined as the number of data points that the model or algorithm identified as relevant but actually were not relevant. .Recall is used to explain how well a model can identify relevant data points within a dataset after learning.

To make the relevance of these metrics clearer, Precision here asks the question: Of all the users that the algorithm predicts will not leave the company, how many of them actually left?Recall asks: What percentage of customers that did not leave the company did the algorithm correctly identify?

The last metric used to evaluate performance is the f1-score. It is derived as a sub contrary mean of the sensitivity value and the number of positive predicted values. For problems where both sensitivity and positive predicted values are important, one can select a model which leverages on the balanced nature of the F-1 score [8]. It is also known as the balanced accuracy metric.Another important visual metric used to describe the skill of the algorithms is the receiver operating characteristic (ROC) curve. It was gotten by plotting true positive rate against the false positive rate (that is, Sensitivity vs. (1-Specificity)).

The model with the best performance has higher the area under the curve (AUC) in an ROC. Based on the results, suggestions were given on the best models to use by the company to predict customer defection in order to take strategic business decisions.

III. RESULTS AND DISCUSSION

The baseline model was applied using k-Fold resampling and cross validation method. The result presented in table 1 shows the performance metrics of the baseline models. This serves as the baseline for measuring the effectiveness of the models after training and testing. The datasets were then fed to the algorithms. The algorithms learned and created models using the training dataset and then the test dataset was used for prediction to yield the performance metrics in the table 2. The optimized training set (SMOTE and RFE applied) was then used for training the algorithms before prediction using the test set which was not tampered with. This process yielded the performance metrics in the table 3. The receiver operating characteristics are shown in figure 4, 5, 6 and 7.

Table 1: Baseline Model Performance Metrics

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Logistic Regression	80%	54	67	60
XGBoost	80	53	67	59
K-Nearest Neighbor	76	52	57	55
Random Forest	78	44	62	51

Table 2: Performance Metrics After First Training And Prediction Without Optimization

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Logistic Regression	81	56	63	59
XGBoost	75	49	63	55
K-Nearest neighbor	76	54	52	53
Random Forest	79	65	31	42

Table 3: Performance Metrics After Second Training And Prediction With Optimization

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Logistic Regression	73	82	47	59
XGBoost	71	74	45	56
K-Nearest neighbor	74	71	48	57
Random Forest	70	86	45	59

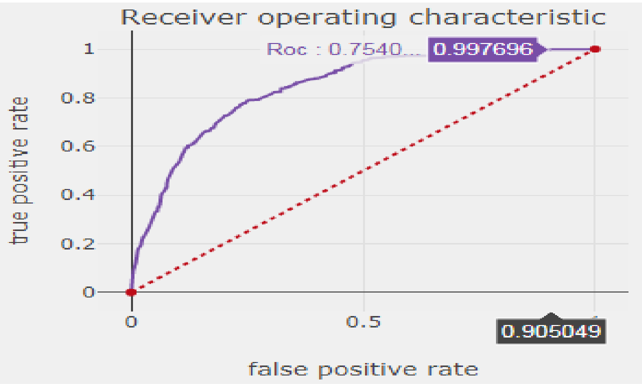

Figure 4: ROC-AUC for Random Forest Algorithm

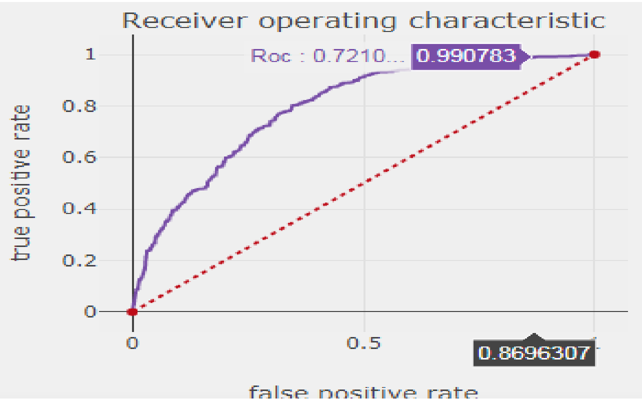

Figure 5: ROC-AUC for XGBoost Algorithm

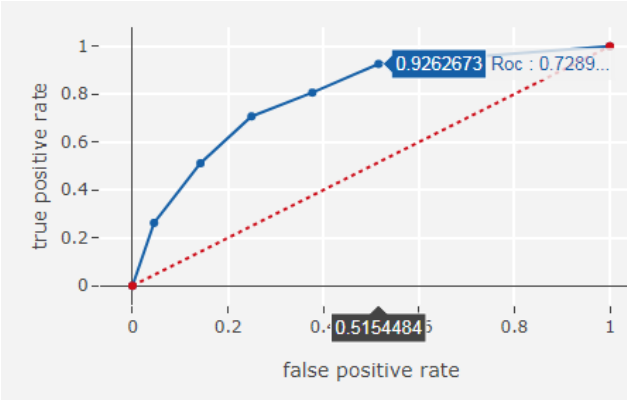

Figure 6: ROC-AUC for KN-Neighbour Algorithm

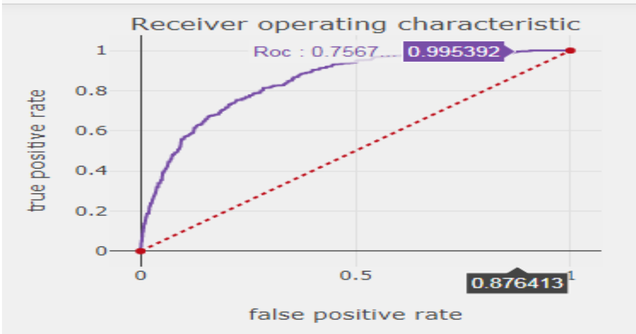

Figure 7: ROC-AUC for Logistic Regression Algorithm

There always has to be a choice between precision and recall as no algorithm can predict customer attrition with 100% accuracy. 100% accuracy will not even be ideal in a random uncertain world. After optimization, the precision metrics for all the algorithms improved. The Random Forest algorithm showed the best improvement in both recall and precision. The f1-score of 58 is also not bad when compared to other algorithms. Affecting one of them results in a change in the opposite direction in the other. So, after the models were tweaked by oversampling and selecting best features, the recall values improved with an opposite effect on the precision.

The choice of one model is made at the expense of another. If the management knows exactly what it wants to do with the results, the choice between precision and recall may be made based on the needs of the enterprise, on the variables identified as triggers to attrition or no attrition.

If the company wants to reach certain promotions towards clients who might ditch them for another company, then precision is the metric to concentrate on. If they want to give promo offers to all clients, without missing any of the customers that may leave in the future, then they have to choose a higher recall value over precision.

The f1-score balances out precision and recall. However, if performance shows improvement in any metric when compared to the baseline, and other metrics are not far apart, the said algorithm may be fit for selection.

The suggestion in this case is that the logistic regression or random forest ML algorithm may be considered by the telecom company to predict which of their customers may leave in the future. This selection may be done based on the calculated choice between precision and recall, or the f1-score. This is because while the logistic regression model yielded 73% accuracy, a 59% f1-score, an 83% recall score and a 47% precision score, the random forest algorithm yielded a 70% accuracy, 58% f1-score, 85% recall score and 44% precision score as indicated in Tables 4.1, 4.2 and 4.3. The ROC curve also affirms that both the random forest and logistic regression algorithms performed better with approximately 75% AUC value each as shown in figure 4.7. However, if the choice is based on accuracy and f1-score alone, then the logistic regression model should be adopted.

IV. CONCLUSION

This paper successfully reports the processes involved in deciding the best machine learning algorithms to adopt for customer-attrition prediction for a telecommunication company. The findings of the study reveal that the logistic regression and random forest algorithms performed better than others. While the logistic regression model yielded 73% accuracy, a 59% f1-score, an 83% recall score and a 47% precision score, the random forest algorithm yielded 70% accuracy, 58% f1-score, 85% recall score and 44% precision score. The ROC curve also affirms that both the random forest and logistic regression machine learning algorithms performed better with approximately 75% AUC value each. Several areas for further research involve using a larger (more than one company) and more diverse population. The models can also be trained for longer periods of time. RFE may be applied during cross validation and datasets may be broken into different train-test proportions to assess how the performance metrics may be affected.

REFERENCES

[1]. Federico Castanedo, Gabriel Valverde, JaimemZaratiegui& Alfonso Vazquez. Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network. (2014)

[2]. TarunRathi. Customer Lifetime Value Measurement using Machine Learning Techniques. (2011)

[3]. AfaqAlam Khan, Sanjay Jamwal, &M.M.Sepehri. “Applying data mining to customer churn prediction in an Internet Service Provider.” International Journal of Computer Applications. Vol 9. (2010).

[4]. Tom Au, Shaomin Li &Guangqin Ma. “Applying and Evaluating Models to Predict Customer Attrition Using Data Mining Techniques.” Journal of Comparative International Management, vol 6, pp. 10-22. (2003)

[5]. Vafeiadisa T., Diamantarasb K. I., Sarigiannidisa G. &Chatzisavvasa K. A comparison of machine-learning techniques for customer churn prediction” Simulation Modelling Practice and Theory.. (2015)

[6]. Samira Khodabandehlou& Mahmoud Zivari Rahman.. “Comparison of supervised machine-learning techniques for customer churn prediction based on analysis of customer behaviour.” Journal of Systems and Information Technology, Vol. 19, pp. 65-93, https://doi.org/10.1108/, (2017)

[7]. Sahar F. Sabbeh. “Machine-Learning Techniques for Customer Retention: A Comparative Study.” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, pp. 275-281. (2018)

[8]. Shruti Saxena. Precision vs Recall. Retrieved from: https://towardsdatascience.com/precision-vs-recall-386cf9f89488 on 30th October, 2019. (2018).

Abarugo Goodness Chidinma"Deciding the Best Machine Learning Algorithm for Customer Attrition Prediction for a Telecommunication Company”. *American Journal of Engineering Research (AJER)*, vol. 9(02), 2020, pp. 1-7.