International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-8, Issue-2, March 2020
https://doi.org/10.21276/ijircst.2020.8.2.1
www.ijircst.org

Enhanced Churn Prediction in the Telecommunication Industry

Awodele Oludele, Adeniyi Ben*, Ogbonna A.C., Kuyoro S.O., Ebiesuwa Seun

ABSTRACT- Prediction models are usually built by applying a supervised learning algorithm to historical data. This involves the use of data analytics system that uses real-time integration and dynamic real time responses data to detect churn risks. Subscribe are increasingly terminating their membership agreement with telecommunication companies through mobile number portability (MNP) in order to subscribe to another competitor companies.

To model the Customer prediction, a Markov Chain Model will be used. The Markov model allows for more flexibility than most other potential models, and can incorporate variables such as non-constant retention rate, which is not possible in the simpler models. The model allows looking at individual customer relationships as well as averages, and its probabilistic nature makes the uncertainty apprehensible. The Markov Decision Process is also appealing, but since dynamic decisions along the lifetime of the customer will not be evaluated the Markov Chain is the simplest model that still meets the requirements. Each state in the Markov Chain will represent a person being a customer for one month, with an infinite number of states. The transition probability to move from one state to the next is equivalent to a customer retaining with the operator to the next month. A customer that has churned will be considered lost forever.

Once the retention and churn rates are determined, the reference churn value for each customer will be computed. The churn rate will be calculated using MATLAB Monte Carlo simulations, running a large number of fictitious customer-company relationship processes, and extracting the results of the average customer.

Using simulation approach gives better result than analytical methods, since an indefinite number of states make matrix algebra complicated. It also allows visualizing the distribution of the results more easily than with algebraic calculation.

To the telecom companies the result of this analysis would improve the level at which they can predict customer churn, because, it will give insight on why a customer would choose to leave one telecommunication industry for another telecommunication industry. In other words, the cost of advertisement and loyalty programs as well as challenges face in retaining loyal customers would be identified. The researcher believes that the enhanced model for churn prediction developed from this study will lead to better retention strategy, improved telecom quality service, enhanced customer loyalty due to the improvement service from applying the information from this model. The study is also significant to researchers, behavioral scientist, business analysts as well as professionals in the computer science domain. The study will serve as a good reference material for research and contribute to the growing objective of developing enhanced model built on data mining techniques that can explain the churn behavior with more accuracy than using single methods.

KEYWORDS- Prediction Models, mobile number portability, Markov Decision Process, churn rate

I. INTRODUCTION

Churn analysis, is useful in many businesses with many customers or high-value customers [4]. Customer churn analytics are being used for a variety of reasons. In the financial services, consumer package goods, energy, manufacturing insurance sector, etc churn analysis is used to measure account holder lifecycle, detect users thinking of switching banks; develop a support model that encourages loyalty, measure how much revenue is at risk of being lost to other providers, measure churn for direct and downstream buyers and predict a user's likelihood to close a policy. In the telecom sector, a churn analysis will show the number of users or accounts that cease using an organizations products or services over a set time period. Churn analysis also identifies customers who are most likely to churn. This identification of valuable customers likely to churn and the execution of proactive steps to retain customers are the characteristics of churn management [10]. Prediction models are usually built by applying a supervised learning algorithm to historical data. This involves the use of data

Manuscript received March 20, 2020

Awodele Oludele, Computer Science Department, Babcock University, Ilishan-Remo, Nigeria.

Adeniyi Ben, Computer Science Department, Babcock University, Ilishan-Remo, Nigeria. (e-mail: ben.adeniyi@gmail.com)

Ogbonna A.C, Computer Science Department, Babcock University, Ilishan-Remo, Nigeria.

Kuyoro S.O., Computer Science Department, Babcock University, Ilishan-Remo, Nigeria.

Ebiesuwa Seun, Computer Science Department, Babcock University, Ilishan-Remo, Nigeria.

.

analytics system that uses real-time integration and dynamic real time responses data to detect churn risks. Subscribe are increasingly terminating their membership agreement with telecommunication companies through mobile number portability (MNP) in order to subscribe to another competitor companies. According to [1], telecommunication companies alone account for 30% of churn rate worldwide. It is cheaper to prevent churning than to acquire, advertise or attract new customers. In order to achieve this, telecommunication companies must be able to manage churn effectively.

Literature show that several solutions have been proffered to detect churn behavior. However, due to firm rivalry new innovations, low switching costs, deregulation by governments, such solutions become ineffective overtime. Some of these solutions were hampered by the restrictions on data collection and data imbalance. Also, in most works, only one data mining method was applied and no room for adequate comparisons. Few authors have attempted to combine techniques; however, these were only able to predict momentary churning behaviors. Hence, the need for a model that can accurately predict churn behavior. The focus of this work is developing an enhanced churn management model by comparing five different algorithms for the prediction of churn behavior. The aim of this work is to develop an enhanced predictive model for churn management in the telecommunication industry using data mining techniques. This would be carried out by a comparative analysis of the data mining classifier algorithms then design an enhanced predictive model based on the outcome of the analysis, which would be implemented as an enhanced predictive model.

II. METHODOLOGY

To carry out a comparative analysis of the existing churn management models, we study the various characteristics of existing models based on techniques used methods of data classification and feature selection processes. Based on this comparison, this study can discover various types of knowledge, including association, classification, clustering, prediction, sequential patterns and decision tree. The knowledge acquired from this comparison will then be classified into general knowledge, primitive-level knowledge, and multilevel knowledge.

The design of the enhanced predictive model will comprise the selection of the following classification algorithms; Decision Tree, Random Forest, Neural Network. Support Vector Machine and Logistic Regression. These classifiers will be evaluated using sensitivity, accuracy, correctly classified instances and specificity. The algorithm with the best result will be used to train and build the model for predicting customer churn in telecommunications sector. Also, the open source data mining software R using Rattle as an interface will be used as the trees produced with the software are less complicated and more compact than some other implementations (such as in WEKA). The imbalanced dataset affects the performance of algorithm. Thus, additional techniques such as under-sampling will be introduced. Furthermore, the data will be trained. The data set will be split into a train and a test set. The train set is a set of examples used for learning. The test set is an independent data set used to assess the performance of the learned classifier. Training set consists of a random hold-out sample of 70% of the total data set. The test set consists of the other 30%. To make sure that each class is represented in the train and test set, stratified sampling is used. With stratified sampling a sampling fraction of each outcome class that is proportional to that of the total population is used. Thus, each set contains approximately the same percentage of samples of each target class as the complete set.

To model the Customer prediction, a Markov Chain Model will be used. The Markov model allows for more flexibility than most other potential models, and can incorporate variables such as non-constant retention rate, which is not possible in the simpler models. The model allows looking at individual customer relationships as well as averages, and its probabilistic nature makes the uncertainty apprehensible. The Markov Decision Process is also appealing, but since dynamic decisions along the lifetime of the customer will not be evaluated the Markov Chain is the simplest model that still meets the requirements. Each state in the Markov Chain will represent a person being a customer for one month, with an infinite number of states. The transition probability to move from one state to the next is equivalent to a customer retaining with the operator to the next month. A customer that has churned will be considered lost forever. Once the retention and churn rates are determined, the reference churn value for each customer will be computed. The churn rate will be calculated using MATLAB Monte Carlo simulations, running a large number of fictitious customer-company relationship processes, and extracting the results of the average customer. Using simulation approach gives better result than analytical methods, since an indefinite number of states make matrix algebra complicated. It also allows visualizing the distribution of the results more easily than with algebraic calculation.

To the telecom companies the result of this analysis would improve the level at which they can predict customer churn, because, it will give insight on why a customer would choose to leave one telecommunication industry for another telecommunication industry. In other words, the cost of advertisement and loyalty programs as well as challenges face in retaining loyal customers would be identified. The researcher believes that the enhanced model for churn prediction developed from this study will lead to better retention strategy, improved telecom quality service, enhanced customer loyalty due to the improvement service from applying the information from this model. The study is also significant to researchers, behavioral scientist, business analysts as well as professionals in the computer science domain. The study will serve as a good reference material for research and contribute to the growing objective of developing enhanced model built on data mining techniques that can explain the churn behavior with more accuracy than using single methods.

III. REPORT

The aggregated telecom data dataset for all variables is presented in Table 1. These variables formed the

benchmark upon which the prediction models for forecasting the behavior of customers was made. The data set containing customer information and a data set containing contractual information was stored each day.

Table 1: Information and Characteristics of the Aggregated

S/N	Features (Column Names)	Value/ Data Type	Types(Values)
1	CustomerID	ID	Unique ID = 7043
2	Gender	Male	3555
		Female	3488
3	SeniorCitizen	No	5901
		yes	1142
4	Married	No	3641
		Yes	3402
5	Dependents	No	4933
		Yes	2110
6	tenure	Integer	Series data type
7	VAS Service	Yes	6361
		No	682
8	MultipleLines	No	4072
		Yes	2971
9	InternetService	4G-LTE	3096
		2G-3G	2421
		No	1526
10	OnlineSecurity	No	5024
		Yes	2019
11	OnlineBackup	No	4614
		Yes	2429
12	Social Media	No	4621
		Yes	2422
13	Streaming	No	4336
		Yes	2707
14	PostPrePaid	PayGo	5348
		Post Paid	1695
15	PaperlessBilling	Yes	4171
		No	2872
16	MonthlyRevenue	Int64	Series % data format
17	TCH Availability	object	Series data format
18	Churn	No	5174
		Yes	1869
	TOTAL		**7043**

The data sets are aggregated on customer ID in order to obtain the customer information data set.

A. Data Preprocessing

The data preprocessing phase or cleaning involve data preparation, data balanced and normalization. The data was cleaned from ambiguities, errors, missing data's, noisy data, redundancies and unique values that do not contribute much in predictive modeling. The data was further prepared along the explanatory variables continuum for modelling using multicollinearity. The scatter plot was used to reveal the probable result from the various data types (numeric, categorical, and binary). Data imbalanced was handled with the use of cost sensitive classifier. The result of these processes is shown in Table 1, Figure 1 and Figure 2

Include a note with your final paper indicating that you The data analysed contain 18 attributes and 7043 instances. The feature selection adopted in this study was the forward selection in the Best First feature selection method as well as matrixes that highlights features that may not affect the model. Churn the target class had two features or binaries that are categorical (Yes) and (No).

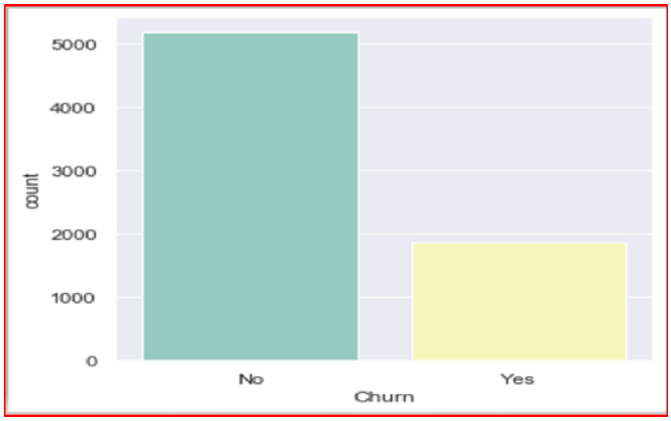

Figure 1: the Target Class (Churn Value)

The correlational matrix evaluated the attributes through the information gain measurement procedure as per the class value and diversifies the selection and ranking of attributes that significantly improves the computational efficiency and classification. The dataset contains 18 conditional features, along with one unique identifier. The output of the correlation, matrix is shown in Figure 2. The following display a vivid matrix using seaborn package. From the heat map a two-dimensional graphical representation of data the individual values that are contained in a matrix are represented

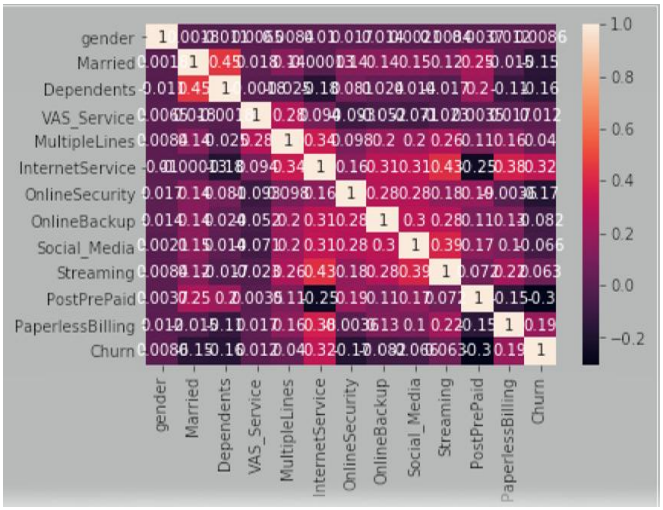

Figure.2: Correlation Matrix

The correlation matrix showed that the correlation coefficients between the variables. The cells show the value of the correlation between two variables. This summary revealed that Post Pre-Paid have a value closer to one than the other variables indicating that it had a higher correlation and is more a predictor of churn.

Exploratory data analysis applied before modeling in this analysis showed that the development of more complex predictive models may not be necessary in the given dataset. Histograms are typically plotted for numeric variables showed the distribution, for categorical variables and counts of categories. The graphical description below showed few overlapping density plots seen when the comparison of different numeric variables grouped by categorical (binary) dependent variables was used. The Two density plots for variables showed differences between churners and non-churners distinctively.

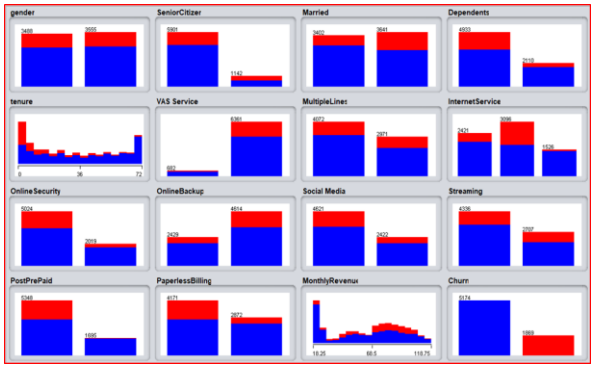

Figure 3: Graphical Attribute Distribution Source; Weka Researchers Dataset

The graphical distribution of attributes shown in Figure 3 showed minor overlap but differing distributions for each of the class values on each of the attributes. This is a good sign that probably shows the attributes and classes can be separated. Notwithstanding two of the attributes has some level of possible Gaussian-like distribution. This shows that more data may likely pull the distribution towards Gaussian. Some attributes such as Monthly revenue and tenure own to their discrete and real characters may tend towards Gaussian distributions with a skew or a large number of observations at the upper right end of the distribution. This also showed a visual indication that the classes are balanced. Furthermore, the categorical variables are graphically represented by bar plots. The numbers inside the bars represent absolute frequencies of a given category in the whole data set. From the graph a few deductions can be made, gender seems to have little effect on churners. VAS, POSTPre paid, Internet service, social media are possible and high attributes with high churners.

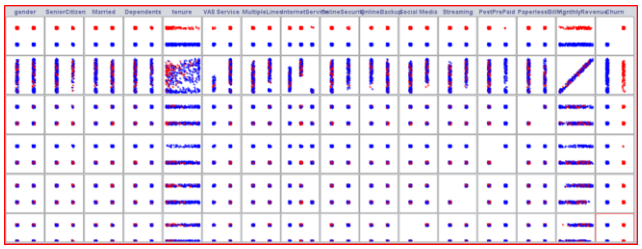

Figure 4: Attributes Interactions Source; Weka Researchers Dataset

Figure 4 shows graph with input variables. It is clearly shown that there is a separation between classes on the scatter plots. This suggests that linear methods, decision trees, etc will do well on this problem. It also suggested that advanced modeling techniques and ensembles may not be needed. More so, the developed enhanced predictive model may consist of simple classifiers.

B. Data Classification

Five classification algorithm or method was used. There were five classification techniques used with different feature selections and classifiers, which include Decision Tree, Naïve Bayes, and Decision Rules, logistic regression, random forest and Neural Network. Here the objects are categories according to their characteristics of the objects. The data was used to apply to unseen data. In order to evaluate classifiers performance for different schemes with their appropriate parameters, the measures of precision, recall, accuracy and F-measure, calculated from the contents of the confusion matrix, shown below was used. Each table shows Error rate and accuracy for each model.

C. Logistic Regression Classification

The Logistic regression was used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. The output of the logistic result is shown in Figure 5

```
Correctly Classified Instances        5617               79.7529 %
Incorrectly Classified Instances      1426               20.2471 %
Kappa statistic                          0.4432
Mean absolute error                      0.2768
Root mean squared error                  0.3723
Relative absolute error                 70.9947 %
Root relative squared error             84.3222 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.487    0.837      0.900    0.867      0.449    0.837     0.932     No
                 0.513    0.100    0.650      0.513    0.574      0.449    0.837     0.641     Yes
Weighted Avg.    0.798    0.384    0.787      0.798    0.789      0.449    0.837     0.855

=== Confusion Matrix ===

   a    b   <-- classified as
 4658  516 |   a = No
  910  959 |   b = Yes
```

Figure 5: Logistic Classification Model Result

From Figure 6 the kappa statistics for interrater reliability was 0.4432. The figure 0.4432 represents the extent in which the data collected are a correct representation of the variables measured. Giving the nature of the data source, the value of 0.4432 is adjudged reliable. The root relative squared error was 84.3%. The classifier had an accuracy of 79.8%. From the confusion matrix values, 7043 data points was used in the analysis, out of which 5617 were correctly classified and 1426 were misclassified. Combining the two metrics into a single metrics from the confusion metrics using the TP rate and FP rate for the classifier into a single graph with FPR values on the abscissa and the TPR values on the ordinate the ROC curve is plotted has shown in Figure 6.

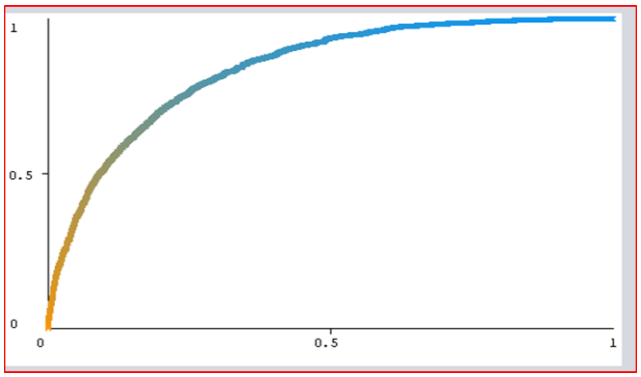

Figure 6: Logistic Regression ROC Curve

Using the metrics of AUC of the curve (AUROC) the following details was observed. The area under the curve (AUROC) with Given P(score(x+)>score(x−)), that is the probability that the classifier will rank a randomly chosen positive example, higher than a randomly chosen negative example had a value of 0.8359 showing that AUC is closer to 1.

```
Correctly Classified Instances        5575               79.1566 %
Incorrectly Classified Instances      1468               20.8434 %
Kappa statistic                          0.4095
Mean absolute error                      0.2084
Root mean squared error                  0.4565
Relative absolute error                 53.4539 %
Root relative squared error            103.4008 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.912    0.541    0.823      0.912   0.865      0.420    0.685     0.816     No
               0.459    0.088    0.652      0.459   0.539      0.420    0.685     0.443     Yes
Weighted Avg.  0.792    0.421    0.778      0.792   0.779      0.420    0.685     0.717

=== Confusion Matrix ===

   a   b   <-- classified as
 4717  457 |   a = No
 1011  858 |   b = Yes
```

Figure 7: Model Result for Support Vector Machine

Interpretation: From the Figure 7 the kappa statistics showed an interrater reliability value of 0.4095. The figure 0.4095 is adjudged reliable based on the source of the data as a correct representation of the variables measured. The root relative squared error was 103%, with accuracy of 79.2%. From the confusion matrix values, 7043 data points, 5575 were correctly classified and 1468 were misclassified. Combining the two metrics into a single metrics from the confusion metrics using the TP rate and FP rate for the classifier into a single graph with FPR values on the abscissa and the TPR values on the ordinate an ROC curve is plotted as shown in Figure 8.

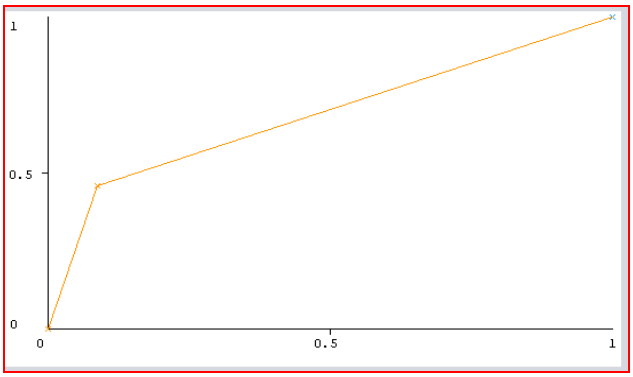

Figure 8: Support Vector Machine ROC Curve

Using the metrics of AUC of the curve (AUROC) the following details was observed. The area under the curve (AUROC) with Given P(score(x+)>score(x−)), the value that the probability of the classifier will rank a randomly chosen positive example, higher than a randomly chosen negative example was 0.685 showing that AUC is midway to 1.

D. Neural Network Classification

Figure 9(a) and Figure 9(b) shows a kappa statistics with value 0.4141, the root relative squared error value 82.2%, and an accuracy of 80.4%. The confusion matrix in the analysis, was 5460 correctly classified and 1583 misclassified. Using the TP rate and FP rate for the classifier an ROC is plotted and it's shown in Figure 10

```
Correctly Classified Instances        5460               77.5238 %
Incorrectly Classified Instances      1583               22.4762 %
Kappa statistic                          0.4272
Mean absolute error                      0.2697
Root mean squared error                  0.3892
Relative absolute error                 69.1725 %
Root relative squared error             88.1542 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.843    0.414    0.850      0.843   0.846      0.427    0.818     0.920     No
               0.586    0.157    0.575      0.586   0.581      0.427    0.818     0.604     Yes
Weighted Avg.  0.775    0.345    0.777      0.775   0.776      0.427    0.818     0.836

=== Confusion Matrix ===

   a   b   <-- classified as
 4364  810 |   a = No
  773 1096 |   b = Yes
```

Figure 9(a): Neural Network Model Analysis

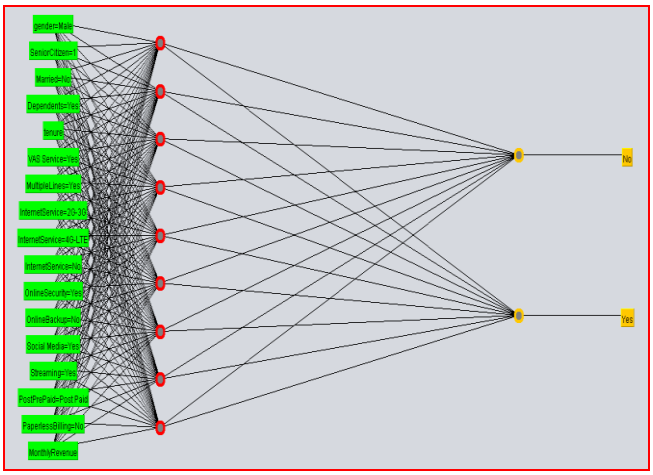

Figure 9(b): Neural Network Path

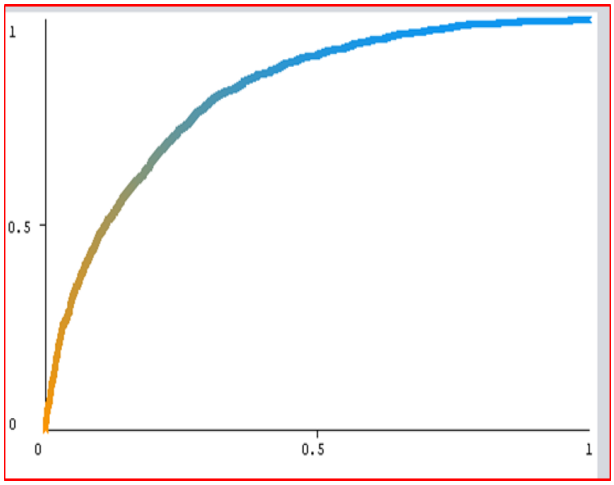

Figure 10: Neural Network ROC Curve

The area under the curve (AUROC) showed a value of 0.818 showing that the AUC comes closer to 1. Hence this model showed a higher AUC.

E. Decision Tree Classification

Figure 11 shows a kappa statistic with a value of 0.3972, the root relative squared error value of 91.2. %, and an accuracy of 78.2%. The confusion matrix in the analysis, was 5508 correctly classified and 1535 misclassified. Using the TP rate and FP rate for the classifier a ROC is plotted and it's shown in Figure 12. More so Figure 13 shows the decision path and the central attributes

```
Correctly Classified Instances        5508               78.2053 %
Incorrectly Classified Instances      1535               21.7947 %
Kappa statistic                          0.3972
Mean absolute error                      0.2767
Root mean squared error                  0.4025
Relative absolute error                 70.9551 %
Root relative squared error             91.1604 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.893    0.524    0.825     0.893    0.858      0.403    0.777     0.867     No
                 0.476    0.107    0.616     0.476    0.537      0.403    0.777     0.538     Yes
Weighted Avg.    0.782    0.414    0.769     0.782    0.772      0.403    0.777     0.780

=== Confusion Matrix ===

   a    b   <-- classified as
 4619  555 |    a = No
  980  889 |    b = Yes
```

Figure 11: Decision Tree Model Analysis

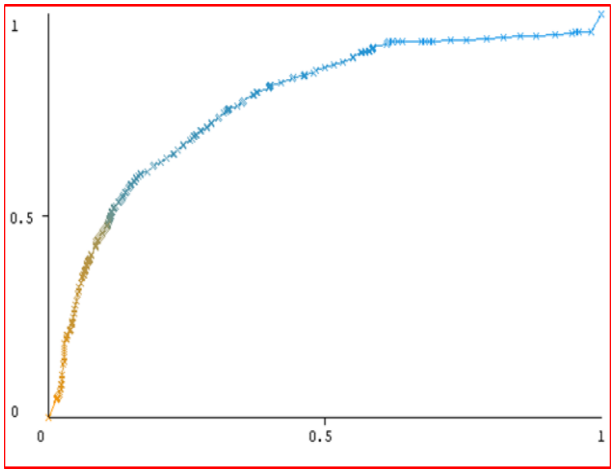

Figure 12: Decision Tree ROC Curve

The area under the curve (AUROC) had a value of 0.777 showing that the AUC comes closer to 1. Hence this model showed a higher AUC

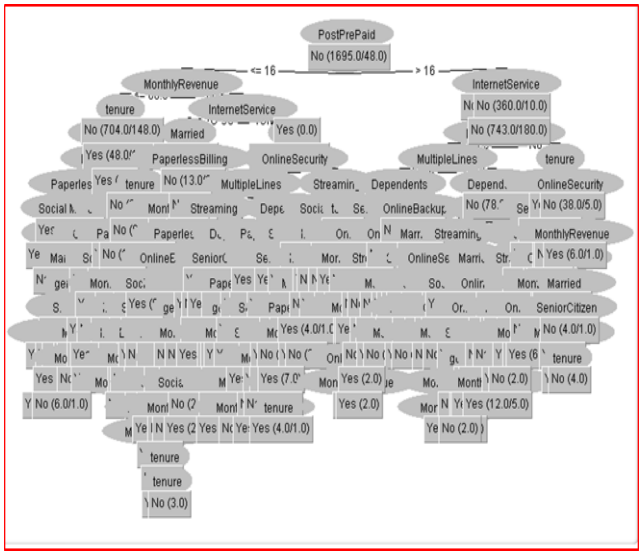

Figure 13: Decision Tree Path

F. Random Forest Classification

Figure 14 shows a kappa statistics value 0.3941, root relative squared 88.6%, and model accuracy of 78%. The confusion matrix in the analysis was 5489 correctly classified and 1554 misclassified. The TP rate and FP in ROC plot it's shown in Figure 15

```
Correctly Classified Instances        5489               77.9355 %
Incorrectly Classified Instances      1554               22.0645 %
Kappa statistic                          0.3941
Mean absolute error                      0.2752
Root mean squared error                  0.391
Relative absolute error                 70.5651 %
Root relative squared error             88.5656 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.887    0.519    0.826     0.887    0.855      0.399    0.804     0.912     No
                 0.481    0.113    0.606     0.481    0.536      0.399    0.804     0.590     Yes
Weighted Avg.    0.779    0.411    0.767     0.779    0.771      0.399    0.804     0.826

=== Confusion Matrix ===

   a    b   <-- classified as
 4590  584 |    a = No
  970  899 |    b = Yes
```

Figure 14: Random Forest Model Result.

International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-8, Issue-2, March 2020
https://doi.org/10.21276/ijircst.2020.8.2.1
www.ijircst.org

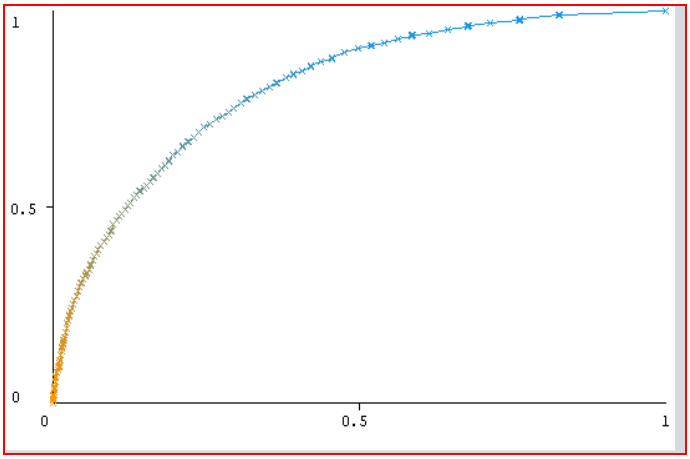

Figure 15: Random Forest Model Result

The area under the curve (AUROC) had a value of 0.804 showing that the AUC comes closer to 1.

G. Summary of Classification Algorithm Performance Metrics

Table 2 shows an extract of the confusion matrices used to compute the performance statistics – Accuracy, Sensitivity, Specificity and F-score. Due to the imbalanced nature of the data set it is not unexpected that accuracy is high for all models. The highest value of sensitivity achieved was by Neural Network, while Logistic regression had the highest for specificity (0.900) and accuracy (79.8). Sensitivity measures the ability of the model to catch customers who, in reality, left the company. Results on sensitivity showed that Logistic regression could catch 83.7 % and Neural Network 85.0% of such customers. Also, F-score, which combines Hit rate and Sensitivity into one measure, was highest for logistic regression and it could mean that less sophisticated model is more suitable for this business case. Thus, the model-based model to be compared with will be logistic regression.

Table 2: Summary of Algorithm Performance Metrics

Model Type	Accuracy	Sensitivity	Specificity		F-Score
Logistic Regression		79.8	0.837	0.900	0.867
Decision Tree		78.2	0.825	0.893	0.858
Neural Network		77.5	0.850	0.843	0.846
Random Forest		77.9	0.826	0.887	0.855
Support Vector Machine		79.2	0.823	0.912	0.865

H. Performance of the Classification Algorithms: Algorithm Evaluation

From Figure 16 all of the models have skill. All the models performed worse. Each model has a score that was worse than earlier performance. Decision Tree (78.29%) was significantly worse than logistic regression (79.87%). Also, Random Forest (78.03%) was worse than Logistic Regression as well as Neural Network (78.37%) at 5% level of significance. The results suggest Logistic Regression (79.87%) is better than neural network, SVM, Random forest. We can certainly infer that at 5% level of significance logistic regression can predict churn to an accuracy of 79.87%. This decision is predicated on the fact that aside having high values both logistic regression and SVM are much simpler model. To this end the Logistic Regression is selected as the model to be enhanced. Therefore, we used the Logistic Regression results as the test base for other models.

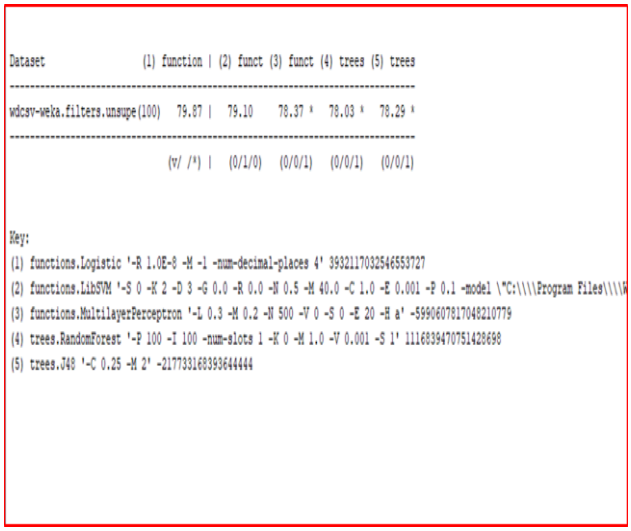

Figure 16: Comparisons of the Performance of the Classification Algorithms

In addition to the use of accuracy, Sensitivity, and specificity are also used to quantify the accuracy of the predictive models. The True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are the TP, FP, TN and FN in the confusion matrix. To access the specificity, ROC analysis, the ROC curve in the equations x=1− specificity (t) and y=sensitivity (t) is shown in Figure 17

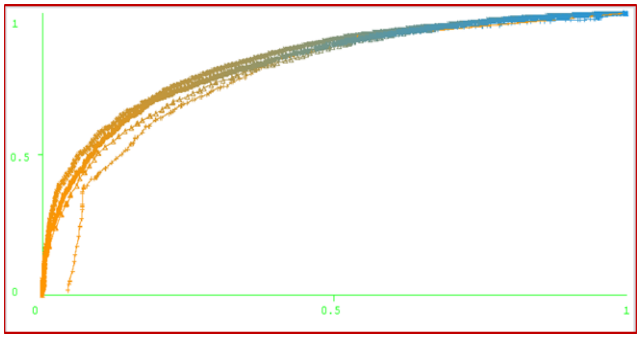

Figure 17: Comparisons of the Performance of Classifiers Implementation of Churn Management System

In this section the result of the trained logistic regression algorithm on unseen data is presented and the performance metric is shown in figure 18 and 19.

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.experiment.R
Analysing:  Percent_correct
Datasets:   1
Resultsets: 5
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       2/6/20, 3:23 PM


Dataset                   (4) functions.Logist
------------------------------------------------
TData                  (100)   79.87(1.49) |
------------------------------------------------
(v/ /*) |


Key:
(4) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
```

Figure 18: Model Description

Using some numbers and parameters the model was further describe in relation to performance of the model on unseen data. From the study the estimated accuracy of the model on unseen data was 79.87% with a standard deviation of 1.49%.

```
Correctly Classified Instances        5632               79.9659 %
Incorrectly Classified Instances      1411               20.0341 %
Kappa statistic                          0.449
Mean absolute error                      0.2759
Root mean squared error                  0.3711
Relative absolute error                 70.7566 %
Root relative squared error             84.0425 %
Total Number of Instances             7043

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area
               0.902     0.483     0.838      0.902     0.869      0.454    0.839     0.933
               0.517     0.098     0.655      0.517     0.578      0.454    0.839     0.646
Weighted Avg.  0.800     0.381     0.789      0.800     0.792      0.454    0.839     0.857

=== Confusion Matrix ===

   a    b   <-- classified as
 4666  508 |    a = No
  903  966 |    b = Yes
```

Figure 19: Enhanced Model Result

By running classification on the data using Logistic regression in WEKA we have the factors that influences churn. The model implemented and the parameters used is shown in Figure 19 and Figure 20. This mode and these parameters form the target variables, management need to focus on. This is used along with Logistic equation y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x)

```
Logistic Regression with ridge parameter of 1.0E-8
Coefficients...
                               Class
Variable                         No
==================================
gender=Male                  0.0205
SeniorCitizen=1             -0.312
Married=No                   0.0135
Dependents=Yes               0.2243
tenure                       0.0414
VAS Service=Yes              0.6745
MultipleLines=Yes           -0.3373
InternetService=2G-3G        0.1773
InternetService=4G-LTE      -0.8308
InternetService=No           0.9702
OnlineSecurity=Yes           0.4765
OnlineBackup=No             -0.1521
Social Media=Yes             0.0818
Streaming=Yes               -0.2663
PostPrePaid=Post Paid        1.1344
PaperlessBilling=No          0.3933
MonthlyRevenue              -0.0041
Intercept                   -0.2474
```

Figure 20: Enhanced Model Implementation

Where
y = Predicted Output
b0 = bias or intercept term
b1 = coefficient for the single input value (x).
x= input (eg gender or VAS, or etc)
That is
y = exp(-0.2474 + b1*x) / (1 + EXP(-0.2474 + b1*x)
y = -0.2474 + 0.0205x1 – 0.312x2 + …… + e Model Equation

IV.DISCUSSION OF FINDINGS

Churn Data are usually noisy and imbalanced, and multiple classifiers have their own limitations. So, the study employed an enhanced model design by using the Weka big data platform that allows for mining, processing and visualization to achieve higher AUC. The algorithms, logistic regression, neural network, Support vector machine, decision tree and random forest were analyzed in the study and it resulted in a high AUC. This means that, compared to random prediction, it is beneficial for telecom providers to implement one of the approaches from this study. Nevertheless, given the value from the enhanced logistic regression model in terms of performance statistics – Accuracy, Sensitivity, Specificity. The Logistic regression model better predict churn. More so, the result showed that internet service, types of contract entered, internet security were major factors that influence churn. The study used various methods of classifiers like earlier researchers [6] [7] [11] who used various methods (Artificial Neural Networks and Decision Trees). The findings from this study showed that the method used were all effective and can be equally strong to predict churn. In terms of variables that causes churn the findings of this study agree with [15] in that many of the variables have correlations with churn and affects it, however, internet services and types of contract affect churn the more. In studies where classifications were not carried out like the study by [7] that adopted a working methodology of Ensemble based Classifiers such as bagging, boosting and random forest, in contrast analysis to

common classifiers such as; Decision Tree, Naïve Bayes Classifier and Support Vector Machine. The study concluded that effectiveness is best with simple classifiers like SVM and logistic regression but the result from logistic regression showed that it was the best Classifier for the Churn Prediction Problem as compared to other models.

Furthermore, this study corroborates [13] study that adopted a predictive models and performance metrics and showed that the various churn prediction methods are of efficient performances. Also, from a qualitative approach the study correlates the findings of [9] that, the determinants of customer churn were varied and firms need to put up strategies to maintaining competitive position within the industry. Furthermore, experimental results confirm that the prediction performance has been significantly improved by using a large volume of training data, a large variety of features. In [18] study the quality of service is highly significant in tandem with, customer satisfaction, possession of superior technology, and cost of change and advertising.

Difference between this findings and other like [16] [5] [16] [21] [20] on logistic regression may be attributable to the rigor on data selection and cleaning as well as the number of classification employed. The study of [23] that adopted the use of data mining tools to select and classify features within the selected customer churn dataset pointed out that Logistic Model is the best method due to its accuracy using neutral network, while this however, do not reflect the outcome of this study and differs from [32] that study showed that the different data mining techniques improve the prediction accuracy of the models because of the combined advantages of the components. These findings differ slightly from the result of this study as only three data mining techniques were shown to have had different predictions. Thus, additional study needs to be carried out to determine and established causes of such discrepancies in result.

There is no doubt that the current churn prediction activities are working towards fulfilling the business needs in telecom companies. Yet it is clear that referring to data mining techniques and methods to identify the possible churners is more precise, unbiased and optimal in churn management. From a marketing viewpoint, the improvement in the classification of the customer, according to their churn-propensity, allows several strategies to be applied. The aftermath effect of this is that the result will allows for better allocation and budget to marketing department so that most likely leaving customer with the acceptable minimum set monetary value and those with propensity of leaving are targeted. By applying the predictive gain of the methods used in this study, there will be a direct effect and improvement on the efficiency of any possible marketing operations. From an operative advantage, the parallel implementation of data mining and enhanced model that result from it along with its methods will be appealing to practitioners that are under pressure to deliver Organizational goals at the set date.

V. SUMMARY AND CONCLUSION

It is imperative that mobile service providers deploy churn predictive models that can reliably identify customers who are about to leave, immediately after the possible churners are identified, intervention strategies should be put in place. To reduce churn, a careful selection of feature sets to be used should be done with this popular classification algorithm and the data cleaned. Also, customer churn behavior should be analyzed by using support vector machine. The attributes should be evaluated in line with the experimental results that showed that proposed model is better than the previous models in terms of the performance of ROC, sensitivity, specificity, accuracy and processing time.

The technique used outperformed and gave the most accurate results, as well as, with minimum risk error, therefore in order to maintain a loyal customer base for all service providers the model should be used to prevent the addition cost of acquiring new customers to retaining the old ones. Factors that causes churn was shown to be varied, but mainly internet security, types of contract and income, hence, telecommunication company must develop ways and means to enhance customer loyalty through the provision of quality services, lower tariffs, utility maintenance to avoid call drop outs, as well as, segmenting the market to target influential age groups.

Although the study focuses on organizational specific situations, but the overall aim of this study is to improve industrial performance with respect to churn prevention. Therefore, it is recommended that institutional measures should be put in place to overhaul telecommunication services provision and improved service delivery in the telecommunications industry especially in Nigeria

REFERENCES

[1]Adebiyi, S.O., Oyatoye, E.O. & Amole, B.B. (2016). Relevant drivers for customers churn and retention decision in the Nigerian mobile telecommunication industry. JC, 8(3), 52-67

[2]Adnan, A., Adnan, Z. Imran, A. Pir, S, Adeel, A., Basit, R., Ahmad, M. & Saif, M. (2017). Optimizing Coverage of Churn Prediction in Telecommunication Industry. International Journal of Advanced Computer Science and Applications, 8(5), 179 - 188

[3]Ahmad, A.K., Jafar, A. & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. JBD, 6(28), 1-24

[4]Albadawi, S. et al (2017). Telecom churn prediction model using data mining techniques. BUJICT, 10(2), 8-14

[5]Ali Dehghan, Theodore B. Trafalis (2012. Examining churn and loyalty using support vector machine. Scienceedu Press, (4), 153161

[6]Axelsson, R. & Notstan, A. (2017). Identify Churn. Unpublished Master's Thesis

[7]Azeem, M., Usman, M., & Fong, A. C. M. (2017). A churn prediction model for prepaid customers in

telecom using fuzzy classifiers. Telecommunication Systems. 66(4), 603–614

[8]Babu, S. & Ananthanarayanan, N.R. (2016). A review on customer churn prediction in telecommunication using data mining techniques. IJSER 4(1), 35-40

[9]Backiel, A., Baesens, B. & Claeskens, G. (n.d.). Predicting time-to-churn of prepaid mobile Telephone Customers using Social Network Analysis.

[10]Balasubramanian, M. & Selvarani, M. (2014). Churn Prediction in Mobile Telecom System using Data Mining Techniques. IJSRP, 4(4), 1-5

[11]Basha, S.M., Khare, A. & Gadipalli, J. (2018). Training and deploying churn prediction model using machine learning algorithms. IJERCSE. 5 (4): 59-64

[12]Bryan, E. & Simmons, L.A. (2009). Family Involvement: Impacts on Post-secondary educational success for first-generation Appalachian college students. JCSD, 50 (4), 391-405

[13]Canale, A. & Lunardon, N. (2014). Churn prediction in telecommunication industry: A study based on Bagging Classifiers. Cellegio Carlo Alberto 350

[14]Chuanqi, W., Ruiqi, L. Peng, W., Zonghai, C. (2017). Partition costsensitive CART based on customer value for Telecom customer Churn Prediction, Control Conference (CCC),

[15]Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. Decision Support Systems. 95(2), 27–36.

[16]Diaz-Aviles, E. et al (2015). Towards real-time customer experience prediction for telecommunication operators.

[17]Eria, K. & Marikannan, B.P. (2018). Systematic review of customer churn prediction in the Telecom. JATI, 2(1), 7-14

[18]Esteves, G.C. (2016). Churn Prediction in the Telecom Business. Unpublished Thesis.

[19]Faris, H. (2018). A hybrid swarm intelligent neural network model for customer churn prediction and identifying the influencing factors. Information Journal. 9 (288)

[20]Fei, T.Y., Shuan, L.H. & Yan, L.J. (2017). Prediction on customer churn in the telecommunications sector using discretization and Naïve Bayes Classifier. IJASCA, 9(3), 23-35

[21]Jae Sik, Lee & Chun Lee, Jin. (2006). Customer churn Prediction by Hybrid Model.

[22]Jamalian, E. & Foukerdi, R. (2018). A hybrid data mining method for customer churn prediction. ETASR 8(3), 2991-2997

[23]Joshi, S. (2014). Customer experience management: An exploratory study on the parameters affecting customer experience for cellular mobile services of a telecom company. Social and Behavioral Sciences, 2(133), 392 – 399.

[24]Karapinar, H.C., Altay, A., & Kayakutlu, G. (2016). Churn detection and prediction in automotive supply industry. IEEE, 1349-1354

[25]Kau, F.M., Masethe, H.D. & Lepota, C.K. (2017). Service Provider churn prediction for telecoms company using data analytics. WCECS 1-4

[26]Skymind (2019). A beginner's guide to neural networks and deep learning. Retrieved from https://skymind.ai

[27]Sufian, A., Khalid, L., Muhammad, M., & Kharbat, F. (2017). Telecom churn prediction model using data mining technique

[28]Tsymbalov, E. (2016). Churn Prediction for Game Industry Based on Cohort Classification Ensemble. MPRA 82871

[29]Umayaparvathi, V. & Iyakutti, K. (2016). A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. IRJET 3(4), 1065-1070

[30]Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory. 55(2), 1–9.

[31]Yu, R., An, X., Jin, B., Shi, J., Move, O. A., & Liu, Y. (2016). Particle classification optimization-based BP network for telecommunication customer churn prediction. Neural Computing and Applications, 1–14.

[32]Zhao, L., Gao, Q., Dong, X. J., Dong, A., & Dong, X. (2017). K- local maximum margin feature extraction algorithm for churn prediction in telecom. Cluster Computing. 20(2), 1401–1409

