Churn is characterized to be the movement of customers leaving the organization and disposing of the administrations offered by it because of the disappointment of the administrations as well as better offering from other network suppliers. 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. 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 purpose of this study was to ascertain the relevant drivers of customers’ churn and retention in the growing telecommunication industry especially in Nigeria and developed an enhanced predictive model to address earlier limitation of accuracy and improved churn prediction. The enhanced churn prediction model performed better than the unenhanced model. Logistic regression had better performance metric than other algorithms: neural network, Support vector machine, decision tree and random forest. Although, all the other algorithm had a high AUC but in terms of generality and simplicity logistic regression resulted in the highest AUC value on performance statistics – Accuracy, Sensitivity, Specificity. More so, the result showed that internet service, types of contract entered, internet security were major factors that influence churn.
Introduction:
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 Rijnen (2018), 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.

Literature Review:
Data mining helps to unearth information related to churning. Demographics, period of service, usage patterns and credit details are related variables that data mining tools use.

Data mining technologies emerged as effective tools for the Identification of the churn indicators in the data. The goal for predictions should then be to increase the weight of meaningful and useful knowledge (Sufian, Khalid, Muhammad and Kharbat (2017)).
Churn refers to the customer movement from one provider to another (Bott, 2014). Churn, also known as turnover, defection or attrition is the loss of clients or customers (Lomax & Vadera, 2017). To retain customers, an appropriate churn management strategy is vital. The first step to manage churn is to predict which customers are most likely to churn. The availability of data makes churn prediction possible. In Nigeria Mobile phone penetration is always on the upward swing, as the number of subscribers grew astronomically in 2017 resulting in 84 per cent penetration from 53 per cent in 2016 for both features and smart phones.

There are two types of churn. Voluntary churn and involuntary churn. In voluntary churn, the customer decides to cancel his contract and to switch to another provider. In involuntary churn involves the company to discontinues the contract itself, e.g. because of fraud or non-
payment of invoices. Voluntary churn can be sub-divided into: incidental churn and deliberate churn. Incidental churn occurs, not because the customers planned for it but because something happened in their lives e.g. a change in financial condition, change in location (Mahajan, Misra & Mahajan, 2015).

Figure 2.4 Types of Churners Source: Sufian, Khalid, Muhammad and Kharbat (2017)

Amin, Anwar, Adnan, Nawaz, Alawfi, Hussain and Huang (2017) used under sampling to balance churners with the non-churners in the training dataset. However, Gui (2017) disregarded Under sampling because a lot of meaningful information was lost with the application of this technique during his analysis of the telecom imbalanced data issue.

Figure 2.5 Churn Prediction Modelling Process (Source: Eria and Marikannan, 2018).

https://escipub.com/american-journal-of-computer-engineering/
Table 2.1 Review of Method and Countries Work on Churn Management

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Method</th>
<th># of customers</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharma and Panigrahi (2011)</td>
<td>Neutral networks</td>
<td>2427</td>
<td>India</td>
</tr>
<tr>
<td>Jadhw and Pawar (2011)</td>
<td>Neutral networks</td>
<td>895</td>
<td>South Korea</td>
</tr>
<tr>
<td>Ahn et al. (2006)</td>
<td>Logistic regression</td>
<td>10 000</td>
<td>Turkey</td>
</tr>
<tr>
<td>Gursoy (2010)</td>
<td>Logistic regression</td>
<td>1 000</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Oghojafor(2012)</td>
<td>Logistic regression</td>
<td>6 000</td>
<td>Nigeria</td>
</tr>
<tr>
<td>Olle and Cai (2014)</td>
<td>Logistic regression</td>
<td>2 000</td>
<td>Asia</td>
</tr>
<tr>
<td>Sebastian and Wagh (2017)</td>
<td>Decision tree</td>
<td></td>
<td>Malaysia</td>
</tr>
<tr>
<td>Khalida et al. (2010)</td>
<td>Decision tree</td>
<td>106 405</td>
<td>Europe</td>
</tr>
<tr>
<td>Kirui et al. (2006)</td>
<td>K-means Clustering</td>
<td>160 000</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Khan et al. (2010)</td>
<td>K-means clustering</td>
<td>2 685</td>
<td>Iran</td>
</tr>
<tr>
<td>Richter et al. (2010)</td>
<td>Social network analysis</td>
<td>16 000 000</td>
<td></td>
</tr>
<tr>
<td>Lu (2002)</td>
<td>Survival analysis</td>
<td>41 374</td>
<td>USA</td>
</tr>
</tbody>
</table>


A number of works have combined the application of some methodologies to determine the rate of customer churn. This study however, aim to combine the application of Random Forest, Decision Trees, Logistic Regression/Model Theory, Neutral Network as well as, Hybrid Model in predicting the rate of customer churn and its management. More so, a number of research studies have not focused on real time customer experience prediction for both telecommunication and all other sectors. Lastly, only a few works have done satisfactorily in the aspect of analyzing the correlation between Social Media Analytics (of interest and behavior) and how they are related to Train and Test Features.

Table 2.2 Review of Methodologies adopted by Researchers

<table>
<thead>
<tr>
<th>S/N</th>
<th>Methodologies Adopted</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Friedman Test</td>
<td>Hejazuna &amp; Kazem (2014).</td>
</tr>
<tr>
<td>10</td>
<td>Cross-validation, Monte Carlo Simulations</td>
<td>Asthana (2018).</td>
</tr>
</tbody>
</table>
Gaps Analysis: From the review above, several predictive modeling approaches were studied and their main characteristics detailed. Besides this overview, a more detailed explanation of some authors approaches to predict churn was presented. Most of the authors focus on training and testing one single model. Parameters were best fit to achieve their main objective. Thus, this study is proposing that using approaches that consist of testing multiple algorithms and comparing the result would bring interesting results to this study. Random Forest, Decision Tree, Naïve Bayes Classifier and Support Vector Machine was adopted by; Diaz-Aviles et al (2015), Mishra & Reddy (2017), Basha et al (2018), Ahmad, Jafar & Aljoumaa (2019), Lomax & Vadera (2017), Saini (2016), Pohjalainen (2016), Fei, Shuan & Yan (2017), Esteves (2016), Nonum, Ezema, & Hyacinth, (2018), Kau, Masethe & Lepota (2017).

Oversampling is however not an ideal technique for handling class imbalance issues given the huge nature of telecom datasets, as Idris, Iftikhar and Rehman (2017) challenged it and proposed PSO based Undersampling. As the class imbalance issue continued to prevail, Amin, et al. (2017) handled it using the random Undersampling method.

In telecom churn prediction datasets, the class of interest is the churners group which is always the minority thus causing a data imbalance problem (Zhu, Baesens & vanden Broucke, 2017; Amin, Gui, 2017). In their empirical comparison among the different class imbalance solutions, Zhu, Baesens and vanden Broucke (2017) stated that this problem affects the performance of the prediction models because the bias caused by the majority class also affects the model performance. Even though, ANNs are known to be stronger churn predictors than Decision Trees, they are known to have disadvantages such as early convergence or being stuck at local optima.

The literature examined regarding customer churn, showed that the majority of the related work focuses on applying only one data mining method to extract knowledge. Only a few authors like Buckinx and den Poel (2005), Verbeke, Dejaeger, Martens, Hur and Baesens (2012) focused on comparing multiple strategies to predict customer churn. Buckinx and den Poel (2005) consists on comparing three classification techniques: Logistic regression, automatic relevance determination (ARD) Neural Networks and Random Forests. The models were evaluated regarding their Area under curve (AUC) and percentage of correctly classified (PCC) instances, in both training and testing sets. The highest AUC value was obtained by the Random Forest model in both train and test sets, with values 0.8249 and 0.8319 respectively.

Sufian, Khalid, Muhammad and Kharbat (2017) study attempted to implement a new prediction model that uses data mining techniques on data obtained from a leading telecom company in Pakistan and then measure the generated benefits. The study did not touch on model that can assign suitable retention strategies for each churner type. While their study covered classification and clustering problem it did not provide any model that can suggest any suitable retention strategies as per churn cluster. That model is expected to try to provide new churn prediction retention model that will use the predicted and clustered data to assign a suitable retention strategy for each churner type. Every researcher has a list of research questions which need to be assessed – this can be done with research design. In this research, the model would predict why customers of telecoms churn, how they churn and most likely as a gap to previous research, when they churn.

It has been empirically understudied that, a number of studies have contributed ideas towards the areas of predictive models, the concept of churn and its causative factors and management, the concept of telecommunication industry as well as, the varying nature of data mining technique with respect to various firms, companies and organization outside Africa. But
little or no attention has been paid towards the peculiarity of the concept in Nigeria as a whole, which will be of immense benefit to service providers and operations in developing countries like Nigeria.

**Methodology**

The research design adopted to study this phenomenon is the predictive data mining technique. The main idea behind the model is to make Telcom industry aware of the best model available to predict churners so that the telecom industry management don’t waste their time and money on wrong customers, who were not even churners. Research design being the framework or methods and techniques chosen by a researcher to reasonably and logically solve a research problem, we thereby define the research problem.

The researcher identified users that are paying customers for at least 30 days at present and, with the help of recent historical data, predict the likelihood that the users will actively choose to discontinue the service sometime within the next 30 days. Furthermore, implementing and deploying this information to identify users that would be valuable to target in a marketing campaign.

The churn predictive model, algorithm and process below will be used to address the research problem. The model is expected to be used in analyzing Customer Relationship Management (CRM) data to deliver customer-based models that depict the probability that a customer will take a specific action.

**Algorithm:** The Algorithm to be use for this study is data mining. Data mining is the science of discovering through large data sets of patterns, potential knowledge, models, rules, insights, identify systematic relationships that are consistent between variables and retrieve meaningful information (Guo & Qin, 2017). Key data mining functionalities are as follows: multivariate statistical analysis (regression

Logistic Regression will be applied in probabilistic classification applications, estimated through the formula

\[ P : y = 1 \{ x_1, x_k \} = \frac{e^{b_0 + b_1 x_1 + \ldots + b_k x_k}}{1 + e^{b_0 + b_1 x_1 + \ldots + b_k x_k}} \]

Where:

- \( y \) is the target variable for each individual \( j \) (customer in churn label (0 or 1);
- \( b_0 \) is a constant
- \( b_j \) is the weight given to the specific variable associated with \( j \) \((j = 1, k)\);
- \( x_1, \ldots, x_k \) are the predictor variables for each customer \( j \),

Customer data sets are analyzed to form the regression equations. An evaluation process for each customer in the data set is then performed. A customer can be at risk of churn if the \( p \)-value for the customer is greater than a predefined value (e.g. 0.5).

From equation, the subset of attributes will be defined as a classification process of the universe into sets such that each object in a set cannot be distinguished from other objects in the set using only the attributes.

**Proposed Enhanced Churn Predictive Model for Telecommunication Industry:** Predictive modeling is essentially concerned with foreseeing how the customer will behave in the future by investigating their past behavior and anticipating customers who are likely to churn. Customer Churn Prediction (CCP) relies on data mining algorithms to develop models that classify telecom customers into churners and non-churners. To predict churners the predictive model will comprise of algorithm, infrastructure and data process.

- **Algorithm:** The Algorithm to be use for this study is data mining. Data mining is the science of discovering through large data sets of patterns, potential knowledge, models, rules, insights, identify systematic relationships that are consistent between variables and retrieve meaningful information (Guo & Qin, 2017). Key data mining functionalities are as follows: multivariate statistical analysis (regression
analysis), relationship mining (frequent pattern mining algorithms), clustering, and classification (decision trees, neural networks), and prediction and outlier detection.

**Infrastructure and Architecture:** Telecom datasets also come in the context of big data due to their large volumes, high velocity, veracity and variety as Ahmed and Maheswari (2017) acknowledged. There are complex big data tools and techniques used to draw meaningful and actionable insights from big data. New technologies developed to handle big data are Hadoop, Hbase and NoSQL, which are open source platforms (Storey & Song, 2017). Telecom datasets is required to be handled carefully because of their imbalanced nature, large volumes and high dimensional structure.

<table>
<thead>
<tr>
<th>DATA WAREHOUSE</th>
<th>DATA ACCESS</th>
<th>SECURITY AND OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data lifecycle</td>
<td>TOOLS: Spark, Map Reduce, SQL</td>
<td>-Zookeeper</td>
</tr>
<tr>
<td>-Atlas</td>
<td></td>
<td>-Cloud Break</td>
</tr>
<tr>
<td>-Sqoop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Flume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NFS</td>
<td>HDFS</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.1 Block Representation of Hortonworks Data Platform (HDP) Architecture

**Hortonworks Data Platform (HDP):** Hortonworks Data Platform (HDP) is chosen because it is a free and an open source framework. HDP is under the Apache 2.0 License. Furthermore, the big data in the study will be handled through other Apache HDP platform variety of open source systems and tools related to big data.

**Apache Flume:** Apache Flume is a distributed system used to collect and move unstructured (CSV and text) and semi-structured (JSON and XML) data files to HDFS. There are three main components in FLUME. These components are the data Source, the Channel where the data moves and the Sink where the data is transported.

Spark engine will be used to explore the structure of this dataset, it is necessary to make the exploration phase and make the necessary pre-preparation so that the dataset will become suitable for classification algorithms. At the spark engine phase, all data that reflect the customer behavior in general will be considered. Data sets related to calls, SMS, MMS, and the internet with all related information like complaints, network data, IMEI, charging, and other will be considered. The data will contain customer transaction before the prediction baseline. The
size of the data is expected to be in tens of terabyte. This will require more than traditional database to perform the feature engineering.

The flowchart described the design processes of the proposed model, starting with dataset which is a telecommunication customer records. This data is then prepared and feature variables classified are used for training the model. Initially, for each attribute, a threshold value is assigned. The attribute values of the training dataset are compared with the attribute’s threshold to declare that a customer will churn or not. Simple if…then …else rules are applied in this process. A model is then constructed for the training dataset. Finally, the model is then applied on the test dataset and the results are listed. The steps are repeated by varying the threshold values of the attributes selected to achieve better accuracy.

Figure 3.3, Flow chart of the proposed system Source: Researchers modification

Figure 3.4 Use case Diagram of the proposed system Source: Researchers modification Data Flow Diagram
There are five phases: 1) Preprocessing the input customer records, 2) Extracting the required features for developing churn models, 3) Construction models using different classifiers and cross validate the models, 4) Calculation of prediction accuracy and variable importance report, and 5) Providing customer retention policies to CRM executives.

**Database Design (ERD):** The database consists of relational datasets hence the use of RMySQL as a “storage house”. The data would be saved by running install.packages (RMySQL) code on R which also loads the DBI Library package automatically. Connecting to the database would require the con (RMySQL::MySQL).

**System Design:** The overall steps to take in churn management in telecom, are similar, it involves analyzing and following the following steps of customer behavior analysis model (Kaur & Vashisht, 2014). Churn modelling includes three major steps after data collection and before model deployment; data preparation, model training and model evaluation (Umayaparvathi & Iyakutti, 2016). In the industry the process of data mining takes six stages or steps. These 6 steps describe the Cross-industry standard process for data mining, known as CRISP-DM. In this study the process of data mining will be employed. The data mining process for the predictive model process are data collection, data preparation or preprocessing and Data classification. Put together the processes includes data collection, data cleaning and integration, data selection, and data transformation, pattern evaluation, and knowledge representation.

- Data Collection
- Data Cleaning and Integration,
- Feature Selection and Transformation.
- Data Classification
- Model implementation
- Model Evaluation
- Deployment
Figure 3.7 Framework of the Churn Prediction Modeling Process

**Telecom Dataset:** In order to examine our model, we will use customer data sets of a telecommunication operation (Globacom) to conduct the study. The datasets were gotten from Globacom Limited, a telecommunications company in Nigeria. There are many types of data used to build churn model. These types are Customer data that contains all data related to customer’s services and contract information. Demographic data like living location, age, etc. In addition to all offers, packages, and services subscribed to by the customer, customer GSMS, Type of subscription, birthday, gender, the location of living, etc. The other kind of data is Towers and complaints database, whose information of action location is represented as Digits mapping into giving the longitude and latitude, sub-area, area, city, and state, Complaints’ database, etc. provides. Network logs data Contains the internal sessions related to internet, calls, and SMS for each transaction in Telecom operator, etc. Call details records contain all charging information about calls, SMS, MMS, etc. Mobile IMEI information contains the brand, model, type of the mobile phone and if it’s dual or mono SIM device. According to the characteristics of the mobile user, customer’s consumption characteristics can be collected by analyzing consumption listing within 3 months. This data will have a large size and there will be a lot of detailed information on it.

**Data Collection:** The data to be use in this research will be collected from multiple systems and databases. The data are expected to be structured, semi-structured (XML-JSON) or unstructured (CSV-Text). The customer consumption listings will be selected from February 2019 to April 2019 and churn data for April 2019 as the model data. The sample data will contain 200,000 records. That means the size of the data is big and the format varied. Dealing with these kinds of data types is very hard without big data platform. Hence big data platform will be use to handle the data. Three month data will be used because it can compensate for churn behavior (as churner needs in average 3 months to be transformed
from non-churner to churner). Also, interval in Telecom Company is huge thus enabling data preprocessing. The data will also be linked to the detailed data stored in relational databases that contain detailed information about the customer.

**Data Preparation:** The essence of data preparation is to turn data into a state suitable for further analysis for model fitting (Coussement, Lessmann and Verstraeten, 2017). In other words data preparation is aimed at making data suitable for model training because the quality of data affects the model performance results.

In telecom churn prediction datasets, the class of interest is the churners group which always the minority is thus causing a data imbalance problem (Zhu, Baesens and vanden Broucke, 2017). Data imbalance affects the performance of the prediction models because the bias caused by the majority class also affects the model performance. The original data record will be in millions and the customer churn record will be in thousand. This differences usually creates a data imbalance problem. That is the small ratio of positive to negative data points and churn rates (1.5% to 9.0%) causes the dataset to significantly become imbalanced.

Data preparation is responsible for removing any bias in the data through class balances and other randomization procedures. Missing value imputation, data cleaning, transformation and general exploration is done in this phase (Federico, 2014). In addition, Pearson correlation coefficient for each pair of numerical features we removed one of the features in highly correlated pairs. In some cases, scatter plots will be used to investigate non-linear correlations between numerical features. Data transformations depend on the input data as well as the algorithms of choice.

The procedure to be choosing for data preparation in this study will involve handling missing and noisy data, data integration, dimension reduction, data normalization. Noisy data will be smoothed, and outliers excluded in the dataset because a model trained on such noisy data might face over fitting problems during the prediction of unseen data. Data will be integrated from all the systems, merge into one homogeneous dataset, and maintained by one needed keys to link these data from different systems.

**Data Classification**

**Handling Missing Data:** Data cleaning involves taking care of inaccurate and missing values, which are common in large datasets. There are various methods for handling missing values. Some of these include ignoring the tuple completely, replacing the value with a constant, the mean or median, or using some other technique to calculate the most probable value. In this respect domain knowledge will be used to carry out discrepancy detection on some of the numerical values. Tuples that were more or less guaranteed to be incorrect such as users over the age of 100 or younger than eight will be ignored. Tuples with missing values can be ignored because useful data in the tuples will be compensated with the size of the big dataset.

**Handling Unbalanced Dataset:** Unbalanced dataset is due to classification problem and occurs where the distribution of a class is not usually homogeneous with other classes. The dominant class is called the basic class, and the other class is called the secondary class. The data set is unbalanced if one of its categories is 10% or less compared to the other one (Chawla, 2005).

There are four widely-used methods to handle data imbalance: Not Balanced Up Sampling, Down Sampling and Weighted Instance. The first method directly train classifier using imbalanced churner and non-churner instances. The second method randomly copies the churner instances to the same number of non-churner instances. The third method randomly samples a subset of non-churner instances to the same number of churner instances. The fourth method assigns a proportion weight to each instance, where higher weights are assigned to churners and lower weights to non-churner instances.
churners. This study will use baseline features to evaluate different methods for data imbalance. Oversampling is however not an ideal technique for handling class imbalance issues given the huge nature of telecom datasets (Idris, Iftikhar & Rehman, 2017).

In order to handle the data imbalance problem, a randomly selected sample of loyal customers and customer churn from the processed data set will be used and set at ratio is 3:1. The ratio of 3:1 is a wide industrial practice. Also studies have shown that this ratio is good for customer churn prediction (Keramati, Jafari-Marandi, Aliannejadi, 2014). The preprocessing stage will exclude abnormal samples. This will result in the original data for the study. Abnormal samples excluded from the preprocessing stage, will result to a data sets for probably about some few thousand figures less than the original records collected. Also, the other record expected from the preprocessing stage is the customer churn records (expected to be about 40,000). Experienced and statistics set customer churn in the industry to about 3% of the whole data (Liu & Yongrui, 2015).

**Steps in using Classification to handle unbalanced dataset**

1. Divide data into two groups: the training group and testing group.
2. Training group will consists of 70% of the dataset and aims to train the algorithms.
3. The test group will contain 30% of the dataset and will be used to test the algorithms.
4. rebalance target class causing negative impact by balancing two classes of the training sample

Any customer attributes which do not affect customer churn prediction will be excluded. In most cases customer attributes such as customer ID, customer address etc, are usually excluded. It is expected that after exclusion of customer attributes that do not affect churn prediction about 15 customer attributes we remain to be used for further analysis.

**Data Normalization:** Secondly, all numerical features can be normalized. Normalization of the input can, for example, help make convergence of neural networks faster and for all distance based algorithms, normalization can prevent specific features from overshadowing others purely based on their scale. There are three normalization methods: min-max normalization, zero-mean standardization and normalization by decimal scaling. Zero-mean standardization will be used used in this study since it is robust against outliers. Zero-mean standardization means subtracting the mean and then dividing by the standard deviation.

**Data Transformation:** Data are transformed before feeding it to classifiers. The dataset contains both numerical and categorical features. Classification algorithms mainly handles numerical data. Therefore, all categorical features needed to be transformed. One-hot encoding is a method that takes care of this problem. Principal component analysis will also be used to transform the data.

**Feature Engineering:** Feature extraction or engineering is making a choice about how to present information to a machine learning model. The study will use it to omit extra and noisy features and those which have little information. Attributes that are highly interdependent are screened so that those with little contribution to the dataset are removed. Thus any off the limit data observed will require using human intuition, and empirical knowledge. Noise removal technique will be used to remove missing values, outliers and unique characters that existed in telecom datasets. Nevertheless techniques such as Univariate analysis that check for the uniqueness of every feature in the dataset and Bivariant/ Multivariant analysis that discover if there is a relationship between 2 or more features in the dataset will help in making informed decisions.

In all, any customer attributes which do not affect customer churn prediction will be excluded. In most cases customer attributes such as customer ID, customer address etc, are usually excluded.
excluded. It is expected that after exclusion of customer attributes that do not affect churn prediction about 18 customer attributes we remain to be used for further analysis. This 18 attributes will be compared using visualization with the features provided by the telecom company (Glo) or published on the internet. A sample will be decided upon and stored in CSV format. The list of attributes or dataset will consist of variable types, namely, nominal, continuous, discrete and Boolean. The unique key value of the dataset is the phone number of each user. While on the other hand, the prediction goal is to successfully classify the customer churn with only binary output; yes or no.

Table 3.1 Dataset Attributes

<table>
<thead>
<tr>
<th>S/N</th>
<th>Field</th>
<th>Data Type</th>
<th>Variable Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Customer ID</td>
<td>Integer</td>
<td>Continuous</td>
<td>Unique identity</td>
</tr>
<tr>
<td>2</td>
<td>Gender</td>
<td>String</td>
<td>Categorical</td>
<td>Male, Female</td>
</tr>
<tr>
<td>3</td>
<td>Senior Citizen (above 65 years)</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>4</td>
<td>Customers Partner</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Yes, No</td>
</tr>
<tr>
<td>5</td>
<td>Dependents</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Yes, No</td>
</tr>
<tr>
<td>6</td>
<td>Tenure (months with Network)</td>
<td>Integer</td>
<td>Continuous</td>
<td>Number of months</td>
</tr>
<tr>
<td>7</td>
<td>Telecom Service</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Data or Voice line</td>
</tr>
<tr>
<td>8</td>
<td>Internet Service</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Is it Data only or not?</td>
</tr>
<tr>
<td>9</td>
<td>Phone service</td>
<td>Integer</td>
<td>Continuous</td>
<td>Duration of contract</td>
</tr>
<tr>
<td>10</td>
<td>Multiple Lines (number)</td>
<td>Integer</td>
<td>Continuous</td>
<td>Yes, No</td>
</tr>
<tr>
<td>11</td>
<td>Device Protection</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Device insured or not</td>
</tr>
<tr>
<td>12</td>
<td>Contract</td>
<td>Integer</td>
<td>Continuous</td>
<td>Duration of contract</td>
</tr>
<tr>
<td>13</td>
<td>Paperless Billing</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Is the address valid</td>
</tr>
<tr>
<td>14</td>
<td>Payment Method (Bank, card)</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Debit or Cash\</td>
</tr>
<tr>
<td>15</td>
<td>Monthly Charges (ARPU)</td>
<td>Integer</td>
<td>Continuous</td>
<td>Subscription</td>
</tr>
<tr>
<td>16</td>
<td>Total Charges</td>
<td>Integer</td>
<td>Continuous</td>
<td>Total Usage</td>
</tr>
<tr>
<td>17</td>
<td>Above Credit-Limit</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>18</td>
<td>Churn</td>
<td>Boolean</td>
<td>Categorical</td>
<td>Yes or No</td>
</tr>
</tbody>
</table>

**Classification Techniques:** The modeling phase then sees prepared data being fed into various modeling algorithms with the parameters usually optimized using cross validation. The solution proposed divided the data into two groups: the training group and the testing group. The training group consists of 70% of the dataset and aims to train the algorithms. The test group contains 30% of the dataset and is used to test the algorithms. As some methods have specific requirements for the form of input data, it is often necessary to go back to the data preparation phase. This phase concentrates on building a model that can predict the churn rate of different types of customers and identify churn patterns and what are the best actions that we can do to retain this segment of customers. Generally, classification can be referred to as a process to categorize objects according to the characteristics of the objects. In data mining,
classification is defined as an analyzing task for a set of pre-classified data objects to study a model or function that can be used or applied to unseen data objects before being placed into one of several predefined classes. Support Vector Machine, Decision trees, Random Forest, logistic regression, and Neural Network will be used for classification. These classifiers are chosen because studies have shown that they effective in churn prediction. The confusion matrix of the generated models. Error rate and accuracy for each model will be developed and used to compare their performance. In this section, logistic regression, random forest, neural networks, decision tree and will be used for model development and comparison purposes.

**Process of using Decision Tree**
- Train Decision Tree algorithm
- Optimize the depth and the maximum number of nodes hyperparameters.
- Experiment with several values

Random Forest (RF) classifier is chosen to make predictions. RF is a supervised learning method that uses the bootstrap aggregating (bagging) technique for an ensemble of decision trees. Given a set of training instances,
\[ x_m = [x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N], \]

With class labels \( y_m = \{\text{non-churner} = 0, \text{churner} = 1\} \),
RF fits a set of decision trees \( f_t, 1 \leq t \leq T \) to the random subspace of features.
The label prediction for test instance \( x \) is the average of the predictions from all individual trees,
\[ y = \frac{f_t(x)}{\text{ }} \]

The step to take when carrying out random forest ([Zumel and Mount, 2014])
- In a decision tree, split points are chosen by finding the attribute and the value of that attribute that results in the lowest cost.
- the best split point,
- evaluate the cost of each value
- Optimized the number of trees of hyperparameter
- build model by experimenting with changing values of parameters, eg, 100, 200, 300 trees.
- Find best result on number of trees
- Draw a bootstrapped sample from the training data.
- For each sample, create a decision tree, and at each node of the tree:
- randomly select a subset of \( m_{try} \) variables from the \( p \) total features (typically, the number of candidate predictors is \( \approx \sqrt{p} \)),
- pick the best variable and the best split from that set of \( m_{try} \) variables,
- continue until the tree is fully grown.

**Logistic Regression** Logistic regression (LR) is a regression analysis widely applied in probabilistic classification applications estimated through the formula.
\[ P(y = 1|x_1, \ldots, x_k) = \frac{e^{b_0 + b_1x_1 + \ldots + b_kx_k}}{1 + e^{b_1x_1 + \ldots + b_kx_k}} \]

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). Logistic regression method has become an integral component of each data analysis related to the description of the relationship between the response variable and one or more explanatory variables. Logistic Regression according to Kotu and Desphande (2015), is used to predict the value of the target variable in the form of binary (0 or 1) using numeric variable input. In other words, the logistic regression model predicts \( P(Y=1) \) as a function of \( X \).
The steps to take in carrying out logistic Regression includes inputing variables, Predict variables (desired Target), Using binary digit, Y(success) = 1, No(failure) = 0, check for variables in dataset that are outliers, with many categories for reduction for better modelling, data exploration, checking for imbalance, data visualization; check for good/strong predictor of the outcome(target) variables, Create dummy variables , that is variables with two values, zero and one, train data with under sampling.

Neural network: neural Network is an approach to predict the customer churn and found that neural networks of medium size their performance is better, when different neural network's topologies were investigated. Neural Networks will also be used because of its performance consistence with large datasets. Neural Networks work in a way that converts data into a brain neuron system (Monani et al., 2016).Neural networks have wide range of applications for prediction and classification problems in industrial and business domains. NN are also more capable for large datasets compared to other techniques. This justifies their increased adoption in the telecom churn prediction where large datasets are the order of the day.

Model Implementation: The model will be executed using the training data. The predictive models trained data using logistic regression, Neural Networks, decision tree and random forest algorithms and are used to predict churn of customers in the testing data set.

Training and Test Set: To test and train the model, the sample data is divided into 70% for training and 30% for testing. The training set is used to train the models which are later tested with the validation set to evaluate their performance accuracy and reliability. On the other hand when constructing a classification model, data will be set aside for testing. The test data will not be used in any way in the training phase so as not to introduce bias, and thus get a result that is too optimistic. There are a few different ways to choose training and test sets to evaluate a model and get an accurate measure of a model's performance.

The first way is using 1 K–fold Cross-validation, the second is hold-out Method (Schill, 2018). The traditional cross-validation method does not compared to the practicality and fitting pattern of the simple hold-out technique. Thus, the study will make us of the hold – out – method. Using the hold – out – method will involve setting aside a third part of the data for validation set; this data set will be used for tuning various hyper parameters affecting the model (Witten, Frank, Hall &Pal, 2016). After the data is prepared, it will be split into a training and a validation set.

The models will be trained on the training set, and the hyper parameters tuned based on a model's performance on the validation set. The feature selection process will also be carried out on the training set. The model that performed best on the validation set will in the end, run on the test set to see how well it is expected to do in general. These datasets will be constructed by randomly partitioning all users (subject to the constraints in the selected data) into three disjoint subsets. Data for the training, validation and test sets at three different points in time will be collected. This will avoid a biased result and will result in the models given better conditions to perform well on general data later.

Model Evaluation

Performance Metrics: To compare multiple classification models, metrics are used to evaluate how well they will likely perform on unseen data and how well equipped they are for solving the problem at hand. Some metrics are:

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \ldots \quad \ldots \quad \ldots \quad 3.4
\]

The proportion of all tuples that were classified as positive that are actually positive.

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \ldots \quad \ldots \quad \ldots \quad 3.5
\]

Recall and true positive rate (TPR) is the proportion of positive tuples that were correctly classified.

False positive rate (FPR)
False Positive Rate (FPR) = FP/FP+TN …….. 3.6

FPR is the proportion of negative tuples that were incorrectly classified.

F1-score = 2 - precision - recall / precision + recall …….. 3.7

True positives (TP): Number of correctly classified tuples from the positive class.
False negatives (FN): Number of incorrectly classified tuples from the positive class.
False positives (FP): Number of incorrectly classified tuples from the negative class.
True negatives (TN): Number of correctly classified tuples from the negative class.

Accuracy: Proportion of correctly classified tuples.

F1-score is defined as the harmonic mean between the two (Han et al., 2011), and is what we used as the primary metric for evaluating models in this study. The definition of precision and recall makes F1-score dependent on which class is defined as positive. High precision is obviously relevant to churn prediction since false positives come with a cost. However, recall is also crucial since finding a decent amount of possibly churning users is what the problem is all about.

**Model Validity**

This is to check with the 30 percent test data if the model accuracy is near optimal. To check if a customer from the “test” would churn or not. Model evaluation process will be carried out using confusion matrix that computes the precision and recall of the results. For the purpose of evaluating each model objectively, the accuracy of each model index will be calculated.

The equation below estimates the overall performance of the model.

OA, % = ncp*100%/N ….. 3.8

This formula attains the overall accuracy (OA) of the model by finding the number of correctly predicted outputs, cp n divided by the total number of samples, N.

The formula to compute the accuracy for the true positive and true negative categories are

True positive rate = number of correctly predicted churn customers * 100% ….. 3.9

Number of correctly predicted non churn customers * 100% ….. 3.10

True Negative Rate = total number of non-churn customers

Confusion matrix will be used to test the validity of the model. Confusion matrix will measure the validity of the model with binary target values (0 or 1). Each value of confusion matrix is obtained from the results of model testing.

<table>
<thead>
<tr>
<th>Predicted Condition</th>
<th>Not Churn</th>
<th>Churn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Churn</td>
<td>TN (True Negative)</td>
<td>FN (False Negative)</td>
</tr>
<tr>
<td>Churn</td>
<td>FP (False Positive)</td>
<td>TP (True Positive)</td>
</tr>
</tbody>
</table>

The values True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) will be used as TP, FP, TN and FN in the confusion matrix. In addition, Area under the Curve (AUC), sensitivity, and specificity will also
be used to quantify the accuracy of the predictive models.

**Sensitivity, Specificity and Accuracy:**
Sensitivity measures the ability of the model to catch customers who, in reality, left the company. Sensitivity to be used will be the proportion of positive cases which are predicted to be positive. The specificity will be the proportion of negative cases which are predicted to be negative. To assess the accuracy of a classifier independent of any threshold, the ROC analysis will be used.

The horizontal axis and the vertical axis, specificity, sensitivity, Accuracy is given by

\[ X = 1 - \text{specificity} \quad (t) \quad \ldots \quad \text{Horizontal axis} \quad \ldots \ldots \quad 3.11 \]
\[ Y = \text{sensitivity} \quad (t) \quad \ldots \quad \text{Vertical axis} \quad \ldots \ldots \quad 3.12 \]
\[ \text{Accuracy} = \frac{TP+TN}{(TP+FN+FP+TN)} \quad \ldots \ldots \quad 3.13 \]
\[ \text{True Positive rate (TPR) Specificity} = \frac{TN}{(TN + FP)} \quad \ldots \ldots \quad 3.14 \]
\[ \text{True Positive rate (TPR) or Sensitivity} = \frac{TP}{TP + FN} \quad \ldots \ldots \quad 3.15 \]

**Summary of Confusion Matrices Churn Prediction Model Measurement indexes**

<table>
<thead>
<tr>
<th>Item</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier Name</td>
<td>Decision Tree, Random Forest, Logistic Regression and</td>
</tr>
<tr>
<td>Matrix</td>
<td>TP, TN, FP, FN, Accuracy, Error rate, Specificity, sensitivity</td>
</tr>
</tbody>
</table>

**Deployment:** Aside academic purposes, the model is supposed to be used in production. Deployment or knowledge representation entails presenting the information gained through the classification process to the end user. For this study, the deployment phase will be a report presentation. Deployment will involve summarizing the insights and outcomes gains as well as reviewing the result. More so, research in the telecom market is to help companies make more profit by laying bare insights for retaining customers. Thus it is expected that the prediction models will have high AUC values. Results obtained can be used by the CRM department to establish whether to maintain a potential churner or let go. The average predictive performance will be obtained using different methods. The outcome will be compare to see which one method outperforms the other methods.

**Development Tools:** The research makes use of five classification techniques named AVM, SVM, Random Forest, Logistic Regression and Decision Trees to build a model for predictions using WEKA. The data sets to be used will be from Globacom Nigeria, reports for broadband and calls, internet penetration from operators for behavioral patterns of customers. The Work system was developed using SQL for datasets scrapping from database, php programming tool. The datasets were gotten from Globacom Limited, a telecommunications company in Nigeria.

**Conclusion**
The importance of this type of research in the telecom market is to help companies make more profit by laying bare insights for retaining customers. It has become known that predicting churn is one of the most important sources of income to telecom companies. Hence, this research aimed to build a system that predicts the churn of customers in telecom companies. These prediction models need to achieve high AUC values. To test and train the model, the sample data is divided into 70% for training and 30% for testing.
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 the 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.

REFERENCE
12. Bradbury, B.L. & Mather, P.C. (2009). The Integration of First-year, first-generation college students from ohio Appalachia. NASPA, 46 (92) 258-281