DATA MINING MODEL FOR DECISION MAKING IN TELECOMMUNICATION INDUSTRY

CASE STUDY OF EMERGING MARKET TELECOMMUNICATION SERVICES (EMTs)

BY

CHINONSO JECINTA ONWUEGBUCHULAM.

20124761688

A THESIS

SUBMITTED TO THE POSTGRADUATE SCHOOL

FEDERAL UNIVERSITY OF TECHNOLOGY, OWERRI

IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE AWARD OF THE DEGREE (MASTER OF SCIENCE), M.SC IN INFORMATION TECHNOLOGY

JANUARY, 2017

CERTIFICATION

This is to certify that this work "**Data Mining Model for Decision Making In Telecommunication Industry; Case study of Emerging Market Telecommunication Services (EMTs)" was** carried out by **Onwuegbuchulam, Chinonso Jacinta with (Reg. No: 20124761688) in partial fulfilment of the requirements for the Award of degree Master of Science (M.Sc.) in Information** Management Technology, School of Management Technology, Federal University of Technology, Owerri .

..... 31/03/17

Dr. (Mrs.) U. F. Eze Date
(Supervisor)

..... 04/04/17

Prof. B.C. Asiegbu Date
(HOD, IMT)

..... 6/4/17

Prof. C. C. Ibe Date
(Dean, SMAT)

....

Prof. (Mrs.) N.N. Oti Date
(Dean, Postgraduate School)

..... 12/01/17

Prof. (Mrs.) F.N. Ogwueleka Date
(External Examiner)

DEDICATION

This project is dedicated to Almighty God and to my beloved parents Mr. and Mrs. Onwuegbuchulam and my relations; Mr Ugochukwu, Mr and Mrs Uzo, Mr and Mrs Chibueze Onwuegbuchulam. Also, to my great soul mate Ifechi Olisakwe.

They are always there for me.

ACKNOWLEDGEMENTS

Praise the Lord, my soul! O Lord, my God, how great you are! You are clothed with majesty and glory;

Ps. 104 vs 1.

Prof. B.C. Asiegbu you are the key to success of this project completion. I will always be grateful, it was pleasure working with you. Dr. Mrs U. F. Eze as my supervisor: despite your tight schedule you still help me to achieve what is expected of me. I am very grateful. I wish to thank the entire scholars that their work is cited in my project and also those people who have directly influenced my professional life. Particular thanks go to Ajunwa Innocent and Chikanma Sonia, Mr. Philip Akhigbe my dealer specialist, Mr. Stanley Diala, Charles Ezeuko, Oluchi Nlemonisa and all my friends whose advice motivated me during my project work.

TABLE OF CONTENTS

CHAPTER FOUR

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This work present a data mining model for decision making in telecommunication industry. Nowadays, most sales and marketing organization, around the world face escalating competition which is forcing them not only to aggressively market special pricing programs aimed at retaining existing customers and attracting new ones but also for effective management and allocation of resources, goods and services. In this project work, several literature works were reviewed in other to identify the key performance indicators that makes data mining technique model a powerful tool to be used for carrying out data analytic tasks that will achieve a good result to support decision making in telecommunication providers. Twelve performance indicators were identified in this study. They are accuracy, interpretability, presentation quality, accessibility, consistency, easy to use, precise, concise, robustness, speed (response in time), reliability and unambiguity. The identified performance indicators were assessed and measured against existing system method use for data analytic in telecommunication providers such as Mobile Telephone Networks (MTN), Global communication (GLO), Airtel and Emerging Market Telecommunication services (EMTs). Delphi method were used as the standard of measurement in the existing system. Six performance indicators were found to be weak indicator. They are Presentation quality, accessibility, easy to use, precise, robustness and speed (response in time). The enhanced system was developed which proves to be stronger than the existing system using also Delphi method as a standard of measurement. The model developed is able to make a sales forecast of the year 2016 performance whereas the training data used for model exploratory analysis range from 2008 to 2015. The financial analysis of the proposed model produce a positive Net present value and Return of Investment is 65.85% which is very outstanding and worthy to be considered.

KEYWORDS: data mining, model, telecommunication industry, performance indicators, data mining techniques.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the Study

Data mining promises to make life easier for business decision makers, analysts and consultants. Data mining can be seen as the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules (Meens, 2012). Ahaiwe and Ukwandu (2012) defined data mining as a computer- assisted process of digging through and analyzing enormous set of data and then extracting the meaning of the data. Hofer et al. (2008) defined data mining as knowledge discovery using a sophisticated blend of techniques from traditional statistics, artificial intelligence and computer graphics. Beyond just predicting future performance, data mining helps to identify relationships in the data that might otherwise be hidden; for example calculating the odds that certain products might sell, explain some observed event or condition, such as why sales of pickup trucks have increased in Northern Nigeria. Additionally, to analyze data for new or unexpected relationships, such as what spending are likely to accompany credit card fraud (Meen, 2012). According to Coenen (2011) data mining is also sometimes called Knowledge Discovery in Databases (KDD). Knowledge discovery differs from traditional information retrieval from databases.

In traditional DBMS, database records are returned in response to a query; while in knowledge discovery, what is retrieved is not explicit in the database, rather, it is implicit patterns. The process of discovering such patterns is termed data mining. It requires intelligent technologies and the willingness to explore the possibility of hidden knowledge that resides in the data. Data mining finds these patterns and relationships using data analysis tools and techniques to build models (Dilly, 1995).

Data mining techniques use advanced statistical methods, such as cluster analysis, classification analysis, regression analysis, association rule learning, prediction, sequential patterns and sometimes employ artificial intelligence or neural network techniques (Thearling, 2005; Rijmenam, 2014). There is a great deal of overlap between data mining and statistics. In fact most of the techniques used in data mining can be placed in a statistical framework (Oracle Help Center, 2016). However, data mining techniques are not the same as traditional statistical techniques (Oracle Help Center, 2016). Traditional statistical methods, in general, require a great deal of user interaction in order to validate the correctness of a model. As a result, statistical methods can be difficult to automate. Moreover, statistical methods typically do not scale well to very large data sets. Statistical methods rely on testing hypotheses or finding correlations based on smaller, representative samples of a larger population. Data mining methods are suitable for

large data sets and can be more readily automated. In fact, data mining algorithms often require large data sets for the creation of quality models (Oracle Help Center, 2016). The major classes of data mining techniques/statistical methods are as follows (Rijmenam, 2014):

1. Classification analysis.
2. Clustering analysis.
3. Regression analysis.
4. Association Rule Learning.
5. Sequence Discovering
6. Visualization.

Classification analysis: Classification analysis is a systematic process for obtaining important and relevant information about data, and metadata – data about data. The classification helps identifying to which of a set of categories different types of data belong. Classification analysis is closely linked to cluster analysis as the classification can be used to cluster data (Rijmenam, 2014).

Clustering analysis: Clustering analysis is the process of identifying data sets that are similar to each other to understand the differences as well as the similarities within the data. Clusters have certain traits in common that can be used to improve targeting algorithms. For example, clusters of customers with similar buying

behavior can be targeted with similar products and services in order to increase the conversation rate (Rijmenam, 2014).

Regression analysis: Regression analysis tries to define the dependency between variables. It assumes a one-way causal effect from one variable to the response of another variable. Independent variables can be affected by each other but it does not mean that this dependency is both ways as is the case with correlation analysis. A regression analysis can show that one variable is dependent on another but not vice-versa. Regression analysis is used to determine different levels of customer satisfactions and how they affect customer loyalty and how service levels can be affected by for example the weather (Rijmenam, 2014).

Association Rule Learning: Association rule learning enables the discovery of interesting relations (interdependencies) between different variables in large databases. Association rule learning uncovers hidden patterns in the data that can be used to identify variables within the data and the co-occurrences of different variables that appear with the greatest frequencies. Association rule learning is often used in the retail industry when finding patterns in point-of-sales data. These patterns can be used when recommending new products to others based on what others have bought before or based on which products are bought together (Rijmenam, 2014).

Sequence Discovering: Sequence discovery is the identification of associations or patterns over time as defined by (Bersonet et al., 2000). Its goal is to model the states of the process generating the sequence or to extract and report deviation and trends overtime (Mitra et al., 2002). Common tools for sequence discovery are statistics and set theory. Sequential patterns are useful method for identifying trends, or regular occurrences of similar events. For example, with customer data you can identify that customers buy a particular collection of products together at different times of the year. In a shopping basket application, you can use this information to automatically suggest that certain items be added to a basket based on their frequency and past purchasing history.

Visualization: Shaw et al. (2001) article defined visualization as the presentation of data so that users can view complex patterns. It is used in conjunction with other data mining models to provide a clearer understanding of the discovered patterns or relationships. Examples of visualization model are 3D graphs. Dashboard is popular visualization tool use in most organization for results display (Hofer et al., 2008). Dashboards are often used to provide an information system in support of business performance management (BPM). BPM system allows managers to measure, monitor, and manage key activities and processes to achieve organizational goals. Often the top dashboard, and executive dashboard, is based

on a balanced scorecard, in which different measures show metrics from different processes and disciplines, such as operations efficiency, financial status, customer service, sales, and human resources etc. Each display of a dashboard will address different areas in different ways. For example, one display may alerts about key customers and their purchases. Another display may show key performance indicators in sales, with "stoplight" symbols of red, yellow, and green to indicate if the measures are inside or outside tolerance limits. Each area of organization may have its own dashboard to determine health of that function (Hofer et al., 2008).

Additionally, prediction in some data mining articles is defined as data mining techniques. According to Brown (2012), Prediction used in combination with the other data mining techniques, involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event. Using the historical sales report of a telecommunication company, for example, you might combine decision tree analysis of a State past sales transactions with classification and historical pattern matches to identify whether the sales of that State is increasing or decreasing. The choice of a particular combination of techniques to apply in a particular situation depends on the nature of the data mining task, the nature of the available data, and the skills and preferences of the data miner according to Berry and Linoff (1999); Carrier and Povel (2003); Grover and Mehra (2008). Data mining tasks can be directed

and undirected as is stated by Berry and Linoff (2004) and Bendoly (2003). In other literature, Lustig et al. (2010) there is made distinction between prescriptive and descriptive data mining tasks, where predictive is similar to directed and descriptive to undirected. Directed data mining attempts to explain or categorize some particular target field such as income or response. Directed data mining tasks are classification, forecasting, regression and description and profiling. Undirected data mining attempts to find patterns or similarities among groups of records without the use of a particular target field or collection of predefined classes. Undirected data mining tasks are association, clustering, description and profiling (Berry and Linoff 2004; Bendoly, 2003).

DATA MINING MODELS: Regardless of the technique used, the real value behind data mining is modeling (SQL Server, 2016). The process of building a model based on user-specified criteria from already captured data. Once a model is built, it can be used in similar situations where an answer is not known. For example, an organization looking to acquire new customers can create a model of its ideal customer that is based on existing data captured from people who previously purchased the product. The model then is used to query data on prospective customers to see if they match the profile. Modeling also can be used in audit departments to predict the number of auditors required to undertake an audit plan based on previous attempts and similar work. Data modeling refers to a

group of processes in which multiple sets of data are combined and analyzed to uncover relationships or patterns (Optimazition Group, 2016).

The goal of data modeling is to use past data to inform future efforts. Data mining is a step in the data modeling process (Optimazition Group, 2016). In data mining, you search for valuable and relevant data to solve the marketing question. You use that data as a basis to build a model to predict future patterns. One of the strengths of data modeling is that it can analyze data from multiple sources and give independent judgments regarding what is relevant or not required that is for the model to decide. According to Gibilisco (2012) there are two kinds of models in data mining. One is predictive models, which use data with known results to develop a model that can be used to explicitly predict values. Predictive modeling is necessary to facilitate the desired outcomes. Predictive modeling is a process used in predictive analytics to create a statistical model of future behaviour (Rouse, 2014). Predictive analytics is the area of data mining concerned with forecasting probabilities and trends. A predictive model is made up of a number of predictors, which are variable factors that are likely to influence future behaviour or results. In marketing, for example, a customer's gender, age, and purchase history might predict the likelihood of a future sale. In predictive modeling, data is collected for the relevant predictors, a statistical model is formulated, predictions are made and the model is validated or revised as additional data becomes available. The model

may employ a simple linear equation or a complex neural network, mapped out by sophisticated software. Predictive modeling is used widely in information technology (IT). In spam filtering systems, for example, predictive modeling is sometimes used to identify the probability that a given message is spam. Other applications of predictive modeling include customer relationship management (CRM), capacity planning, change management, disaster recovery, security management, engineering, meteorology and city planning. Predictive models are born whenever data is used to train a predictive modeling technique. To put it formally, data + predictive modeling technique = model. The second kind of data mining model is descriptive models, which describe patterns in existing data. Gibilisco (2012) defines descriptive modeling as a mathematical process that describes real-world events and the relationships between factors responsible for them. The process is used by consumer-driven organizations to help them target their marketing and advertising efforts. In descriptive modeling, customer groups are clustered according to demographics, purchasing behavior, expressed interests and other descriptive factors. Statistics can identify where the customer groups share similarities and where they differ. The most active customers get special attention because they offer the greatest ROI (return on investment). All the models are abstract representations of reality and can be a guide to understand business and suggest actions.

Data Mining Functionality: According to Han and Kamber (2001) data mining functionalities include data characterization, data discrimination, association analysis, classification, clustering, outlier analysis, and data evolution analysis. Data characterization is a summarization of the general characteristics or features of a target class of data. Data discrimination is a comparison of the general features of target class objects with the general features of objects from one or a set of contrasting classes (Han and Kamber, 2001). Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Classification is the process of finding a set of models or functions that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. Clustering analyzes data objects without consulting a known class model. Outlier and data evolution analysis describe and model regularities or trends for objects whose behavior changes over time (Han and Kamber, 2001).

According to Hofer et al. (2008) data mining techniques can be performed against the database, data marts or the enterprise data warehouse. A data warehouse is an integrated and consistent store of subjected oriented data obtained from a variety of sources and formatted into a meaningful context to support decision making in an organization. Most data warehouses today follow three-layer architecture. The first layer consists of data distributed throughout the various operational systems. The

second layer is an enterprise data warehouse, which is a centralized, integrated data warehouse that is the control point and single source of all data made available to end users for decision support applications. The third layer is a series of data marts. A data mart is a data warehouse whose data are limited in scope for the decision making needs of a particular user group (Hofer et al, 2008). A data mart can be independent of an enterprise data warehouse (EDW), derived from the EDW, or logical subset of the EDW. The data layer in the enterprise data warehouse is called the reconciled data layer. The characteristics of this data layer (ideally) are the following: It is detailed, historical, normalized, comprehensive, and quality controlled. Reconciled data are obtained by filling the EDW or operational data store from various operational systems. Reconciling the data requires four steps: capturing the data from the source systems, scrubbing the data (to remove inconsistencies), transforming the data (to convert it to the format required in the data warehouse), and loading and indexing the data in the data in the data warehouse. Reconciled data are not normally accessed directly by end users (Hofer et al, 2008). There are many different types of analysis that can be done in order to retrieve information from big data. Each type of analysis will have a different impact or result. The type of data mining technique you should use really depends on the type of business problem that you are trying to solve. Different analyses will deliver different outcomes and thus provide different insights. One of the common

ways to recover valuable insights is via the process of data mining (Hofer et al, 2008).

Data mining is not specific to any industry (Meens, 2012). In recent years, several economic sectors have gained a competitive edge by mining their expanding databases for valuable, detailed transaction information applying these data mining techniques. Examples of such uses are as follows:

1. Retail: Through the use of store-branded credit cards and point-of-sale systems, retailers can keep detailed records of every shopping transaction. This enables them to better understand their various customer segments. Some retail applications are performing basket analysis, sales forecasting, database marketing and merchandise planning and allocation (Rygieslski, 2002).
2. Banking: Banks can utilize knowledge discovery for various applications such as card marketing, card holder pricing and profitability, fraud detection and predictive life-cycle management (Rygieslski, 2002).
3. Telecommunications: Telecommunication companies around the world face escalating competition which is forcing them to aggressively market special pricing programs aimed at retaining existing customers and attracting new ones. Knowledge discovery in telecommunications involves call detail

record analysis, customer loyalty, fault detection and fraud detection (Rygieslski, 2002).

4. Bioinformatics/ biological data analysis and electrical power engineering: for instance in bioinformatics which involve analyzing large biological data sets that require making sense of the data by inferring structure or generalizations from the data (Raza, 2012).

In the field of purchasing and sales department of Telecommunication Company it is however still barely used. This is odd as in these environment Enterprise Resources Planning (ERP) systems are used. In Meens (2012) Modern Enterprise resource planning (ERP) is a business management software, usually a suite of integrated applications that a company can use to collect, store, manage and interpret data from many business activities, including:-

1. Product planning, cost and development
2. Manufacturing or service delivery
3. Marketing and sales

Therefore, data mining techniques seems to be a good technology to support the data processing, information retrieval and knowledge generation process in order to support decision making.

1.2 Statement of the Problem

In most companies, purchasing and sales department face a number of problems when attempting to analyze their data. There is generally no lack of data. In fact, many businesses feel they are drowning in data. The problems are as follows:

1. The difficulty in gaining a precise view of target area in a collated voluminous business transaction data by decision makers of an organization and presenting the information gained in a real time.
2. Inability to translate and formulate business question correctly to data mining tasks in order to have a regular in-depth analytical review of the data within organization repositories for a better understanding of their business data.
3. Problem of addressing data quality; Even if you can find and analyze data quickly and put it in the proper context for the audience that will be consuming the information, the value of data for decision-making purposes will be jeopardized if the data is not accurate or consistent.

In view of the above, this work is geared designing a model to automate the process that enhance effective decision making in organization.

1.3 Aim and Objectives of the Study

The aim of this project is to design an enhanced system of data mining in telecommunication service provider to support decision making. The specific objectives are as follows:

1. To identify the performance indicators of system under study.
2. Evaluate the existing system using identified performance indicators.
3. To build a data warehouse and data marts which is populated with historical sales records for data analysis.
4. Design the improved system that captures the performance indicators.
5. Implementation.
6. Testing running the enhanced system and evaluate its performance in relation with the indicators.

1.4 Research Questions

In relation to this specific objectives of this study, the researcher poses questions to guide the study as follows:

1. What are the key performance indicators for evaluation of data mining model?

2. To what extent does the existing model not measure up with required standard of performance indicators?
3. To what extent have an enhanced system been designed based on established tasks?
4. To what extent have the designed enhanced system performed to measure up with required standard of performance indicators?
5. What policy recommendations can be made?

1.5 Scope of the Study

The focus of this study is to develop a model of data mining techniques for telecommounication service provders in performing data analytic tasks to support decision marking. The study area focus in purchasing and sales department of Emerging Telecommunication Services (EMTs) also known as etisalat. Training data sets which is used for the model exploratory data analysis and evaluation is eight years sales records that range from 2008 to 2014 from Purchasing and Sales department of EMTs. The data mining model used by Etisalat make use of oracle/SQL programming language. This study laid more emphasis on the usefulness and exploratory powerful skills of data mining techniques on business performance evaluation. The scheme work in generating and answering business questions was mainly based on interviews with the top management staff of EMTs.

In the study the existing system was analyzed and the necessary tasks needed in order to develop the enhanced systems were immensely studied and applied.

1.6 Significance of the Study

The organization such as Emerging Marketing Telecommunication Services (etisalat), MTN Nigeria, Airtel Nigeria and GLO Nigeria will benefit from the output of the designed model. The output will be useful as follows:

1. The organization will be able to have access to its data warehouse resource and mine out refined information in a real time for effective decision making.
2. Easy evaluation of performance management: Managers can easily gain the multiple perspectives of how the business is doing; ask questions such as what year the organization and regions made the highest or lowest sales in general or specific product with regards to historical records and get answer in real time that can be used to make short term and longer term business plans.
3. There will be significant improvement in measuring and reporting performance that can help take company to the next level of growth. Inventory level and turnover can be check.
4. The output from the designed model can help organization in forecasting and prediction of future outcome.

Finally, output from the designed model is interpretable, this quality of it will greatly enhance and assist the analyst/consultant of an organization to easily formulate and answer business questions correctly.

CHAPTERTWO

2.0 LITERATURE REVIEW

2.1 Historical Development of Data Mining Techniques.

Early methods such as Bayes' theorem in the 1700s and regression analysis in the 1800s were some of the first techniques used to identify patterns in data (Kantardzic and Mehmed, 2003). After the 1900s, with the proliferation, ubiquity, and continuously developing power of computer technology, data collection and data storage were remarkably enlarged. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, clustering, genetic algorithms decision trees in the 1960s and support vector machines in the 1980s. Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns (Kantardzic and Mehmed, 2003).

Data mining techniques (DMT) have formed a branch of applied artificial intelligence (AI) since the 1960s (Liao et al, 2012). According to Ha et al. (2000) during the intervening decades, important innovations in computer systems have led to the introduction of new technologies. Data mining allows a search, for valuable information, in large volumes of data (Weiss and Indurkhya, 1998).

The explosive growth in databases has created a need to develop technologies that use information and knowledge intelligently. Therefore, DMT has become an increasingly important study area. Data mining or data mining technology has been used for many years by many fields such as businesses, scientists and governments. It is used to sift through volumes of data such as airline passenger trip information, population data and marketing data to generate market research reports, although that reporting is sometimes not considered to be data mining (Kantardz and Mehmed, 2003). According to Coenen (2004) the data that data mining techniques were originally directed at was tabular data and, given the processing power available at the time, computational efficiency and particular the number databases accesses was of significant concern. As the amount of processing power generally available increased, processing time although still an issue became less of a concern and was replaced with a desire for accuracy and a desire to mine ever larger data collections. Today, in the context of tabular data, we have a well-established range of data mining techniques available. It is well within the capabilities of many commercial enterprises and researchers to mine tabular data, using software such as SPSS clementine or Weka, on standard desktop machines (Coenen, 2004). However, the amount of electronic data collected by all kinds of institutions and commercial enterprises, year on year, continues to grow and thus there is still a need for effective mechanisms to mine ever larger data sets.

According to Wesis, (2009) telecommunications industry was one of the first to adopt data mining technology. This is most likely because telecommunication companies routinely generate and store enormous amounts of high-quality data, have a very large customer base, and operate in a rapidly changing and highly competitive environment. Telecommunication companies utilize data mining to improve their marketing efforts, identify fraud, and better manage their telecommunication networks. However, these companies also face a number of data mining challenges due to the enormous size of their data sets, the sequential and temporal aspects of their data, and the need to predict very rare events such as customer fraud and network failures and decision for effective allocation of resources in real-time. The popularity of data mining in the telecommunications industry can be viewed as an extension of the use of expert systems in the telecommunications industry according to Liebowitz (1988). These systems were developed to address the complexity associated with maintaining a huge network infrastructure and the need to maximize network reliability while minimizing labor costs. The problem with these expert systems is that they are expensive to develop because it is both difficult and time consuming to elicit the requisite domain knowledge from experts. Data mining can be viewed as a means of automatically generating some of this knowledge directly from the data (Wesis, 2009).

2.2 Empirical Framework of the Study

Data mining technique is finding increasing applications in expertise orientation and the development of applications for data mining technique is a problem-oriented domain (Liao et al, 2012). This present work design a model for data mining techniques for carrying out analytic task in order to improve an existing system and identify the several performance indicators that a data mining model should be evaluated with to ascertain its efficiency. The performance indicator are also the major characteristics if an existing system do not possess can make system is a weak one. Several scholars have explored the application of data mining techniques for series of research analysis and to provide solutions to myriad of problems confronting human race (Alhassan and Lawal, 2015). Model evaluation performance of some scholar was based on a specific criteria but the present work established not only one criteria but identify the key performance indicators of a model of data mining techniques.

Oyeniyi and Adeyemo (2015) developed a data mining model for customer churn analysis in banking sectors using data mining techniques. In their study, a real-life customers records provided by major Nigeria banks was used as a training data set. The raw data was extracted from the bank's customer relationship management database and transactional data warehouse which contained more than 1,048,576 customer records described with over 11 attributes. The raw data was cleaned, pre-

processed and then analyzed using Waikato Environment for Knowledge Analysis (WEKA), a data mining software tool for knowledge analysis. Simple Knee Means (K-Means) was used for the clustering phase while a rule-based algorithm, RIPPER (JRip) Error reduction was used for the rule generation phase. Oyeniyi and Adeyemo (2015) in their study adopt cross-industry process for data mining (CRISP-DM) methodology in the model development. According to NG DATA (2014) churn has an equal or greater impact on Customer Lifetime Value (CLTV) when compared to one of the most regarded Key Performance Indicator (KPI's) such as Average Revenue Per User (ARPU). As one of the biggest destructors of enterprise value, it has become one of the top issues for the banking industry. Customers churn prediction is aimed at determining customers who are at risk of leaving, and whether such customers are worth retaining. The constructed model was validated using the test validation dataset. Performance evaluation of the applied data mining techniques was carried out to test the goodness of fit, and adequacy of the constructed models in customer churn and non-churn prediction and analysis. The outcome of the model validation and performance evaluation was the ability of the applied model to accurately predict churn and non- churn customers (Oyeniyi and Adeyemo, 2015). The present work tend to identify others performance indicators apart from accuracy which is also one of the attributes under data quality that data mining model should possess.

Also, Fashoto et al., (2013) carried a project work on application of data mining technique for fraud detection in health insurance scheme in Nigeria using knee-point k-means algorithm. Nigerian Health Insurance Scheme (NHIS) was introduced by the Nigerian government to make healthcare affordable to all citizens, irrespective of economic situation or occupation. This scheme is, however, known to be beset by fraudulent claims from health practitioners within the system. To reduce or possibly eliminate this fraud, Fashoto et al., (2013) applied Knee-point K-means Clustering method, which is capable of detecting fraudulent claims from Health providers. Cluster-based outliers were examined. Health providers claims submitted to Health Maintenance Organization (HMO) were grouped into clusters. Claims with similar characteristics were grouped together. Clusters with small populations were flagged for further investigations. Clustering differs from classification and regression by not producing a single output variable, which leads to easy conclusions, but instead requires that you observe the output and attempt to draw your own conclusions (Kantardzic, 2011).

Fashoto et al. (2013) concluded that the model produced six clusters, but it was up to them to interpret the data within the clusters and draw conclusions from this information. In this respect, it was difficult for them to get clustering model correctly without determining the value of k clusters first. They were able to carve out some interesting information from the results on health insurance claims. The

results from the data collected from an HMO in Lagos Nigeria show that the total number of claims identified as possible anomalies from cluster-based outliers is 7 in Nigeria health insurance using probability of 0.6 as the cutoff point. The model this present work is to developed aims to produce result that is easy to interprets and make use of a visualization tool that provide high levels of understanding and trust.

Cortes and Pregibon (2001) developed signature-based methods and which was applied to data streams of call detail records for fraud detection. Telecommunication companies maintain data about the phone calls that traverse their networks in the form of call detail records, which contain descriptive information for each phone call. Call detail data is useful for marketing and fraud detection applications. A signature of a user corresponds to a vector of feature variables whose values are determine during a certain period of time. During Cortes and Pregibon (2001) work on data stream, they generated a signature from a data stream of call detail records to concisely describe the calling behavior of customers and then they used anomaly detection to "measure the unusualness of a new call relative to a particular account." Because new behavior does not necessarily imply fraud, this basic approach was augmented by comparing the new calling behavior to profiles of generic fraud and fraud is only signaled if the behavior matches one of these profiles. Customer level data also aid in identifying

fraud. Customer level such as price plan and credit rating information can be incorporated into the fraud analysis Rosset et al. (1999). More recent works using signatures has employed dynamic clustering as well as deviation detection to detect fraud, Alves et al. (2006) narrated in its article. In Cortes and Pregibon (2001) work each signature was placed within a cluster and a change in cluster membership was viewed as a potential indicator of fraud. Cortes and Pregibon (2001) exploited this behavior by recognizing that certain phone numbers are repeatedly called from compromised accounts and that calls to these numbers are a strong indicator that the current account may be compromised. Another method for detecting fraud exploits human pattern recognition skills. Cox et al. (1997) built a suite of tools for visualizing data that was tailored to show calling activity in such a way that unusual patterns are easily detected by users. These tools were then used to identify international calling fraud. Cortes and Pregibon (2001) signature-based model was experiment with France Telecom, AT&T, and SBC databases of 29, 26, and 25 Terabytes, respectively and was successful but the model have few drawbacks. Firstly, the signature-based method cannot support fraud incidences that did not follow the profiles. Secondly, these systems require upgrading to keep them up to date with current frauds methods. Upgrade and maintenance costs are high and mean continual dependence on system vendors. Thirdly, they require very accurate definitions of thresholds and parameters. Also, it require extensive

training using labeled data sets for formulation of evaluative models against which to assess newly arriving transactional instances. Adopted learning algorithms must therefore be continually retrained with labeled fraud data to support the extraction of emerging fraud threats resulting in a highly time consuming and costly business operation during which new fraud instances may go undetected. The present work identify this constraints in the course of study and tends to develop a model which eliminate all such kind of constraints.

Also, Avcilar and Yakut (2014) proposed the of use association rules in data Mining on a Clothing and Accessory Specialty Store. In this study, association rules were estimated by using market basket analysis and taking support, confidence and lift measures into consideration. In the process of analysis, by using of data belonging to the year of 2012 from a clothing and accessory specialty store operating in the province of Osmaniye, a set of data related to 42,390 sales transactions including 9,000 different product kinds in 35 different product categories (SKU) were used. Analyses were carried out with the help of SPSS Clementine packet program and hence 25,470 rules were determined. In Avcilar and Yakut (2014) study the useful association rules were determined between the product groups with regard to understanding what kind of purchase behavior customers' exhibit within a certain shopping visit from both in-category and different product categories for the specialty store in question. Association rule

mining (ARM) has attracted the attention and interest of a great number of researchers (Agrawal et al., 1993; Agrawal and Srikant, 1994; Srikant and Agrawal, 1995; Han and Fu, 1995; Liu et al., 1998; Changchien and Lu, 2001; Tsai and Chen, 2004; Chen et al., 2005; Zhou and Yau, 2007; Tang et al., 2008; Romero et al., 2011; Ahn, 2012). ARM analysis is to use to discover patterns that describe strongly associated features in the data.

According to Avcilar and Yakut (2014) work utilizing the association rules which they discovered as a result of the analyses, the retail store manager was able to develop and apply effective marketing and sales promotion strategies. support quite important strategic decisions of the retailing businesses which effect their success such as market segmentation, cross-sales, determination of the product mix to be offered for sale in the store, effective product assortment management, determination of the product prices and management of product price discounts, planning and management of product promotions, stock management, visual presentation of the products in the store, allocation the products on the shelves in the store. In Avcilar and Yakut (2014) work time based consecutive and patterned association rules were not determined. The present work did not only emphasis on pattern discovering and time based consecutive but also identify other performance indicators.

Furthermore, Idowu et al. (2013) worked on the use of data mining techniques for predicting immunize-able disease using Nigeria as a case study. The study develop a Mathematical Model (MM) for predicting immunize-able diseases that affect children between ages 0 - 5 years. The methodology used to build the predictive model is the CRISP-DM. A detail of this methodology is available in Sundar et al. (2012) and Viaene et al. (2014). The model was adapted and deployed for use in six selected localized areas within Osun State in Nigeria. Using the MATLAB's artificial neural network (ANN) toolbox, the Statistics toolbox for classification and regression, and the Naïve Bayesian classifier the MM was developed. The MM is robust in that it takes advantage of three data mining algorithm: ANN, Decision Tree Algorithm and Naïve Bayes Classifier. These data mining techniques provided the means by which hidden information were discovered for detecting trends within databases, and thus facilitate the prediction of future disease occurrence in the tested locations. Results obtained showed that diseases have peak periods depending on their epidemicity, hence the need to adequately administer immunization to the right places at the right time. The above reviewed work developed a model which possess a great quality “Robust” that most data mining technique model lacks. The present work embrace such quality and many more.

In Telecommunication Monitoring and maintaining telecommunication networks is an important task. As these networks become increasingly complex, expert systems were developed to handle the alarms generated by the network elements according to Weiss et al. (1998). However, because these systems are expensive to develop and keep current, data mining applications have been developed to identify and predict network faults. Fault identification can be quite difficult because a single fault may result in a cascade of alarms, many of which are not associated with the root cause of the problem. Thus an important part of fault identification is alarm correlation, which enables multiple alarms to be recognized as being related to a single fault. Klemettinen et al. (1999) came up with Telecommunication Alarm Sequence Analyzer (TASA).TASA is a data mining tool that aids with fault identification by looking for frequently occurring temporal patterns of alarms. Patterns detected by this tool were then used to help construct a rule-based alarm correlation system. Another effort, used to predict telecommunication Switch failures, employed a genetic algorithm to mine historical alarm logs looking for predictive sequential and temporal patterns; this was also cited in (Weiss and Hirsh, 1998). One limitation with the approaches just described is that they ignore the structural information about the underlying network. Devitt et al. (2005) stated that the quality of the mined sequences can be improved if topological proximity constraints are considered in the data mining process or if substructures in the

telecommunication data can be identified and exploited to allow simpler, more useful, patterns to be learned according to Baritchi et al. (2000). Also, Sterritt et al. (2000) narrates that another approach is to use Bayesian Belief Networks to identify faults, since they can reason about causes and effects. The present work studied the structure of the existing system and the historical sales obtained before carrying out data cleansing, data transformation and data compression in order to produce a quality result which will not eliminate any useful facts about what is expected from the result. The present study goes further to address quality of the result produced.

Nevertheless, in stock market the application of data mining techniques model cannot be overemphasized. In a stock market, how to find right stocks and right timing to buy has been of great interest to investors. To achieve this objective, Muh-Cherng et al. (2006) present a stock trading method by combination of filter rule and the decision tree technique. They provide an effective structure in which alternative decisions and the implications of taking those decisions can be laid down and evaluated. They also help investors to form an accurate, balanced picture of the risks and rewards that can result from a particular choice. The filter rule, having been widely used by investors, is used to generate candidate trading points. These points are subsequently clustered and screened by the application of a decision tree algorithm. Muh-Cherng et al. (2006) work is distinct because it

incorporates the future information into the criteria for clustering the trading points. Taiwan and NASDAQ stock markets in Iran were used to justify the proposed method. Experimental results show that the proposed trading method outperforms both the filter rule and the previous methods used by investors as written by Muh-Cherng et al. (2006). The present work is distinct from the reviewed work in the sense that the model will evaluate the existing events in order to predict future event and also identify other performance indicator.

Also, listed companies' financial distress prediction is important to both listed companies and investors. Jie and Hui, (2008) presented a data mining method which combines attribute-oriented induction, information gain, and decision tree that are suitable for preprocessing financial data and constructing decision tree model for financial distress prediction. On the basis of financial ratios attributes and one class attribute, adopting entropy-based discretization method, a data mining model for listed companies' financial distress prediction was designed. The empirical experiment with 35 financial ratios and 135 pairs of listed companies as initial samples got satisfying result, which testifies to the feasibility and validity of the proposed data mining method for listed companies' financial distress prediction (Jie and Hui (2008). This present model to be developed, its result will be tested for validity using the performance indicator.

Also, accurately, forecasting stock prices was extensively studied by (Hajizadeh et al. (2010). Jar-Long and Shu-Hui, (2006) provided a proposal to use a two-layer bias decision tree with technical indicators to create a decision rule that makes buy or not buy recommendations in the stock market. A novel method designed for using two-layer bias decision tree to improve purchasing accuracy. Comparison with random purchases, the results indicate the system presented not only have excellent out-of sample forecasting performance, but also delivers a significant improvement in investment returns for all listed companies. Additionally, the proposed system has few parameter requirements, stable learning, and fast learning speed. Increasingly, the system presented has high accuracy given large amounts of varied test data, with testing periods that experienced structural change including both bull and bear markets. Based on all of the above, they believe the proposed bias decision model is very flexible, modular and easily understandable. The main purpose of the study is to present a model with the attributes as narrated in (Jar-Long and Shu-Hui, 2006) and with more features.

Ehsan and Jamal (2010) reviewed the works of Lu and Chen (2006) that employed decision tree-based mining techniques to explore the classification rules of information transparency levels of the listed firms in Taiwan's stock market. The main purpose of their study was to explore the hidden knowledge of information disclosure status among the listed companies in Taiwan's stock market. Moreover,

the multi-learner model constructed with decision tree algorithm was applied. The numerical results show that the classification accuracy was improved by using multi-leaner model in terms of less Type I and Type II errors. In particular, the extracted rules from the data mining approach were developed as a computer model for the prediction or classification of good/poor information disclosure potential. By using the decision tree-based rule mining approach, the significant factors with the corresponding equality/inequality and threshold values were decided simultaneously, so as to generate the decision rules. The decision tree approach is able to provide the explicit classification rules. According to Lu and Chen (2006) constructed a multi-learner model boosting ensemble approach with decision tree algorithm was used to enhance the accuracy rate in their work. Based on the extracted rules, a prediction model was built to discriminate good information disclosure data from the poor information disclosure data with great precision. Moreover, the results of the experiment have shown that the classification model obtained by the multi-learner method has higher accuracy than those by a single decision tree model. Also, the multi-learner model has less Type I and Type II errors. It indicates that the multi-learner model is appropriate to elicit and represent experts' decision rules, and thus it has provided effective decision supports for judging the information disclosure problems in Taiwan's stock market. By using the rule based decision models, investors and the public can

accurately evaluate the corporate governance status in time to earn more profits from their investment. It has a great meaning to the investors, because only prompt information can help investors in correct investment decisions. Although a model to be developed is not a multi- layer type but a adopt classification data mining technique and decision tree algorithm. Classification as well known in producing precise and accurate result unlike clustering analysis as narrated in Fashoto et al. (2013), work on fraud detection in NHIS.

Basaltoa et al. (2005) applied a pair wise clustering approach to the analysis of the Dow Jones index companies in Iran, in order to identify similar temporal behavior of the traded stock prices. The objective of this attention was to understand the underlying dynamics which rules the companies' stock prices. In particular, it would be useful to find, inside a given stock market index, groups of companies sharing a similar temporal behavior. To this purpose, a clustering approach to the problem may represent a good strategy. To this end, the chaotic map clustering algorithm was used, where a map were associated to each company and the correlation coefficients of the financial time series to the coupling strengths between maps. The simulation of a chaotic map dynamics gives rise to a natural partition of the data, as companies belonging to the same industrial branch are often grouped together. The identification of clusters of companies of a given stock market index were exploited in the portfolio optimization strategies. Graph

representation of the stock market data and interpretation of the properties of this graph gives a new insight into the internal structure of the stock market (Basaltoa et al, 2005). The present work explores more powerful tool of visualization to be used to display results other than graphs. Furthermore, Balasubramanian and Selvarani (2014) worked on churn prediction in mobile telecommunication system were reviewed. The study used predictive mining or predictive modeling in carrying out their churn analysis. Given a predefined forecast horizon, the goal is to predict the future churners over that horizon, given the data associated with each subscriber in the network. The input for this problem includes the data on past calls for each mobile subscriber, together with all personal and business information that is maintained by the service provider. The model is trained with highest accuracy, the model must be able to predict the list of churners from the real dataset. According to Balasubramanian and Selvarani (2014), the knowledge discovery in database (KDD) function for problem was defined to be the classification problem. The predictive model shows 98.88% accuracy and error rate 1.116667% for our decision tree model and neural network model. Also it has the false positive of 0.93% and false negative of 2.23%. This study limits itself with prediction of churn and no steps were analyzed to include retention policies. The first step in predictive modeling is the acquisition and preparation of data. Having the correct data is as important as having the correct method Balasubramanian and Selvarani (2014).

The present work adopt this step in the course of development in order to come up the right data that will produce accurate results.

According to Olle and Cai (2013) on hybrid churn prediction model in mobile telecommunication, an efficient churn prediction model that raises the trilogy: Who wants churn; why does he or she want to churn and when would that happen, is very less found in the research of churn prediction. Accordingly, their goal of the study is to show that hybrid models built on DM techniques can explain the churn behavior with more accuracy than single methods; and that in some extend the reason of churn can be revealed, as well as explaining the gap between the decision to churn and the deactivation time. Hybrid model uses Logistic Regression in parallel with Voted Perception for classification, and combined with clustering. The model is built using WEKA, a well-known tool of Machine Learning. 2000-instances of a real world dataset with 23 variables from an Asian mobile operator are used for evaluation. In particular, the moment of deactivation is subsequent to an overseen grace period, favorable to the measures adopted to proactively and efficiently interfere. Hybrid methods in Predictive Data Mining techniques follow two strategies of serial combination. In the first strategy, the input from the preprocessing step passes through a classification technique and subsequently to a clustering phase to get the final output. The present work follows a step of data

preparation down to design of the model which incorporate several classes of data mining techniques.

Qasem and Eman, (2012) studied the use of data mining techniques to build a classification model for predicting employees' performance. In their work data mining techniques were utilized to build a classification model to predict the performance of employees. To build the classification model the CRISP-DM data mining methodology was adopted. Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. To validate the generated model, several experiments were conducted using real data collected from several companies. The model was intended to be used for predicting new applicants' performance. The present work did not only develop a model but it was also evaluate and assess using certain key performance indicators.

KNOWLEDGE GAP

Several literature reviewed so far have demonstrated the great work of data mining in problem solving but none of the literature reviewed so far have identified the overall performance indicators that a data mining model should possess and also failed to demonstrate its benefit on prediction of unkown entitities.

2.3 Concepts, Theories, Model approach and Technology.

The theory of data mining is statistics (as a science) according to Mannila (1997). Research in data mining and knowledge discovery in databases has mostly concentrated on developing good algorithms for various data mining tasks. Some parts of the research effort have gone to investigating data mining process, user interface issues, database topics, or visualization as written by Fayyad et al. (1996). Relatively little has been published about the theoretical foundations of data mining. The area is at its infancy (Mannila, 2000).

First of all one has to answer questions such as "Why look for a theory of data mining? Data mining is an applied area, why should we care about having a theory for it?" Probably the simplest answer is to recall the development of the area of relational databases. Given that theory is useful, what would be the properties that a theoretical framework should satisfy in order that it could be called a theory for data mining? The example of relational model can serve us also here. First of all, the theoretical framework should be simple and easy to apply; it should (at least some day) gives us useful results that we could apply to the development of data mining algorithms and methods (Mannila, 2000). A theoretical framework should also be able to model typical data mining tasks (clustering, rule discovery, classification), be able to discuss the probabilistic nature of the discovered patterns and models, be able to talk about data and inductive generalizations of the data,

and accept the presence of different forms of data (relational data, sequences, text, web). Also, the framework should recognize that data mining is an interactive and iterative process, where comprehensibility of the discovered knowledge is important and where the user has to be in the loop, and that there is not a single criterion for what an interesting discovery is *(*Mannila, 2000).

Several theories for the basis of data mining include the following according to (Mannila, 2000):

Data Reduction/Reductionist Approach: In this theory, the basis of data mining is to reduce the data representation, data reduction trades accuracy for speed in response to the need to obtain quick approximate answers to queries on very large database. Data reduction includes singular value decomposition that is the driving element behind principal components analysis, wavelets, regression, log-linear models, histograms, clustering sampling and the construction of index tree. Reductionist approaches sees data mining as a part of some existing area, such as statistics or Machine learning. Many data mining tasks naturally may be formulated in statistical terms, and many statistical contributions may be used in data mining in a quite straightforward manner (Mannila, 2000). The reductionist approach of viewing data mining in terms of statistics has advantages of the strong theoretical background and easy-formulated problems; this very beneficial to the present study. Accuracy is the most common evaluation parameter, but we don't

have to look out the other parameters. To select the most appropriate method, other aspects apart from accuracy should be considered as for example: robustness, speed, interpretability and ease of use (Clemente et al., 2001). According Clemente et al. (2001) *even* if a model is accurate, it cannot be said to be of good quality if it spends too much time for processing data or if it is not easily interpretable. Thus, quality is a multi-faceted concept. The most important characteristics depend on user perspectives, needs and priorities, which may vary across user groups. Cai and Zhu (2015) proposed a hierarchical data quality standard from the user which is used for quality criteria assessment. Also, Piprant and Ernst (2008) provided a list to demonstrate a need to address data quality assessment throughout the solution's system development life cycle (SDLC) called generic SDLC and quality assessment. Nevertheless, Statistical summaries of all sorts are also common and useful for gathering insights for assessing model trust. Pairwise scatter-plots and low-dimensional density estimates are especially common. Summaries can be particularly useful for comparing relative trust of two models, by allowing analysis to focus on subsets of features for which their interrelationships differ most significantly between two models (Thearling et al., 2011).

Data Compression Approach: The data compression approach to data mining is simply to state: the goal of data mining is to compress the data set by finding some structure for it. That is, data mining looks for knowledge, where knowledge is

interpreted as a representation that makes it possible to code the data using few bits. For example, the minimum description length (MDL) principle (Mehta et al., 1995) can be used to select among different encodings accounting to both the complexity of a model and its predictive accuracy. If desired, the minimum description length (MDL) principle can be used to select among different encodings. To yield structure that is comprehensible to the user, we have to specify compression methods that are based on concepts that are easy to understand (Mannila, 2000). Several simple data mining techniques can be viewed as instances of this approach. For example, association rules can be viewed as ways of providing compression of parts of the data (Agrawal et al., 1996). Also, Clustering approaches can also be considered as a way of compressing the dataset (Imielinski and Mannila, 1996).

This theory is beneficiary to present study because it have a strong analytic background. Understanding as a performance indicator is undoubtedly the most fundamental motivation behind visualizing the model (Thearling et al., 2011). The more interesting way to use a data mining model is to get the user to actually understand what is going on so that they can take action directly. Unless the output of the data mining system can be understood qualitatively, it won't be of any use. In addition, the model needs to be understood so that the actions that are taken as a result can be justified to others. Understanding means more than just

comprehension; it also involves context (Thearling et al., 2011). If the user can understand what has been discovered in the context of their business issues, they will trust it and put it into use. Three components are essential for understanding a model: representation, interaction, and integration. Representation refers to the visual form in which the model appears. A good representation displays the model in terms of visual components that are already familiar to the user. Interaction refers to the ability to see the model in action in real time, to let the user play with the model as if it were a machine. Integration refers to the ability to display relationships between the model and alternate views of the data on which it is based. Integration provides the user context (Thearling et al., 2011).

Constructive induction approach: Constructive induction is a learning process that consists of two intertwined phases, one of which is responsible for the construction of the "best" representation space and the second concerns with generating hypothesis in the found space (Michalski and Wnek, 1993). Constructive induction methods are classified into three categories: data driven (information from the training examples is used), hypothesis-driven (information from the analysis of the form of intermediate hypothesis is used) and knowledge-driven (domain knowledge provided by experts is used) methods. Any kind of induction strategy (implying induction, abduction, analogies and other forms of non-truth preserving and non-monotonic inferences) can be potentially used.

However, the focus usually is on operating higher-level data-concepts and theoretical terms rather than pure data. Michalski (1997) considered constructive (expands the representation space by attribute generation) and destructive (contract the representational space by feature selection or feature abstraction) operators that can be applied to produce a better representation space comparing to the original one. Constructive induction approaches have relatively strong analytical background, as well as connections to the philosophy of science. This attribute make the theory very useful in the present study. Analysis yield good data quality. According to Piprant and Ernst (2008) the true cause of poor data quality can be attributed to a lack of supporting business processes and insufficient analysis techniques.

Probabilistic Approach: A possible theoretical approach to data mining is to view data mining as the task of finding the underlying joint distribution of the variables in the data. Typically one aims at finding a short and understandable representation of the joint distribution, e.g., Bayesian network (Heckerma, 1997) or a hierarchical Bayesian model (Gelman et al., 1995 and Gilks et al., 1996). This approach is obviously closely related to the reductionist approach of viewing data mining as statistics. The advantages of the approach are that the background is very solid, and it is easy to pose formal questions. Tasks such as clustering or classification fit easily into this approach. What seems to be lacking, as in most of the approaches,

are ways for taking the iterative and interactive nature of the data mining process into account. Hierarchical Bayesian models (Gelman et al 1995 and Gilks et al 1996) seem a very promising statistically sound approach to data mining. Such a model describes the structural part of the distribution independently of the actual functional form of the distribution. It is ability to generate a short and understandable representation of the joint distribution is an attribute of this theory that make it very useful to the present work. Data mining approaches emphasize database integration, simplicity of use, and understandability of results. Last but not least Mannila (2000) pointed out that the theoretical framework of statistics does not concern much about data analysis as a process that generally includes data understanding, data preparation, data exploration, results evaluation, and visualization steps. However, there are persons (mainly with strong statistical background) who equate DM to applied statistics, because many tasks of DM may be perfectly represented in terms of statistics (Mannila, 2000).

MICROECONOMIC VIEW OF DATA MINING: The microeconomic view of data mining introduced by Kleinberg et al. (1998) is a very interesting approach. The starting point is that data mining is about finding actionable patterns: the only interest is in patterns that can somehow be used to increase utility. Kleinberg et al., (1998) give a decision theoretic formulation of this principle: the goal of the organization is to find the decision x that leads to the maximum utility $f(x)$. The

form of the utility *f(x)* is typically a sum of utilities *fi(x)* for each customer i. This function *fi(x)* is actually a complex function of the decision *x* and the data *yi* on customer *i*, and can often be represented using a single function, i.e., as *fi(x) = g(x, yi)*. Thus the task is to find the decision *x* maximizing the sum of the terms *g(x, yi)* over the customers *i*. The basic observation of Kleinberg et al. (1998) is that data mining is useful if and only if the function g is nonlinear. They are able to describe pattern discovery, clustering, etc. as instantiations of the framework, and demonstrate also interesting connections to sensitivity analysis. Furthermore, they also show how the framework gives useful suggestions for research problems. The approach is clearly very promising to the present study because seem to satisfy most of the requirements.

INDUCTIVE DATABASES: The basic idea of inductive databases is that the query concept should be applied also to data mining and knowledge discovery tasks. In the slogan form from (Imielinski 1995, Imielinski and Mannila, 1996): there is no such thing as discovery, it is all in the power of the query language. That is, one can benefit from viewing the typical data mining tasks not as dynamic operations constructing new nuggets of information, but as operations unveiling hitherto unseen but pre-existing pieces of knowledge. The term inductive database refers to a normal database plus the set of all sentences from a specified class of

sentences that are true of the data (Boulicaut et al., 1998, Boulicaut et al., 1999 and Mannila, 1999).

In model-theoretic terms according to Chang and Keisler, (1990) the inductive database contains the data and the theory of the data. The approach can be compared to the idea of deductive databases, which contain a normal database plus a set of rules for deriving new facts from the facts already existing in the database. The user of a deductive database can act as if all the facts derivable from the database would be actually stored there. Of course, this set might be infinite, or finite but very large, so in practice it cannot be represented. But the idea of treating stored and derived facts in the same way is crucial for deductive databases. In the same way, an inductive database does not contain all the rules that are true about the data stored in it; the user is just able to assume that all these rules are there. In practice, the rules are constructed on demand. The schema of an inductive database consists of a normal relational database schema plus a schema for the generalizations. It is relatively easy to design a query language that works on such schemas (Boulicaut, 1998). The result of a query on an inductive database is again an inductive database, so we have the closure property that has been so useful for relational databases. The process view on data mining is directly built in to the concept of inductive databases. It also suggests architecture for data mining systems. Association rules and other simple pattern formalism fit quite easily into

the framework, and there are some good partial solutions that can be viewed as partial implementations of inductive databases. However, e.g., clustering is harder to describe in a useful way. The probabilistic nature of data mining can be incorporated by having the underlying concept class support probabilistic concepts. It is possible to design a query language that works on inductive databases (Boulicaut et al., 1998). Certainly, there might be a need to find a solution about what should be presented to a user and when to stop the recursive rule generation while querying. This theory is very vital to the present study because it is database oriented approach and the present work is also related to database.

In general the essence of theory is to guide and help in producing quality model which can pass a test of fitness. In model assessment and evaluation, data quality is consider as a performance indicators. According to Oyeniyi and Adeyemo, (2015) the performance of machine learning algorithms is typically evaluated using predictive accuracy. Clemente et al., (2001) identified accuracy as an indicator in data quality. In Clemente et al. (2001) work other parameters for model evaluation and assessment are also identified and they are as follows: Robustness, speed, interpretability, and ease of use. Cai and Zhu, (2015) also identified several parameters such as consistency, integrity, understandable, trust and presentation quality which are classified into dimension called data quality dimension. Clemente et al., (2015) proposed the use of a methodology for evaluating statistical

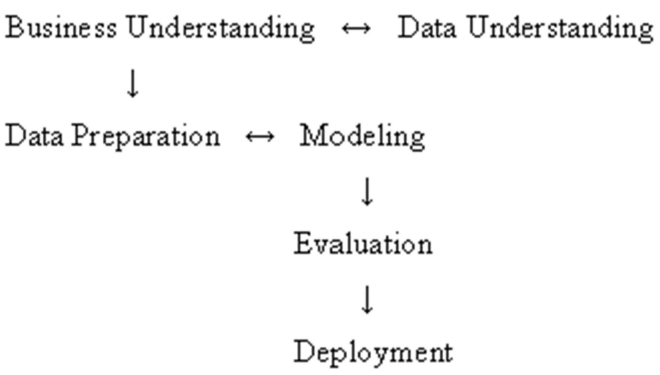
models for classification with the use of a composite indicator. The composite indicator measures multidimensional concepts which cannot be captured by a single parameter and help decision makers to solve complex problem. The parameters accuracy, robustness, speed, interpretability, and ease of use are joined to form a composite indicator.

In general terms, an indicator is a quantitative or a qualitative measure derived from a series of observed parameters (Clemente et al., 2001).To distinguish between the quantitative and qualitative parameter; accuracy, speed and robustness are a quantitative measure, because they are easily quantifiable. However, the interpretability and the ease of use are difficult to measure, because it is a subjective measure. In the case of the interpretability, it is generally recognized that there are some methods more interpretable than others so we transform it in a quantitative parameter (Clemente et al., 2001).

Model approach for Data mining: According to DELL Inc. (2016) in the business environment, complex data mining projects may require the coordinate efforts of various experts, stakeholders, or departments throughout an entire organization. In the data mining literature, various general frameworks have been proposed to serve as blueprints for how to organize the process of gathering data, analyzing data, disseminating results, implementing results, and monitoring

improvements (DELL Inc, 2016). Model approach proposed by (DELL Inc, 2016) are as follows:

Cross-Industry Standard Process for data mining (CRISP): was proposed in the mid-1990s by a European consortium of companies to serve as a non-proprietary standard process model for data mining. This general approach postulates the following general sequence of steps for data mining projects (DELL Inc, *2016*):



Chapman et al., (2000) explain the above steps as follows:

Business understanding: This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

Data understanding: The data understanding phase starts with initial data collection and proceeds with activities that enable you to become familiar with the

data, identify data quality problems, discover first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information (Chapman et al., 2000).

Data preparation: The data preparation phase covers all activities needed to construct the final dataset. Data that will be fed into the modeling tool(s) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools (Chapman et al., 2000).

Modeling: In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, going back to the data preparation phase is often necessary (Chapman et al., 2000).

Evaluation: At this stage in the project, you have built a model (or models) that appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to thoroughly evaluate it and review the steps executed to create it, to be certain the model properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a

decision on the use of the data mining results should be reached (Chapman et al., 2000).

Deployment: creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the management can use it. It often involves applying "live" models within an organization's decision making processes for example, real-time personalization of Web pages or repeated scoring of marketing databases. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise (Chapman et al., 2000).

The present work adopt CRISP data mining approach because it is comprehensive, every little detail are considered. At the first stage of this methodology, the objectives are noted and the performance indicators are identify. Throughout the modelling process parameter needed to achieve valid results are considered.

SIX SIGMA: according to (DELL Inc, 2016) the Six Sigma methodology is a well-structured, data-driven methodology for eliminating defects, waste, or quality control problems of all kinds in manufacturing, service delivery, management, and other business activities. This model has recently become very popular due to its successful implementations in various American industries, and it appears to gain

favor worldwide. It postulated a sequence of, so-called, DMAIC steps (DELL Inc, 2016) -

Define → Measure → Analyze → Improve → Control

The traditional DMAIC Six Sigma process, as it is usually practiced, which is focused on evolutionary and continuous improvement manufacturing or service process development, usually occurs after initial system or product design and development have been largely completed. DMAIC Six Sigma as practiced is usually consumed with solving existing manufacturing or service process problems and removal of the defects and variation associated with defects. It is clear that manufacturing variations may impact product reliability. However, despite that DMAIC as a methodology has been successfully used as an end-to-end technical project frameworks for analytic and mining projects, this has been observed by domain experts to be somewhat similar to the lines of CRISP-DM.

SEMMA APPROACH: According to SAS Institute (2003), data mining is often seen as an unstructured collection of methods, or as one or two specific analytic tools, such as neural networks. However, data mining is not a single technique, but an iterative process in which many methods and techniques may be appropriate. And like data warehousing, data mining requires a systematic approach. Beginning with a statistically representative sample of the data, you can apply exploratory

statistical and visualization techniques, select and transform the most significant predictive variables, model the variables to predict outcomes, and affirm the model's accuracy. To clarify the data mining process SAS Institute (2003), has mapped out an overall plan for data mining. This step-by-step process is referred to by the acronym SEMMA: sample, explore, modify, model, and assess.

Step 1: Sample

Extract a portion of a large data set big enough to contain the significant information yet small enough to manipulate quickly. For optimal cost and performance, SAS Institute advocates a sampling strategy, which applies a reliable, statistically representative sample of the full detail data. Mining a representative sample instead of the whole volume drastically reduces the processing time required to get crucial business information. If general patterns appear in the data as a whole, these will be traceable in a representative sample. If a niche is so tiny that it's not represented in a sample and yet so important that it influences the big picture, it can be discovered using summary methods (SAS Institute, 2003).

Step 2: Explore

Search speculatively for unanticipated trends and anomalies so as to gain understanding and ideas. After sampling your data, the next step is to explore them visually or numerically for inherent trends or groupings. Exploration helps refine the discovery process. If visual exploration does not reveal clear trends, you can

explore the data through statistical techniques including factor analysis, correspondence analysis, and clustering. For example, in data mining for a direct mail campaign, clustering might reveal groups of customers with distinct ordering patterns. Knowing these patterns creates opportunities for personalized mailings or promotions (SAS Institute, 2003).

Step 3: Modify

Create, select, and transform the variables to focus the model construction process. Based on your discoveries in the exploration phase, you may need to manipulate your data to include information such as the grouping of customers and significant subgroups, or to introduce new variables. You may also need to look for outliers and reduce the number of variables, to narrow them down to the most significant ones. You may also need to modify data when the "mined" data change. Because data mining is a dynamic, iterative process, you can update data mining methods or models when new information is available (SAS Institute, 2003).

Step 4: Model

Search automatically for a variable combination that reliably predicts a desired outcome. Once you prepare your data, you are ready to construct models that explain patterns in the data. Modeling techniques in data mining include neural networks, tree-based models, logistic models, and other statistical models such as time series analysis and survival analysis. Each type of model has particular

strengths, and is appropriate within specific data mining situations depending on the data. For example, neural networks are good at combining information from predictors which support nonlinear associations with a target (SAS Institute, 2003).

Step 5: Assess

Evaluate the usefulness and reliability of findings from the data mining process.

The final step in data mining is to assess the model to estimate how well it performs. A common means of assessing a model is to apply it to a portion of data set aside during the sampling stage sometimes known as validation data. For a model to be considered successful and useful, it should work for this validation sample as well as for the training data used to construct the model (SAS Institute, 2003).

Similarly, you can test the model against known data. For example, if you know which customers in a file had high retention rates and your model predicts retention, you can check to see whether the model selects these customers accurately. In addition, practical applications of the model, such as partial mailings in a direct mail campaign, help prove its validity. (SAS Institute, 2003)

SEMMA approach have a great significant to the present work. The benefit of SEMMA approach is that at every stage and step taken in modelling process, the key performance indicator that will prove the model efficiency is considered. The parameters that will help to achieve it are also evaluated and put into consideration

before carrying out any process in order not to alter any necessary information that will lead to produce a valid model.

Data Mining Technology: According to Zentut (2011) there are several major data mining *techniques* have been developing and using in data mining projects recently including *association, classification, clustering, prediction, sequential patterns and decision tree.* The major classes of data mining techniques/statistical methods are as follows (Rijmenam, 2014):

1) Classification.
2) Clustering
3) Regression.
4) Association Rule Learning.
5) Sequence Discovering
6) Visualization.

Classification: Ahmed (2004); Berry and Linoff (2004) and Carrier and Povel (2003) defined classification as one of the most common learning models in data mining. For instance, it aims at building a model to predict future customer behaviors through classifying database records into a number of predefined classes based on certain criteria.

Clustering: Is a common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data (Jain and Dubes, 1988). The categories can be mutually exclusive and exhaustive or consist of a richer representation, such as hierarchical or overlapping categories. Examples of clustering applications in a knowledge discovery context include discovering homogeneous subpopulations for consumers in marketing databases and identifying subcategories of spectra from infrared sky measurements enumerated in Cheeseman and Stutz (1996) article.

Regression: is learning a function that maps a data item to a real-valued prediction variable. Regression applications are many, for example, predicting the amount of biomass present in a forest given remotely sensed microwave measurements, estimating the probability that a patient will survive given the results of a set of diagnostic tests, predicting consumer demand for a new product as a function of advertising expenditure, and predicting time series where the input variables can be time-lagged versions of the prediction variable (Rijmenam, 2014).

Association: Ahmed and Jiao (2004) states that association aims to establishing relationships between items which exist together in a given record. Market basket analysis and cross selling programs are typical examples for which association modeling is usually adopted. Common tools for association modeling are statistics and prior algorithms.

Sequence discovery: Sequence discovery is the identification of associations or patterns over time as defined by Bersonet et al. (2000). Its goal is to model the states of the process generating the sequence or to extract and report deviation and trends overtime (Mitra et al., 2002). Common tools for sequence discovery are statistics and set theory.

Visualization: Shaw et al. (2001) article defined visualization as the presentation of data so that users can view complex patterns. It is used in conjunction with other data mining models to provide a clearer understanding of the discovered patterns or relationships. Examples of visualization model are 3D graphs.

2.3.1 Data Mining Algorithm

According to SQL Server (2016) an *algorithm* in data mining (or machine learning) is a set of heuristics and calculations that creates a model from data. To create a model, the algorithm first analyzes the data you provide, looking for specific types of patterns or trends. The algorithm uses the results of this analysis over many iterations to find the optimal parameters for creating the mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics. According to SQL Server (2016) the mining model that an algorithm creates from your data can take various forms, including:

1. A set of clusters that describe how the cases in a dataset are related.

2. A decision tree that predicts an outcome, and describes how different criteria affect that outcome.
3. A mathematical model that forecasts sales.
4. A set of rules that describe how products are grouped together in a transaction, and the probabilities that products are purchased together.

Most wide used algorithm are predictive analytic algorithm. Predictive analytic is a branch of data mining concerned with the analysis of data to identify underlying trends, patterns, or relationships to predict future probabilities and trends (Nyce, 2007). In predictive modelling, data is collected, a statistical model is formulated, predictions are made and the model is validated or revised as additional data becomes available (Mishra et al., 2000). Predictive data mining automatically create classification model from training dataset, and apply such model to automatically predict other classes of unclassified datasets (Bharacheesh and Iyenger, 2004). Predictive data mining modeling involves algorithms and the predictive data mining (PDM) algorithms are as follows (Danjuma and Osofisan, 2015):

Decision Tree (DT) Algorithm: According to Aftarczuk (2007) decision tree is a predictive data mining techniques often used in several models to easily visualize, and understand resistant to noise in data. And is applicable in both regression and association data mining tasks capable of handling continuous attributes, which are

essential. In case of medical data e.g. blood pressure, temperature, etc. It is a non-parametric supervised learning method used for classification to create models that predicts the value of a target variable by learning simple decision rules inferred from the data features according to Venkatalakshmi and Shivsankar (2014). Decision trees are significantly faster than neural networks with a shorter learning curve that is mainly used in the classification and prediction to represent knowledge. The instances are classified by sorting them down the tree from the root node to some leaf node. Milovic (2012), the nodes are branched based on if-then condition. The variants of decision tree algorithm include CART, ID3, C4.5, SLIQ, and SPRINT as written by Osamanbegovic and Suljiic (2012).

Artificial neural networks (ANN's) Algorithm: The main function of Artificial Neural Networks is prediction (Aftarcruz, 2007). Despite its complexity and difficulty in understanding the predictions, it has been successfully applied in clinical several fields such as in medicals in prediction of coronary artery disease, EEG signals processing and the development of novel antidepressants. ANN is biologically inspired, highly sophisticated analytical techniques, capable of modelling extremely complex non-linear functions as written by Fathima et al. (2011) that are modeled based on the cognitive learning process and the neurological functions of the human brain, consisting of millions of neurons interconnected by synapses and capable of predicting new observations after

learning from existing data. The process of learning in ANN is to solve a task, having a set of observations and a class of functions, which is to find as the optimal solution to the task. The most popular of the ANN is Multilayer Perception algorithm. Osmanbegovic and Suljic (2012) MLP is most suitable for approximating a classification function, and consists of a set of sensory elements that make up the input layer, one or more hidden layers of processing elements, and the output layer of the processing elements. The Multi-Layer Perception (MLP) with back-propagation (a supervised learning algorithm) is arguably the most commonly used and well-studied ANN architecture capable of learning arbitrarily complex nonlinear functions to arbitrary accuracy levels as (Kumari and Godara, 2011) wrote.

Naïve Bayes Algorithm: Naïve Bayes is a Bayesian Network based on Bayes rules of simple conditional probability that estimate the likelihood of a property given the set of input data. According to Bhardwaj and Pal (2011) Naïve Bayes require small amount of training data to estimate parameters such as mean and variance necessary for classification. The structure of a Naïve Bayes model forms a Bayesian network of nodes with one node for each attribute. As written by Aftarczuk (2007) the nodes are interconnected with directed edges and form a directed acyclic graph. Bayesian networks uses directed acyclic graphs to model the dependencies among variables (Floyd, 2007). According to Bellazzi and Zupan

(2008) each node in the acyclic directed graph represents a stochastic variable and arcs represent a probabilistic dependency between a node and its parents. The Bayesian Network is a way of representing probabilistic relationships between variables associated with an outcome of interest. The outcome of interest could be uncertainty essential during clinical diagnosis, prediction of patient's prognosis and treatment selection. The probabilities applied in the Naïve Bayes algorithm are calculated according to the Bayes' Rule.

2.4 Conceptual Framework of the Study

In narrowing down the theoretical framework of the study, the benefit of each data mining technique in relation to the data mining theories, models and algorithm was considered. The conceptual framework of this study is presented in the Figure 2.1.

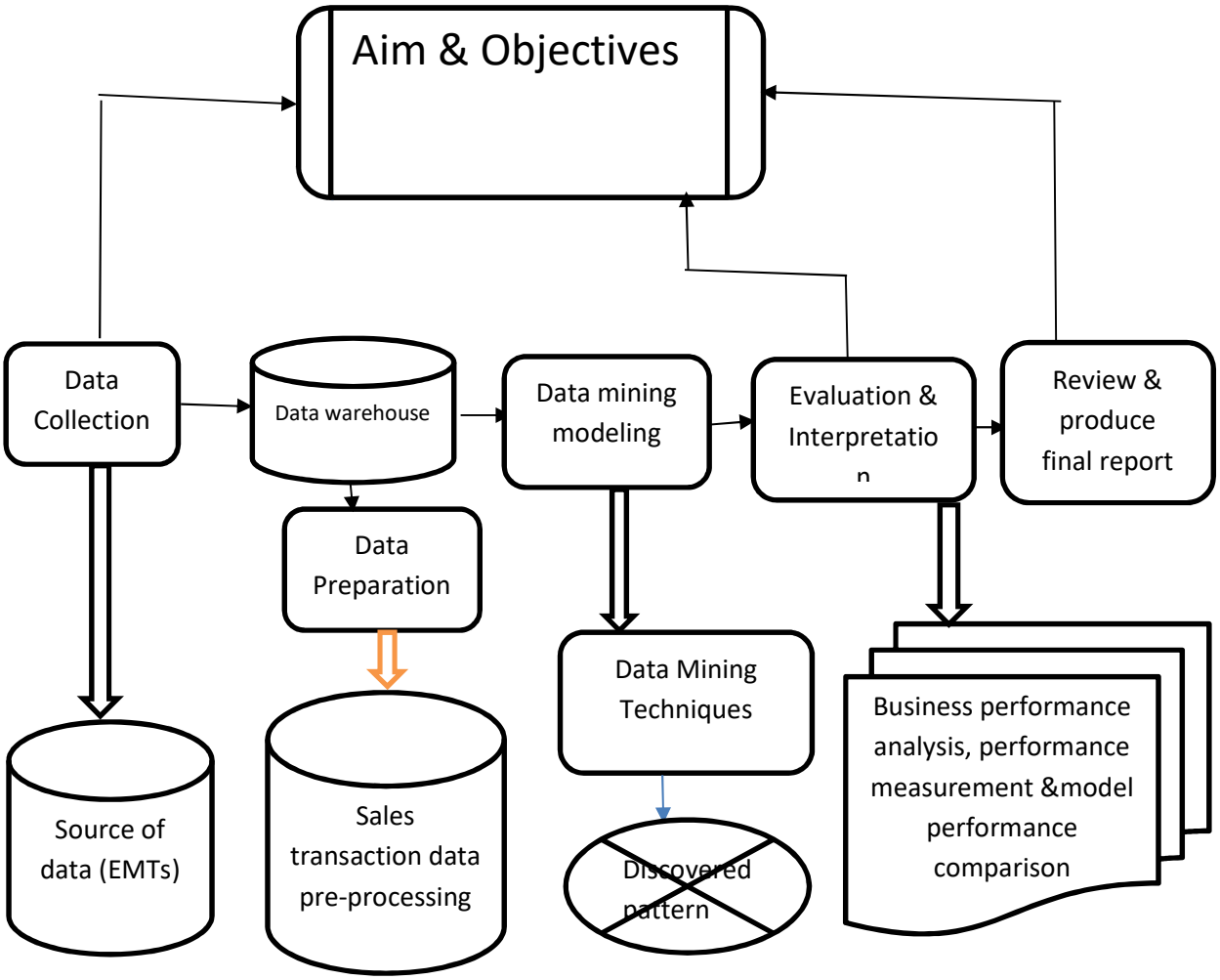


Figure 2.1: Conceptual Framework of the Study

As shown in Figure 2.1, the aim of the study is to design model for data mining techniques in performing analytic tasks and support decision making in telecommunication service providers. The concept involves identifying performance indicators that enhances existing system. On this basis the enhanced model for data mining is designed as presented in Figure 2.1. The performance

indicators are reviewed in section 2.3 of this chapter. The specific objectives are presented in boxes below the aim and objectives as shown in the Figure 2.1.

Data collection: The dataset used for this study for business performance analysis was acquired from Emerging Market Telecommunication Service (etisalat) Nigerian. The raw data was extracted from the purchasing and sales department database and transactional data warehouse which contained sales records ranging from 2008 to 2015 describe over 3 attributes.

Data warehouse: the study build a data warehouse and data marts for the data mining model. The warehouse is populated with the data collected from emerging market telecommunication Service

The data preparation phase: in this study covers all the activities needed to construct the final datasets from the raw sample data obtained from the etisalat. Data preparation tasks consider transforming acquired datasets to remove noise, inconsistencies, incoherence, bias and redundancies. The data preparation tasks includes table, record, and attribute selection as well as transformation and elimination of data for modelling, that can be performed multiple times, in no prescribed order.

Data Mining Modelling involves the application of data mining techniques such as classification, association rule learning, regression and sequence pattern

discovering in order to discover pattern and obtain information from the training dataset.

Evaluation and interpretation involves analysis, examination and assessment of the designed model to ascertain if it meets the requirement specification. In this stage the model is be evaluate using performance indicators to test its fitness.

Review and produce final result: this involves generating of result obtained through analysis and proper presentation using visualization tool such as dashboard as discussed in the section 1.1 to improve understanding and interpretation.

2.4.1 Telecommunication Service Provision Variables.

Data for data mining comes in many forms: from computer files typed in by human operators, business information in SQL or some other standard database format, information recorded automatically by equipment such as fault logging devices, to streams of binary data transmitted from satellites. For purposes of data mining we will assume that the data takes a particular standard form (Bramer, 2013).

Bramers (2013) stated that for any data mining application we have a universe of objects that are of interest. The universe of objects is normally very large and we have only a small part of it. Usually we want to extract information from the data available to us that we hope is applicable to the large volume of data that we have not yet seen. Each object is described by a number of *variables* that correspond to

its properties. In data mining variables are often called *attributes.* The set of variable values corresponding to each of the objects is called a *record* or (more commonly) an *instance*. The complete set of data available to us for an application is called a *dataset*. A dataset is often depicted as a table, with each row representing an instance. Each column contains the value of one of the variables (attributes) for each of the instances (Bramer, 2013).

In the present work, eight years historical sales records ranging from 2008 to 2015 of emerging market telecommunication services (EMTS) was used as a training dataset and this form the data that resides in its data warehouse. EMTS classifies its operating localities into regions; thereby forming seven sub-regions which are Lagos South (LS), Lagos North (LN), South West (SW), South South (SS), North 1(N1) and North 2 (N2). The major product been sold by the company according to the historical records are namely Airtime, Electronic top up (e-top up) and Sim cards and these are the product considered at the course of analysis. The product mentioned are the variables in the study. A sample of a dataset used in this study is presented in the Table 2.1:

Table 2.1: Sales Report on Airtime.

MTD DP Total Airtime Sales Summary as at OCTOBER 2008			
Total Airtime Sales-October 2008			
Region	**Total Airtime Rev.Target**	**Total Airtime Purchase**	**% Achieved**
LS	**1,483,854,412**	**800,000,000**	**53**
LN	**1,483,854,412**	**624,514,210**	**42**
SW	**1,409,124,954**	**324,814,112**	**23**
SE	**1,385,512,060**	**423,008,214**	**30**
SS	**1,484,302,326**	**618,218,412**	**41**
North 1	**1,631,289,908**	**900,814,218**	**55**
North 2	**1,309,522,321**	**200,800,212**	**15**
Grand total	**10,187,460,393**	**3,892,169,378**	**38**

Source: EMTS sales record 2008

Definitions of the Variables

Airtime: This is time spent communicating using a mobile phone. Usage includes sending or receiving calls and other wireless transmission such as faxes, e-mail or data files. Most carriers charge for a whole minute even if only part of a minute is used.

Electronic Top Up (E-top up): E-top up is the most advanced and efficient way of recharging prepaid accounts. It is both a fast and readily accessible means of recharging mobile accounts and a highly cost effective means for channel partners and service providers to manage the entire flow of top up process.

A subscriber identity module (SIM): is a smart card inside of a GSM cellular phone that encrypts voice and data transmissions and stores data about the specific user so that the user can be identified and authenticated to the network supplying the phone service. The SIM also stores data such as personal phone settings specific to the user and phone numbers.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

3.1 Preamble

The process used to collect information and data for the purpose of making business decisions. The methodology may include publication research, interviews, surveys and other research techniques, and could include both present and historical information is term Research Methodology (Business dictionary,2016). Methodology implies more than simply the methods you intend to use to collect data. It is often necessary to include a consideration of the concepts and theories which underlie the methods. In regard to system development CMS (2008) defined methodology as a framework that is used to structure, plan, and control the process of developing an information system. A wide variety of such frameworks have evolved over the years, each with its own recognized strengths and weaknesses. One system development methodology is not necessarily suitable for use by all projects. Each of the available methodologies is best suited to specific kinds of projects, based on various technical, organizational, project and team considerations. According to CMS (2008) there are several system development methodology and they are as follows:

1. Rapid Application Development (RAD)

2. Prototyping Methodology
3. Incremental Methodology
4. Spiral Methodology
5. Waterfall Methodology

According to James (1991), **Rapid application development** (**RAD**) is both a general term used to refer to alternatives to the conventional waterfall model of software development as well as the name for James Martin's approach to rapid development. In general, RAD approaches to software development put less emphasis on planning tasks and more emphasis on development. In contrast to the waterfall model, which emphasizes rigorous specification and planning, RAD approaches emphasize the necessity of adjusting requirements in reaction to knowledge gained as the project progresses. This causes RAD to use prototypes in addition to or even sometimes in place of design specifications. According to CMS (2008) RAD framework type is iterative in nature.

Bowman (2009) stated that a **prototyping methodology** is a software development process which allows developers to create portions of the solution to demonstrate functionality and make needed refinements before developing the final solution. It is framework type is it iterative in nature (CMS, 2008).

Iterative and Incremental development is any combination of both iterative design or iterative method and incremental build model for software development. The combination is of long standing and has been widely suggested for large development efforts. During software development, more than one iteration of the software development cycle may be in progress at the same time. This process may be described as an evolutionary acquisition or incremental build approach. The relationship between iterations and increments is determined by the overall software development methodology and software development process. The exact number and nature of the particular incremental builds and what is iterated will be specific to each individual development effort (Craig, 2003).

The Spiral Lifecycle Methodology is a sophisticated lifecycle model that focuses on early identification and reduction of project risks. A spiral project starts on a small scale, explores risks, makes a plan to handle the risks, and then decides whether to take the next step of the project - to do the next iteration of the spiral. It derives its rapid development benefit not from an increase in project speed, but from continuously reducing the projects risk level - which has an effect on the time required to deliver it. Success at using the Spiral Lifecycle Model depends on conscientious, attentive, and knowledgeable management .It can be used on most kinds of projects, and its risk-reduction focus is always beneficial (Technologies Professionals, 2016).

Explanation for waterfall methodology is on section 3.2 of this chapter.

3.2 Methodology Adopted

The methodology adopted in this study is a waterfall methodology. According to Eleven40 Pro Theme (2016) the Waterfall Model was first process model to be introduced. It is also referred to as a **linear-sequential life cycle model**. It is very simple to understand and use. In a waterfall model, each phase must be completed fully before the next phase can begin. This type of model is basically used for project which is small and there are no uncertain requirements. At the end of each phase, a review takes place to determine if the project is on the right path and whether or not to continue or discard the project. In this model the testing starts only after the development is complete. In waterfall model phases do not overlap. The Figure 3.1 is the overview of waterfall methodology.

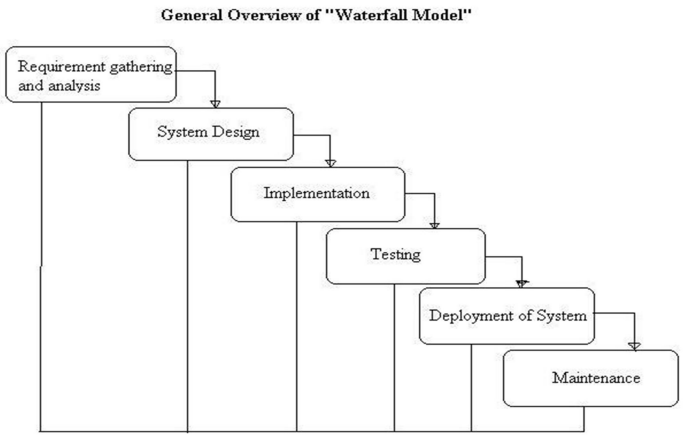


Figure 3.1: Waterfall Methodology (Eleven40 Pro Theme, 2016)

Requirement gathering and analysis as shown in Figure 3.1is known as conception, initiation and analysis (Choudhury, 2011). According to Choudhury (2011) this triggers when a problem is perceived. This phase involves identifying goals to be achieved after the problem is solved, estimating benefits in the new system over the current system, and identifying other areas that are affected by the solution. This phase also involves and developing the business case for the project. A business case provides the information that a manger needs to decide whether to support a proposed project, before resources are committed to its development. In simple term it includes problem definition and data collection of the domain under study. In the present work the problem has been stated in chapter one (section 1.2) and methods for data collection is narrated in this chapter (section 3.3). Initiation involves a macro level study of the user requirements. This phase also involves defining alternative solutions to the user requirements and cost-benefit justification of these alternatives. Analysis involves carrying out detailed study of the user requirements and arriving at the exact requirements of the proposed system and this is discussed in detail in section 3.4. The phase involves freezing the requirements before the design phase begins (Choudhury, 2011).

System Design: According to Choudhury (2011) system design Involves translating the identified requirements into a logical structure, called design that

can be implemented in a programming logic. Thus, the design of the proposed/new system will be discussed in the next chapter.

Implementation: Involves converting the new system design into operation. This may involve implementing the software system and training the operating staff before the software system is functional (Choudhury, 2011).

Testing: Involves integrating and testing all the modules developed in the previous phase as a complete system (Choudhury, 2011).

Deployment: is all of the activities that make a software system available for use.

Maintenance: is a continuous process following system deployment. The activities here includes, system updates, interface modifications, and any other issue to be done as fired from the organizational policies (Choudhury, 2011).

Waterfall Methodology advantages over other mentioned methodologies are as follows (Choudhury, 2011):

1. This model is simple and easy to understand and use.
2. It is easy to manage due to the rigidity of the model – each phase has specific deliverables and a review process.
3. In this model phases are processed and completed one at a time. Phases do not overlap.

4. Waterfall model works well for smaller projects where requirements are very well understood.

The waterfall methodology was chosen in this study because of its orderly sequence of development steps and strict controls for ensuring the adequacy of document and design reviews helps the quality, reliability and maintainability of the developed software. Secondly, progress of the system development is measureable.

3.3 Data Collection

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same (Knatterud, 1998). In this study the data to use for model analysis and evaluation is collected from Telecommunication Service Provider in Nigeria.

3.3.1 Characteristics of the Population

In Nigeria, there many telecommunication service providers but the major telecommunication service providers considered in this study is Emerging Marketing Telecommunications (EMTs) also known as etisalat, Mobile Telephone Networks (MTN), AIRTEL and Globacom (GLO). According to Uduchukwu (2013), it is a widely known fact that life in the world today has been made easier through Information Communications Technology, popularly referred to as ICT. Information communication telecommunication has made every business in the world so easy that many now describe the world as a global village. The Telecommunication service providers listed out operates in the all thirty six states in Nigeria having its offices and outlets in all these states. Nigerian Mobile Telecommunication is the fastest growing market in Africa (Abdul, 2014). It was reported by Ndukwe (2005) that between 1998 and 2000, the number of mobile lines was 35,000 but grew to over 11million as of March 2005, with a growth rate of more than a million lines annually since 2002. This translated to an increase from the total density of 0.4 lines per 100 inhabitants in 1998 to 9.47 lines per 100 inhabitants currently. Additionally, the sector has attracted an investment of over US$8billion.

All the telecommunication service providers in Nigeria offers similar product and service but package in different ways. The operations within the company are almost the same. Using Etisalat for illustration; Etisalat Nigeria divides it operating locality into seven regions namely: Lagos North, Lagos South, South West, South East, South South, North 1 and North 2. The organization has two major department: Retail Sales department and Channel Sales department. Retail Sales department derives its sales from subscribers that walked into etisalat experience centers. While sales from Channel sales department are mainly from distribution partners. Etisalat major product are airtime, electronic recharge and subscriber identity module (simcard) which every other telecommunication providers offers. The major source of revenue to these service provider are sales made from their distribution partners located in all their regions. The training dataset used for this study is from the channel sales department of etisalat.

3.3.2 Sampling Design and Procedure

A sample is a subset of people, items, or events from a larger population that you collect and analyze to make inferences. To represent the population well, a sample should be randomly collected and adequately large (Minitab, 2016). The population considered in the study are decision makers of staff of telecommunication service providers mentioned in section 3.3.1 in this chapter

which involves directors, distribution managers, analysts, state coordinators, specialists, managers, supervisors and team leaders etc. According to career portals of the telecommunication service providers considered in the study, MTN Nigeria number of employee is 17,509, GlO number of employees is 2,545, Etisalat Nigeria number of employee is 5000+ and Airtel Nigeria number of employee is 1000. Adding up the number of employee of these companies, we have a total of 26,054.

In order to calculate the sample size, the Yamane (1964:280) formula for finite population was applied stated as thus-

$$n = \frac{N}{1+N(e)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where

I = constant value

N = population size

e = co-efficient of confidence or margin of error or allowable error or level of significance

n = sample size

$$n = \frac{N}{1+N(e)^2}$$

$$n = \frac{26{,}054}{1+26{,}054\ (0.05)^2}$$

n=394

Therefore the questionnaire will be administered to 394 respondents as the representative of the population (the listed telecommunication providers). The target in administering the copies of questionnaire is the decision makers of the organization.

3.3.3 Data Collection Instruments

Data collection instruments are the means used in collecting data that is the tool used in data collection such as interviews, questionnaire, pilot survey and observation.

1. **Interview:** Some managers of the company were interviewed. The results support the need to embarking on this project. Sample questions used can be referenced at the appendix E.

2. **Questionnaires:** 394 copies of the questionnaires were administered to decision makers of MTN, etisalat, GLO and Airtel telecommunication service provider for information gathering.

3. **Observation:** the telecommunication providers' offices and outlets were visited to observe ongoing activities and operations.

4. **Pilot Survey:** A pilot survey is a strategy used to test the questionnaire using a smaller sample compared to the planned sample size. In this phase of conducting a survey, the questionnaire is administered to a percentage of the total sample population, or in more informal cases just to a convenience sample (Sincero, 2008).

3.4 System Analysis Procedure

The present study adopted system analysis procedure of structural system analysis and design methodology (Gangolly, 1997). According to Gangolly (1997) in Systems analysis phase the present system is investigated and its specifications documented. They should contain our understanding of HOW the present system works and WHAT it does. The Figure 3.2 is system analysis and procedure of SSADM/Waterfall Model according to Gangolly (1997).

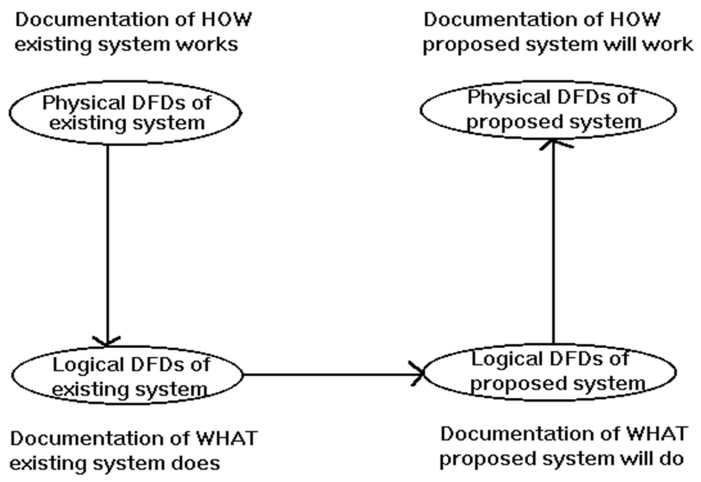


Figure 3.2: System analysis Procedure SSADM/Waterfall (Gangolly, 1997)

The steps of system analysis and procedure of Gangolly (1997) used in this study are presented below.

STEP 0: Defining the scope of the system under study.

As stated on section 1.5 of chapter one, the study area focus in purchasing and sales department of emerging telecommunication services (EMTs) also known as etisalat. Etisalat data warehouse was build with oracle programining language, in order to carryout analysis on the captured data, sequel query language (SQL) is been used to carry out query on the captured historical sales records.

STEP 1: Documentation of how the existing system works. Data will be collected based on these questions:

- Who performs the tasks?
- How they are performed?
- When or how often they are performed?
- How the data is stored (media)?
- How the data flows are implemented (media)?

The data will be collected from the management and products/services. Those in the management are:

The Director: The director oversees the sales affairs of every region and the head of decision making in the organization. It has the responsibility for determining and implementing the company's policy.

Head of Region (HOR): Each region of the organization is been control and manage by the head of region. The head ensures that the region meet its stipulated sales target. Head of region is also part of decision makers of the organization.

Distribution Manager (DM): Reports to head of region the affairs of the region. Here, all dealer specialist operations are been control and manage. A DM is also a part of decision makers in the organization.

Sales Analyst: The sales analyst keeps records and manages the sales activities of the region. It gives the region update of amount sales achieved on daily and monthly basis. The sales analyst works at the regional level, each region have its own sales analyst.

Dealer Specialist (DS): The dealer specialist is the account manager of the distribution partners of the organization. Each state within the region has its own dealer specialist. The dealer specialist ensures that its distribution partner achieved its sales target which is also the state target. The DS overseas the sales made in the state and all organization operational activities within the state.

Distribution Partner is an individual that have a registered company under Nigeria Communication Commission to operate and sales telecommunication product and services. The distribution partner offers the product and services of the company in which it partners with. He operates based on the rules and regulation of the organization in which he is a partner.

The main product and services are as Airtime, electronic recharge (e-top up) and subscriber identity modules (sim cards). Definitions of these terms are stated in section 2.4.1 of chapter two.

STEP 2: Documentation of what the existing system does.

STEP 3: Documentation of what the proposed system will do.

STEP 4: Documentation of how the proposed system will work.

Financial Feasibility Study of the Proposed System

According to Fonollera (2009), a feasibility study (FS) is an evaluation tool used to determine the viability/profitability of a certain idea. It is a tool that systematically explores whether a given idea will work and whether it should be pursued further for implementation. The financial feasibility study determine the amount of money required in realization of the project. The source of financial and cost involved. According to Schwalbe (2012), financial considerations are often an important aspects of the project selection process, especially during tough economic times. There are three primary methods for determining projected financial value of projects (Schwalle, 2012):

1. Net present value analysis (NPV)
2. Return on investment (ROI)
3. Payback analysis

Net present value analysis: is a method of calculating the expected net monetary gain or loss from a project by discounting all expected future cash inflows and

outflows to the present point in time. The NPV is the sum of all the present value of the money that gets into the business (considered positive) and the present value of all the money that gets out of the business (considered negative) at a certain discount rate (Fonollera, 2009). The discount rate is set by the goals of the business. In general a higher positive NPV is preferable. If the NPV is negative it indicates that it is not worth pursuing after all. If NPV is zero, it suggests that you should better off putting the money in a bank. Steps involved in performing the calculation manually for NPV are as follows (Schwalbe, 2012):

1. Determine the estimated costs and benefits for the life of the project and the projects it produces.
2. Determine the discount rate. A discount rate is the rate used in discounting future cash flow. It is also called the capitalization rate or opportunity cost of capital.
3. Calculate the net present value. There are several ways to calculate NPV. The mathematical formula for calculating NPV is:

$$\mathrm{NPV}(i) = \sum_{t=0}^{N} \frac{R_t}{(1+i)^t} \quad \text{.....(2)}$$

Where t equals the year of cash flow. N is the last year of the cash flow. R is the amount of cash flow each year and r is the discount rate.

The formula for the discount factor is $1/(1+r)^t$

Where r or i is the discount rate.

Return on Investment: According to Schwalbe (2012) ROI is the result of subtracting the project costs from benefits and then dividing by costs. ROI is always a percentage. It can be positive or negative. In calculating ROI it better to consider discounted costs and benefits for multi-year projects. Mathematical formula for ROI.

ROI= (total discounted benefits-total discounted)/discounted costs.

The higher the ROI, the better.

Payback Analysis: According to Schwalbe (2012) payback period is the amount of time take to recoup, in the form of net cash inflows, the total invested amount in the project. Payback analysis determine how much time will lapse before accrued benefits overtake accrued costs. It occurs when the net cumulative benefits equals the net cumulative benefits costs equals zero. As stated by Fonollera (2009) the payback period is computed by cumulating the estimated annual cash inflows and determining the point in time at which they equal the investment outlay provided the periodic cash flows are not uniform.

The Cash Flow Statement: According to Fonollera (2009) a cash flow statement is a financial statement that provides information about cash inflows (receipts or sources of cash) and cash outflows (payments or uses of cash) of the business for the given period of time. Cash inflows are money transactions that go into the business while cash outflows are cash transaction that go out of the business. The net cash flow formula is stated in equation (3) (*Schwalbe,* 2012):

NET CASH FLOW= INFLOWS – OUTFLOWS

OR …….. (3)

CASH FLOW= BENEFITS - COSTS

Cost estimate: According to Investorword (2016) cost estimate is an appropriate value of the total cost of a service, product, resources or project used for planning, sales quotes or resource allocation. Cost estimate are generally prepared as accurately as possible to prevent misallocation of resources or negative perceptions from clients, managers or potential customers. Following the steps involves in the ground rule and assumptions for the cost estimates according to Schwalbe (2012) the steps listed below will be used for cost estimate of the proposed system. The steps are as follows:

1. This project was preceded by a detailed study and proof of concept to show that it was possible to develop a model for data mining to assist consultant in data analytic tasks in telecommunication providers.
2. The main goal of this project is to design an enhanced system of data mining in telecommunication service provider to support decision making, continuing developing the software (especially the user interfaces), test the proposed system in telecommunication companies and train 110 data analyst in selected regions to use the proposed system.
3. The project has the following working breakdown structure (WBS):
    a. Project management
    b. Hardware : laptop devices and Servers
    c. Software development
    d. Testing
    e. Training and Support
    f. Reserves
4. Cost must be estimated by WBS and by month. The project manager will report progress on the project using earned value analysis.
5. Costs will be in monetary value. Since the project length is one year, inflation will not be included.

6. The project will be managed by the top management involving directors, head of region, distribution managers, chairman of the executives and vice presidents etc. There will be a part-time project manager and four team members assigned to the project. The team members will help manage various parts of the project and provide their expertise in the areas of software development, training, and support. Their total hours will be allocated as follows: 25 percent to project management, 25 percent to software development, 25 percent to training and support and 25 percent to no-project work.
7. The project involves purchasing of laptops. The laptops will be used by 110 analysts. The cost rate is estimated to be ₦80000 per unit. The project will require four additional servers to run the software required for the devices and for managing the project.
8. The software development includes developing a graphical user interface for the devices, an online help system, and a new modules.
9. Testing cost should be low. An estimate based on 10 percent multiplied by the total hardware and software estimate should be sufficient.
10. Training will include instructor-led classes in seven different locations.
11. Because there are several risks related to this project, include 20th percent of the total estimates as reserves.

12. A computer model will be develop for the estimate, making it easy to change several inputs, such as labour hours for various activities or labor rates.

Benefit Estimation: According to Goldsmith (2006) benefit estimate is a technique used for estimating the direct (tangible) or indirect (intangible) favorable results of an action taken. Tangible benefits can be measured in terms of their direct monetary value, such as increase in revenue; intangible benefits are the favorable results that may not be measurable in money terms, such as improved morale of the employees, but can be estimated using a qualitative approach. Goldsmith (2006) present the procedure below for benefit estimate:

1. Define and calculate all tangible benefits consistent with the impact of the proposed change. Identify potential recurring and nonrecurring benefits associated with the process. Examples include:

   a. Nonrecurring benefits:

      i. Cost reduction resulting from reduced resource requirements and increased productivity.

      ii. Value-added enhancement reflected in reduced product defects and increased revenue from faster time-to-market.

b. Recurring benefits - the savings realized by lower operating costs for items such as:

i. Salaries

ii. Equipment lease

iii. Training and travel

2. Define all intangible benefits.
3. Spread the money associated with the benefits across the planning period in a tabular format.

3.4.1 Performance Indicator of the Existing System

The performance indicators identified in this study for evaluation and assessment are data qualities which comprises accuracy, interpretability, presentation quality (understanding/visualization), accessibility, consistency, easy to use, precision, concise, robustness, correctness, completeness, speed (response time), relevance, reliability and unambiguous.

Data quality: According to Rouse (2005) data quality is a perception or an assessment of data fitness to serve its purpose in a given context.

Accuracy: The accuracy of a model is an indicator of its ability to predict the target class for future observations. The most basic indicator is the proportion of observations of the test set correctly classified by the model. Similarly, the error

rate is calculated as the ratio between the number of errors and the number of cases examined (Clemente et al., 2017).

Speed: Speed up data mining process is an important economic goal (Skalak, 2002). In this present work result are to be generated in real- time in order to close the gap of the existing system. The ability for the model to process its input in order to produce its output in a real time.

Visualization: The main objective of data visualization is the overall idea about data mining model. In data mining most of the times retrieving data from repositories which are in hidden form is a difficult task for a user "so this visualization of the data mining helps to provide high levels of understanding and trust. Data visualization is to let the user understand what is going on (Fayyad et al., 2001).

The visualization tool used in the course of study is Dashboard. Dashboard is self-explanatory.

ROBUSTNESS: Robust is characteristic describing a model's, test's or system's ability to effectively perform while its variables or assumptions are altered. A robust concept can operate without failure under a variety of conditions. Robustness can relate to both economic and statistical concepts. For statistics, a test is claimed as robust if it still provides insight to a problem despite having its assumptions altered or violated. In financial markets that continue to perform

despite alterations in market conditions in general, being robust means system can handle variability and remain effective.

Accessibility: The extent to which data is available, easily or quickly retrievable (Wang, 1996).

Completeness: The extent which data is not missing and is of sufficient breadth and depth of the task at hand (Wang, 1996).

Interpretability: The extent to which data is in appropriate languages, symbols and units the definitions are clear (Wang, 1996).

Easy to use: The extent to which data is easy to manipulate and apply to different tasks (Wang, 1996).

Consistent Representation: The extent to which data is presented in the same format (Wang, 1996).

Concise Representation: The extent to which data is compactly represented (Wang, 1996).

Understandability: The extent to which data is easily comprehended (Wang, 1996).

Relevance: Every piece of information stored is important in order to get a representation of the real world (Bobrowski et al., 1999).

Reliability: The data stored is trustable, i.e., it can be taken as true information (Bobrowski et al., 1999).

Unambiguity: each piece of data has a unique meaning (Bobrowski et al., 1999).

Standard of Measurement.

The study adopted a questionnaire-based data quality methodology as measurement standard for evaluating the performance indicators identified (Vaziri and Mohsenzaden, 2012). This method can be as well as called Delphi methods (Berry, 1994).

According to Vaziri and Mohsenzaden (2012) a single aspect of data quality is defined as a "dimension" such as "consistency", "accuracy", "completeness", or "timeliness". In order to assess and improve data quality, "methodologies" have been defined. Data quality methodologies are sets of guidelines and techniques that are designed for measurement assessment, and perhaps, improving data quality in a given application or organization. If an appropriate list of dimensions is available for the specific needs of an organization, a questionnaire-based methodology can be designed in order to 1. Measure dimensions and identify "weak" dimensions in the organization 2. Select a proper "strategies" to improve data quality. Batini (2009), divided methodologies into three main "phases and steps". The three steps and phases are the following:

1. **State reconstruction**: which collects contextual information on organizational data, processes and services.

2. **Assessment/Measurement**: which measures the quality of data along relevant "dimensions". The term "measurement" refers to measuring the values of data itself, and the term "assessment" refers to comparison against reference values.

3. **Improvement**: which proposes techniques and strategies for reaching higher levels of data quality, perhaps levels specified by the organization's management.

Vaziri and Mohsenzaden (2012) proposed a questionnaire-based state reconstruction phase to identify the relevant dimensions for the organization or application at hand. Measuring the relevant dimensions by a questionnaire-based methodology could serve several purposes.

1. It gives an overall measurement of the current quality of data in the organization.

2. It identifies "weak" dimensions, namely dimensions that require urgent and thorough attention in the organization.

3. It provides values for assessment purposes. The assessment is when we compare data against reference values. In the present study the reference value are noted below.

Measurement: Now we identify three groups of subjects and present them with a comprehensive measuring questionnaire. The three groups of subjects are:

Information Professionals (IPs), Information Consumers (ICs) and Independent Experts (IEs) (Vaziri and Mohsenzaden, 2012).

Information Professionals (IPs): These are the people who collect and maintain the information for an organization. They are also responsible for designing the systems where information resides.

Information Consumers (ICs): These are the people who use the information such as decision makers.

Independent Experts (IEs): These are defined as experts that have appropriate amount of practical or academic experience in the practices of the organization being evaluated. Also they are called independent because they have no vested interest in the organization being evaluated and thus can present an unbiased opinion.

First the questionnaire collects general information about each subject, such as Name, Family Name, Organization, etc. It also asks about the "Role" of each subject in the organization for later "gap analysis". Then the questionnaire covers all the relevant dimensions, and asks the subjects to rate each dimension from a scale of 1 to 5 with 0 being rated as "extremely" and 5 as "extremely high". The exact "definition" of each dimension is included in the questionnaire to avoid incoherent interpretation of the dimensions. The measurement value of each dimension can be asked by several items (questions). The number of items for each dimension should not be too big. Too many questions can tire the subjects and lower the effectiveness of the questionnaire. The groups of items that target

specific dimensions can be mixed at random so that the subjects will not detect that each dimension is being questioned several times. This way, we could have several independent measurements for each dimension. For each single dimension these are the items that we propose (Vaziri and Mohsenzaden, 2012):

Direct Question:

Is the Data in your organization [Dimension name]?

Reverse Question

Is the Data in your organization [Opposite of dimension]?

Synonymy Question:

Is the Data in your organization [Dimension Synonym]?

Definition Question:

Is the Data in your organization [Definition of Dimension]?

The first item directly asks the subjects about the specific dimension. This item measures the very first impression of the subjects about a dimension clearly and directly. Its simplicity and directness is its feature advantage. The second item asks about the value of a dimension, but in reverse. This item is intended to obtain the subjects opinion from a negative point of view. Of course, the value of this item must be subtracted from 5. The third item measures the dimension if it were given another name. This items tests to see if the subjects are biased towards a term that they have seen lots of times in the literature before. And the fourth item measures

the value of a dimension in terms of its definition. The definition of the dimension could be based on a well-known paper such as (Wang, 1996). This items tests whether the subjects have a good understanding of the meaning of the dimension. For each dimensions the "mean value" of its related items makes up its measured value.

Improvement

Once the "relevant dimensions" have been measured the next task at hand is to identify a proper method to improve data quality. Since we have followed a questionnaire-based study, we will continue with the same strategy in the third and final phase of our methodology namely the improvement. For each dimension's measured value a system of classification like the following can be devised:

Strong: 3.0 <Dimension Value

Intermediate: Equal to 3

Weak: Dimension Value < 3.0

Obviously, for data quality improvement the methodology should begin by the "Weak" dimensions. Depending on the resources and funding available in the organization for data improvement, the organization can decide whether to move on to "Intermediate" and "Strong" dimensions as well. A questionnaire must be developed for each of the "weak" dimensions.

Evaluation

THE CONTROL MATRIX: Pierce (2004) proposed a control matrix which can be used to relate data quality problems with data quality controls. This could also serve as an evaluation tool for our questionnaire results. In such "Control Matrices" the rows identify the various quality checks or controls that are available or proposed and the columns identify the data quality problems that have been identified in an organization. Each cell in the control matrix specifies how effective a data control or check is for a data problem. Here is a simple example in Table 3.1:

Table 3.1: The Control Matrix for Improvement Strategies (Pierce, 2004)

	DQ Problem 1	DQ Problem 2	DQ Problem 3	DQ Problem 4
Control 1		Medium		
Control 2	Medium			
Control 3		Strong		Weak
Control 4		Strong		
Control 5			Weak	Weak

The previously-discussed questionnaires can help us fill out the control matrix. The columns, which are the data quality problems, in the case of our methodology are the dimensions that have been identified as *weak* by the subjects. On the other hand, the rows are controls or improvement techniques that have been proposed by

the improvement questionnaire. Notice that the results from the same questionnaire can also be used to rate the effectiveness of the strategy so that the associated cell in the matrix could be filled as **Weak**, **Medium**, or **Strong**. Filling out the above matrix can additionally help with the improvement techniques. For instance, in the above matrix DQ Problem 2 has two strong controls associated with it, thus it is safe to assume that the problem is getting good attention. On the other hand, DQ Problem 3 has only one weak control associated with it. So in the improvement stage, the IE's, and perhaps IP's, must introduce new controls for DQ Problems that have week or no controls.

Gap Analysis and Ranking

Since we are using three groups of subjects to measure and improve data, gap analysis and ranking could be done on both.

1. Gap Analysis: If all groups of subjects agree on a measurement or improvement technique it is safe to say that a reliable result has been reached. However, if there is serious disagreement between the various groups further analysis is required to explain the differences. For instance, if there is a large gap between IP's and IC's, the information professionals are not aware of the consumers' concerns. On the other hand, if there is large gap between the IE's and other two groups, there seems to be a lack of expertise among professionals and consumers.

2. Ranking: Since all the dimensions and improvement strategies are measured by a scale of 1 to 5 a ranking of dimensions from weakest to strongest could be performed. This is very useful when an organization has limited resources in terms of data quality improvement. In such cases, an organization could concentrate on the weakest dimensions first and, depending on the resources, continue on to the other dimensions. On the other hand, an organization finds the most benefit by following the improvement strategies that have been identified as the most useful.

3.4.2 Performance Indicator of the Proposed System

In assessing the performance of the proposed system, the performance indicator identified are the same of that of existing system. The performance indicator accuracy, interpretability, presentation quality, accessibility, consistency, easy to use, precision, concise, robustness, correctness, completeness, speed, relevance, reliability and unambiguous.

Standard of Measurement. The questionnaire-based data quality methodology (Delphi Method) will also be used for measuring the performance indicators.

CHAPTER FOUR

4.0 DESIGN AND IMPLEMENTATION

This chapter presents the physical/logical design, system analysis of the existing and proposed system, financial analysis, implementation, testing and results achieved were also discussed.

4.1 Data Preparation and System Analysis

4.1.1 Data Preparation

This is the process of gathering, combining, structuring and organizing data so it can be analyzed as part of business intelligence (BI) and business analytics (BA) programs (Rouse, 2016). The components of data preparation include data discovery, profiling, cleansing, validation and transformation; it often also involves pulling together data from different internal systems and external sources (Rouse, 2016).

In this study data collected for preparation falls into two categories, which are secondary data and primary data.

Secondary data: In this study, a historical sales data starting from 2008 to 2015 was collected from Emerging Market Telecommunication services (etisalat) as a training data set for the model. The data was prepared into the required format in

order to build the data warehouse. The existing system choose of programmining language used to build the warehouse is oracle, there is no embedded or unified tool used for data analysis. The database design is presented in section (4.2.1) of this chapter

Primary data: In this study information was collected through personal interviews, pilot survey, observation and questionnaires. 394 copies of questionnaire was distributed to telecommunication providers namely MTN, Airtel, GLO, Etisalat and experts in the field of study. However, due to several factors a total of 247 respondents returned their responses. This represent 62.69% returns. It is worthy to note here that method of return of questionnaire was by hand delivery and via e-mail address. The analysis of data collected for the study is presented in this chapter (section 4.1.2)

4.1.2 System Analysis

The present study adopted system analysis and procedure of structural system and design methodology (Gangolly, 1997). The steps involved in the analysis was enumerated in chapter three (section 3.4).

Analysis of Data Collected from the Existing System.

The present system adopted Delphi method (questionnaire-based data quality methodology) in data collection and standard measurement for evaluation of the

existing system and proposed system. The steps involved in the process was enumerated in chapter three (section 3.4.1).

Assessment/Measurement: **Information Professionals (IPs)**

A total number of 135 copies of questionnaire was received from people identified in this role. Table 4.1 present control matrix of the data collected in these role.

Table 4.1: Control Matrix for Information Professionals.

Performance indicator	**Extremely low**	**Low**	**Neutral**	**Very high**	**High**	**Rank**
Likert scale grade	**1**	**2**	**3**	**4**	**5**	
Accuracy	0	0	0	47	88	**5**
Interpretability	0	0	7	71	57	**5**
Presentation	1	67	55	10	0	**2**
Accessibility	20	59	56	0	0	**2**
Consistency	0	0	0	67	68	**5**
Easy to use	0	57	40	33	5	**2**
Precise	40	59	20	10	6	**2**
Concise	0	0	0	85	50	**4**
Robust	17	20	48	50	0	**4**
Reliable	0	0	25	50	60	**5**
Unambiguous	0	9	57	58	11	**4**
Grade in minutes	**1hr/more**	**45mis**	**30mins**	**20min less**	**5mins**	**Rank**
Speed (response in time)	0	0	26	109	0	**4**

The weak performance indicators identified in this role are presentation quality, accessibility, precise and easy to use.

Assessment/Measurement: **Information Consumers (ICs)**

A total number of 58 copies of questionnaire was received from people identified in this role. Table 4.2 present control matrix of the data collected in these role.

Table 4.2: Control Matrix for 1nformation Consumers.

Performance indicator	**Extremely low**	**Low**	**Neutral**	**Very high**	**High**	**Rank**
Likert scale grade	**1**	**2**	**3**	**4**	**5**	
Accuracy	0	0	0	3	55	**5**
Interpretability	0	0	0	2	56	**5**
Presentation	0	28	20	0	0	**2**
Accessibility	8	40	10	0	0	**2**
Consistency	0	0	0	2	56	**5**
Easy to use	0	30	13	15	0	**2**
Precise	23	26	6	2	1	**2**
Concise	0	0	0	10	48	**5**
Robust	0	28	15	5	10	**2**
Reliable	0	0	0	23	28	**5**
Unambiguous	0	18	5	15	20	**5**
Grade in minutes	**1hr/more**	**45mis**	**30mins**	**20min less**	**5mins**	**Rank**
Speed (response in time)	14	27	13	4	0	**2**

The weak performance indicators identified in this role are presentation quality, accessibility, precise, easy to use, speed (response in time) and robust.

Assessment/Measurement: **Independent Experts (IEs)**

A total number of 27 copies of questionnaire was received from people identified in this role. Table 4.3 present control matrix of the data collected in these role.

Table 4.3: Control Matrix for Independent Experts.

Performance indicator	**Extremely low**	**Low**	**Neutral**	**Very high**	**High**	**Rank**
Likert scale grade	**1**	**2**	**3**	**4**	**5**	
Accuracy	0	0	0	2	25	**5**
Interpretability	0	0	0	6	21	**5**
Presentation	3	18	6	0	0	**2**
Accessibility	2	19	7	0	0	**2**
Consistency	0	0	0	5	23	**5**
Easy to use	0	0	23	5	0	**3**
Precise	5	18	5	0	0	**2**
Concise	0	0	0	4	24	**5**
Robust	0	13	14	0	0	**3**
Reliable	0	0	2	12	13	**4**
Unambiguous	0	2	13	12	0	**3**
Grade in minutes	**1hr/more**	**45mis**	**30mins**	**20min less**	**5mins**	**rank**
Speed (response in time)	0	0	25	2	0	**3**

The weak performance indicators identified in this role are presentation quality, accessibility and precise.

Gap Analysis: Based on the outcome of the assessment carried in the existing system, all the identified roles produce same weak performance indicators expect Information consumers which identified some additional indicators as weak which

others did not identified as weak indicators. The indicators identified as weak by this role are robust, easy to use and speed (response in time). Since the proposed system is to support decision making and the information consumer are the decision maker as defined in chapter three (section 3.4.1). Therefore, robust and speed are also a weak performance indicator of the existing system.

Weak Performance Indicator of the Existing System

The identified weak performance indicators of the existing system are presentation quality, accessibility, easy to use, precise, and robust and speed (response in time).

4.1.2.1 Financial Analysis

In cost estimate ground rule and assumptions for the cost estimates discussed in section 3.4 of chapter three was applied (Schwalbe, 2012).

Cost Estimation: The details of the work base structure (WBS) are as follows:

1. Project management: estimate based on compensation for the part-time project manage and 25% of the team members' time. The budget experts for this project suggests using a labor rate ₦1000/hour for the project manager and ₦750/hour for each team member, based on working an average of 160 hours per month, full time. Therefore, the total hours for the project manager under this category are 960 (160/2 *12=960). Costs are also

included for the four project team members working 25% of their time each or a total of 160 hours per month for all project personnel (160 *12 = 1920). An additional amount will be added for all contracted labor, estimate by multiplying 10% of their total estimates for software development and testing costs (10% * (5,940,000 + 1,554,000))

2. Hardware
    a. Laptops devices: 110 devices estimates for the analysts is ₦80000 per unit.
    b. Servers: four servers estimates ₦200, 000 each.
3. Software: software development will use two estimate approach; a labor cost and function point estimate. The approach that produce higher result, its value will be considered in the calculation.
4. Testing: based on similar projects, testing will estimate as 10% of the total hardware and software cost.
5. Training and support: based on similar projects, training will be estimated on a per- trainee basis, plus travel costs. The cost per trainee (1000*110) will be ₦110, 000 and travel be ₦700/day/person for the instructors and project team members. It is estimate that there will be a total of 12 travel days (700*12) will be ₦8, 400. Labor costs for the project team members will be

added to this estimate to assist in training and providing support after the training. The labor hour's estimate for team members is 1,920 hours total.

6. Reserves: reserves will be estimated at 20% of the total estimate.

The Table 4.4 shows the cost model using the above information.

Table 4.4: Proposed system cost estimate.

	#Units/Hrs	Cost/Unit/Hr	Subtotals	WBS level (1 totals)	% of Total
WBS items					
1. Project management				₦1,709,400	
Project manager	960	₦1000	₦960,000		
Contractors (10% of software dev & testing)			₦749,400		
2. Hardware				₦9,600,000	
a. Laptops devices	110	₦80000	₦8,800,000		
b. Servers	4	₦200000	₦800,000		
3. Software				₦5,940,000	
Software development			₦5,940,000		
4. **Testing** (10% of total hardware and			₦1,554,000	₦1,554,000	

software costs)					
5. Training and support				₦1,558,400	
Trainee cost	110	₦1000	₦110000		
Travel cost	12	₦700	₦8400		
Project team members	1920	₦750	₦1,440,000		
6. Reserves(20% of total estimate)				₦4,072,360	
Total project cost estimates				₦24,434,160	

Software development Estimate: the Table 4.5 present the cost estimate for the software development.

Table 4.5: Software development cost estimate (Roetzheim, 2003)

	#Units/Hrs	**Cost/Unit/Hr**	**Subtotals**	**Calculations**
1. Labour Estimate				
Contractor labor estimate	3000	₦1,500	₦4,500000	3000*1500
Project team member estimate	1920	₦750	₦1,440,000	1920*750
Total labor estimate			₦5,940,000	Sum above two values
2. Function estimate				
Source lines of codes (SLOC)			19,527	Functions points*program lang

				equivalency value
Productivity KSLOC penalty (in months)			29.28	3.13*8.05*1.072
Total labour hours (160hrs/month)			4,684.65	29.28*160
Cost/labour hour (₦1200)			₦1200	Assumed value from budget expert
Total function point estimate			5,621,580	4684.65*1200

Benefit Estimate

In benefit estimate the procedure reviewed in section 3.4 of chapter three was applied (Goldsmith, 2006).

Procedure: The tangible benefits identified that are consistent with the proposed system are:

1. Cost reduction: Cost resulting from reduced requirement in the subsequent year after implementation and increase productivity. The cost involves are software development cost and hardware cost, testing cost. The amount calculated from the cost estimate is to be used.
2. Value – added enhancement reflected in reduced product defects and increase revenue from the faster time-to market. The estimate value used is the organization total sales figure of 2014 gotten from the secondary data

from the management before implementation of proposed system divided by 12 to obtain a month amount (N129,966,533,964/12= ₦10,830,544,497)

3. Salaries: The salaries of project manager, team members and contractors. In most cases, IT projects are being handle by outsource firms for organization. The development team may not be staff of the organization, so after project completion salaries pays to this group will be saved by the organization.
4. Training and travel: The staff involved will be trained after the proposed system is developed, in the subsequent years before any other modification or development on the system that will not be need for training and travelling, the organization will not spend such amount again which is also a benefit to the organization. The benefit estimate will be the initial cost of training and travel obtained in cost estimation in Table 4.4. Table 4.6 represent the benefit estimation analysis worksheet

Table 4.6: Benefit estimate worksheet (Goldsmith, 2006)

	Cost/Unit	**Subtotals**
Testing cost	N1,554,000	₦1,554,000
Software development cost	N5,940,000	₦5,940,000
Hardware cost	N9,600,00	₦9,600,000
Value-added enhancement	N10,830,544,497	₦10,830,544,497
Project manager	N960,000	₦960,000
Contractors	N749,400	₦749,400
Team member	N110,000	₦110,000

cost		
Training cost	N8400	₦110,00
Total tangible benefits		₦10,849,466,297

Net Present Value Analysis (NPV).

The investment year or years for project cost is denoted as Year 0 and no discount costs will be done in Year 0 and ongoing system costs and projected benefits are included for Years 1, 2, and 3. The discount rate can also vary, often it is based on the prime rate and other economic considerations. Based on the advice of the financial experts in EMTs, the discount rate used is 8%. The Table 4.7 below is the NPV analysis:

Using mathematical formula for calculating NPV as stated in chapter 3 (section 3.4)

$$\mathrm{NPV}(i) = \sum_{t=0}^{N} \frac{R_t}{(1+i)^t} \text{---------- (1)}$$

Year 0: discount factor = $1/ (1 +0.08)^0 =1$

Year 1: discount factor = $1/ (1+ 0.08)^1 = .93$

Year 2: discount factor = $1/ (1+ 0.08)^2 = .86$

Year 3: discount factor= $1/ (1+ 0.08)^3 =.79$

Subsequent cost is obtained by reducing initial cost by 28.57% (Schwalbe, 2012)

Table 4.7: Net Present Value and ROI analysis.

Discount rate	**8%**					
The project is completed in Yr 0			**Year**			
	0	**1**	**2**	**3**	**Total**	
Costs	24,434,160	6,980,840	6,980,840	6,980,840		
Discount factor	1	0.93	0.86	0.79		
Discounted cost	**24,434,160**	**6,492,181**	**6,003,522**	**5,514,864**	**42,444,727**	
Benefits	0	10,849,466,297	10,849,466,297	10,849,466,297		
Discount factor	1	0.93	0.86	0.79		
Discounted benefits	**0**	**10,090,003,656**	**9,330,541,015**	**8,571,078,375**	**27,991,623,046**	
Discounted benefits-costs	(24,434,160)	10,083,511,475	9,324,537,493	8,565,563,511	**27,949,178,319**	**NPV**
Cumulative benefits-costs	(24,434,160)	10,059,077,315	19,383,614,808	27,949,178,319		
		Payback in Yr 1	Yr 2	Yr 3		
Return of investment (ROI)	65.85%					

An ROI of 65.85% is outstanding and the project produce a positive and higher NPV. Therefore, it is worth to be considered.

Problems of the Existing System

1) The difficulty in gaining a precise view of target area in a collated voluminous business transaction data by decision makers of an organization and presenting the information gained in a real time.
2) The difficulty in accessing the system for information retrieval and the existing system seems to be difficult to be used for data analysis.
3) The existing system is not robust and this pose difficulty in data exploratory analysis and manipulation.
4) The need to enhance the presentation quality of information used for decision making to increase understanding and trust in information assimilation.

4.2 System Design

System design is *the process of defining the components, modules, interfaces, and data for a system to satisfy specified requirements* (Blanchard and Fabrycky, 2010).

4.2.1 Logical System Design

Logical system design pertains to an abstract representation of the data flows, inputs and outputs of the *system.* This is often conducted via modelling, using an over-abstract (and sometimes graphical) model of the actual *system*. In the context of systems, *designs* are included *(Ulrich and* Eppinger*, 2000)*.

4.2.1.1 Data Flows Model of the existing System.

The components in the data flow diagrams presents below has its definitions in section 3.4 of chapter three.

Level 0 logical dataflow diagram: Data warehouse and processing system of sales records. The data warehouse contains the historical sales record of the purchasing and sales department. This records are been retrieved when needed for business analysis and evaluation by the management.

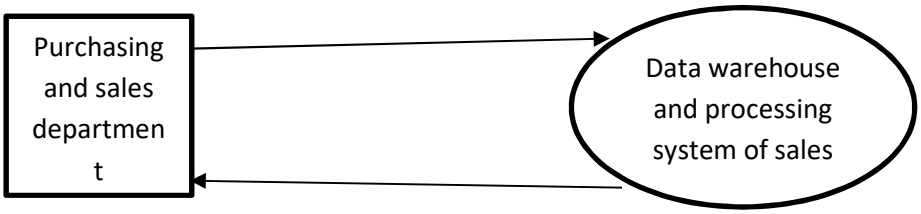


Figure 4.1: Level 0 DFD of data warehouse and processing system of sales records

Level 1 logical DFD: The relationship between the features of the existing system. The component identified in this level their definitions, roles and operation were

discussed in chapter three (section 3.4). The logical diagram depicts the actors that have access and flow of information in the entire organization and regional level.

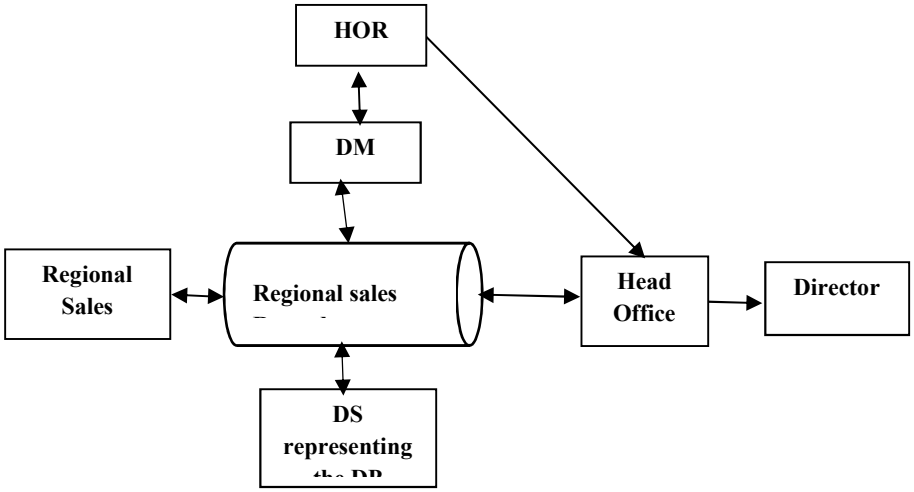


Figure 4.2: Level 1 DFD the relationship between the features of the existing system

Level 2 logical DFD: The distribution partners flow of products within the existing system. The regional sales record is aggregated sales records from all the region operational warehouse stores in the data warehouse. The sales records are derived from each states and it is based on records of sales on products made by the distribution partners of the company. In this level the process in which the distribution partner purchases goods and services from the organization is described. Figure 4.3 shows the Level 2 DFD.

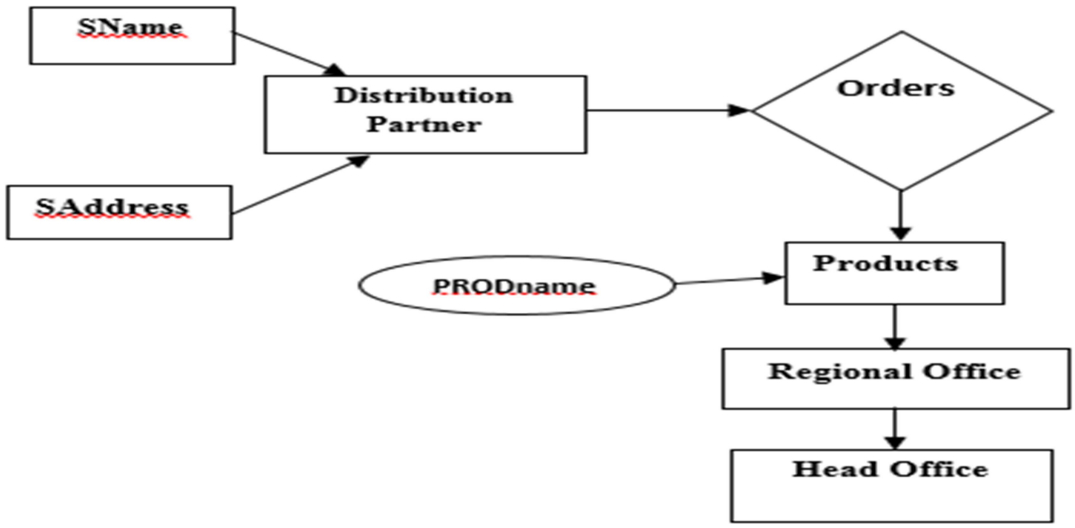


Figure 4.3: Level 2 DFD the distribution partners' flow of products within the existing system

Level 3 logical DFD: Business operation in the existing system. The items identified in this level has its definition in chapter three (section 2.4.1). Sales record comprises of three products namely Etop-up, Sim cards, and Airtimes. In business performance evaluation and analysis in existing system; they are three major products that made up the sales record that are been accessed and analyzed by the analyst and members of decision makers. Figure 4.4 is the Level 3 DFD.

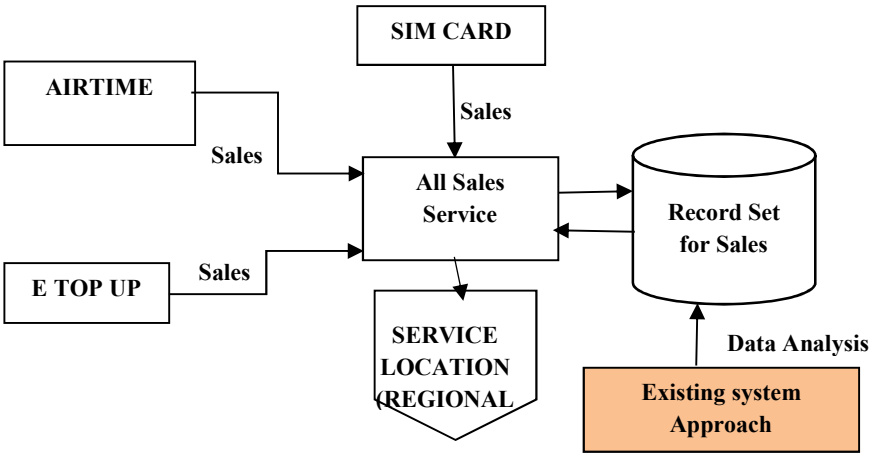


Figure 4.4: Level 3 DFD business operation in the existing system

4.2.1.2 Use Case of the Existing System

Therefore, figure 4.5 presents the use case diagram for the main entities of the existing system operational activities. The roles of the actors and their activities were enumerated in chapter three (section 3.4). The use case represent the level of each actor involvement in accessing and retrieval of information in the existing system.

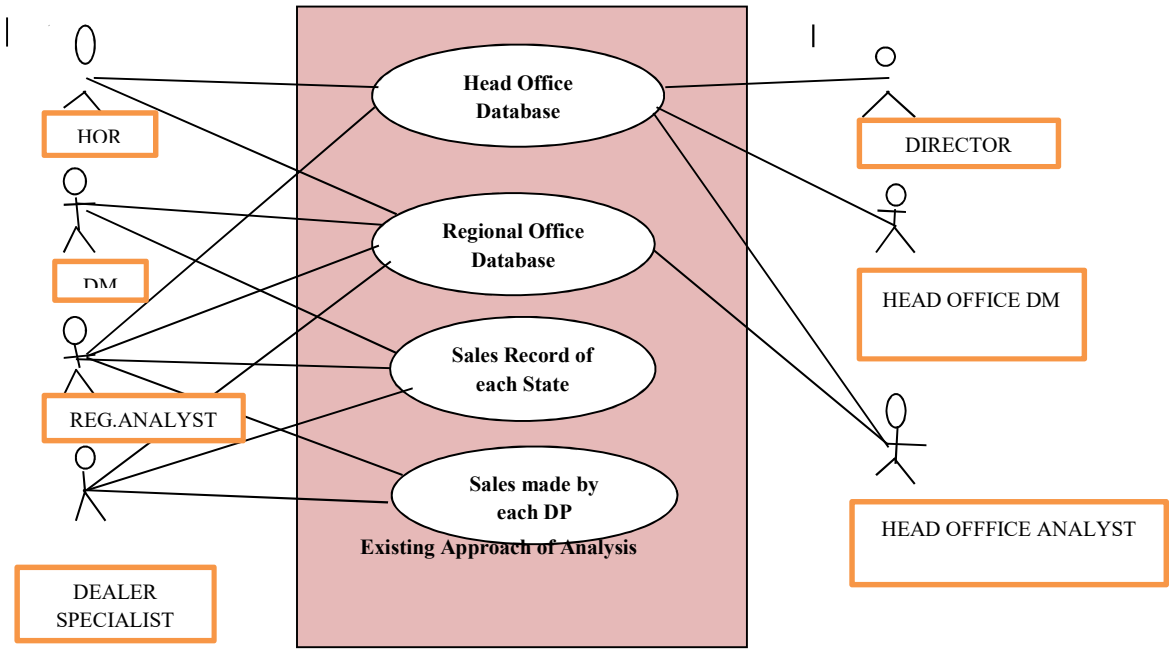


Figure 4.5: Use Case of the existing system

4.2.1.3 Logical Data Flow Model of the Proposed System

Level 1 Data flow Diagram (DFD): The business operation of the proposed system as discussed in level 3 logical DFD in (subsection 4.2.1.3). The process of accessing and analyzing the record in repository has been replaced by proposed system data mining approach. Figure 4.6 is the Level 1 DFD.

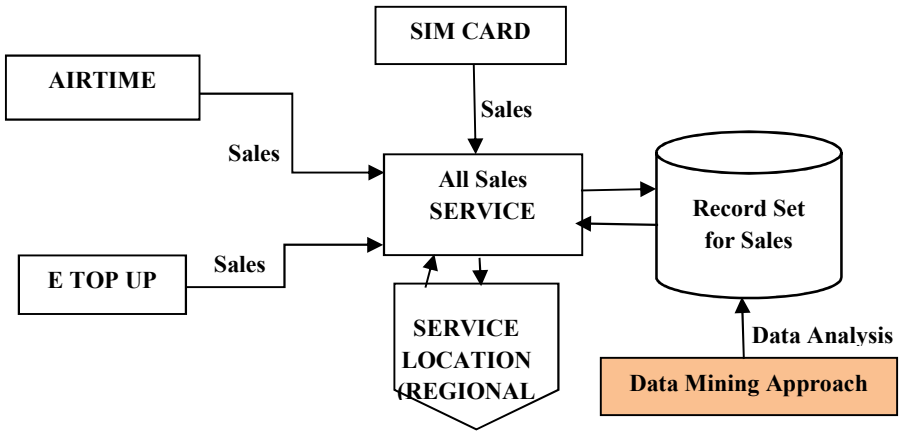


Figure 4.6: Level 1 DFD business operation in the proposed system

4.2.1.4 Use Case of the Proposed System

Figure 4.7 represents the use case diagram for the main entities of the proposed system operational activities they perform. Here data mining approach of data analysis is employed for carrying analysis to overcome the bottlenecks of the existing system approach. The data mining techniques alleviate the difficulty in accessing records in data warehouse, eliminate delay in response in time and allow the actors identified to have a precise view of the targeted area.

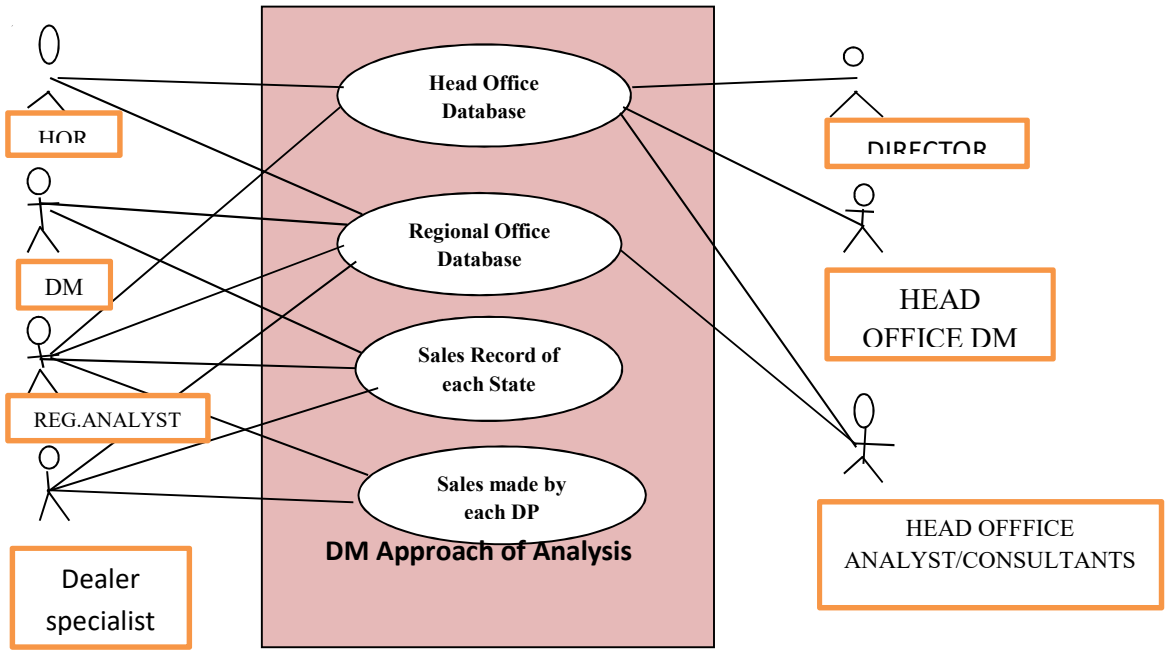


Figure 4.7: Use case of the proposed system

4.2.1.5 High Level Models

System Architecture with Server Technology

This work imitates a typical four tier system and the components have its meaning as follows:

1. **Server Suite:** This comprises of three component Network server which manages the entire network processes in the system; HTTP server which manages all http requests in the system and also all processes linking the system to another network; data Mining server which creates and processes complex data mining activities/procedures

2. **Process Server:** This comprises of the two components DBMS (Database Management System) which is responsible for the retrieval, inserting, operation and maintenance of Data in the data repository while data repository which store the actual data in the system.
3. **Process Node:** This module is responsible for channeling users request to the proper channel for effective and efficient responses.
4. **User Interface:** This is the Graphical User Interface (GUI) component of the system which includes where the administrator can input data can carry out other operations as required. This is also includes screens and forms for users to view the profiles and carry out other operations as deemed fit. Tools for developing this component include HTML, CSS and JavaScript.

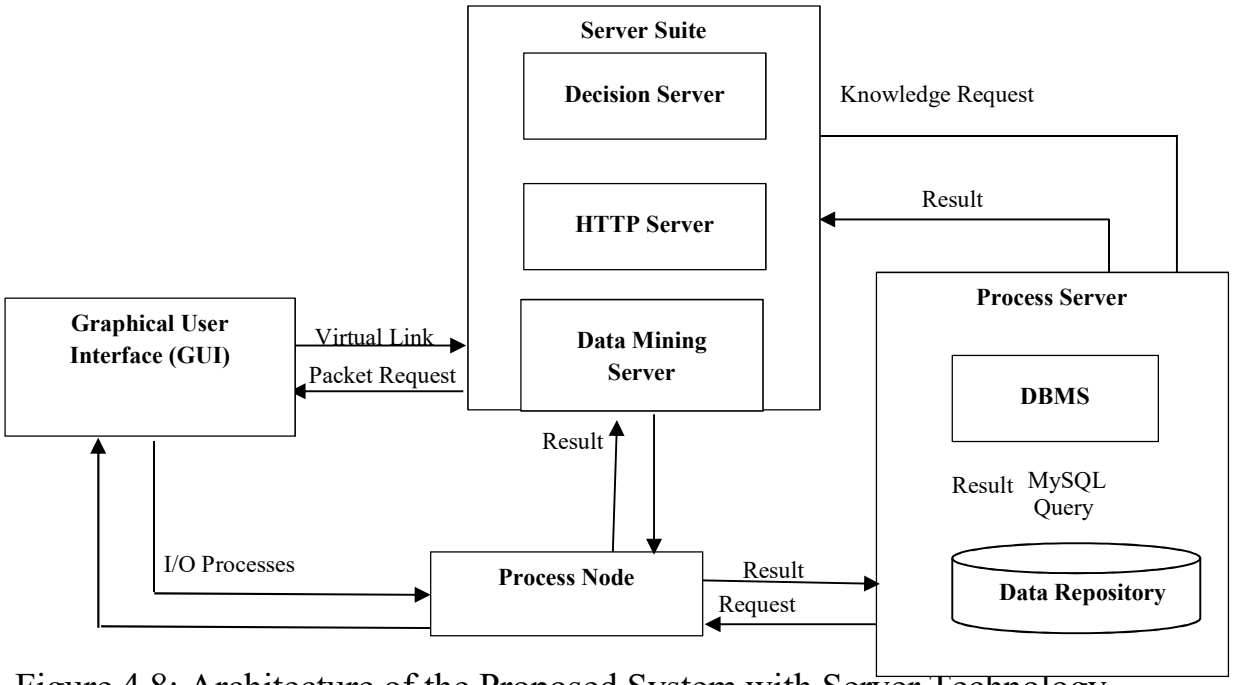


Figure 4.8: Architecture of the Proposed System with Server Technology

Framework of the Proposed System

Figure 4.9 illustrates the operational 3 –tier Framework of the proposed system. This diagram shows the communication existing between the users (from the interface), Sales records, Database storage, and Data mining approach with respect to the server infrastructure that controls all the entities.

Graphical User Interface: This module is responsible for the interaction/communication between the system and humans. The users are the top management and the Analyst.

Sever Infrastructure: Enabling component communication and access to privilege users to access the dynamic web document.

Database Server: This module is responsible for storage of data within the system to be used and sent to/from the system to other devices.

Data Mining Approach: Involves the application of data mining techniques such as classification, association, sequence discovery and prediction etc on the mine out data in order to carry out analysis on the record.

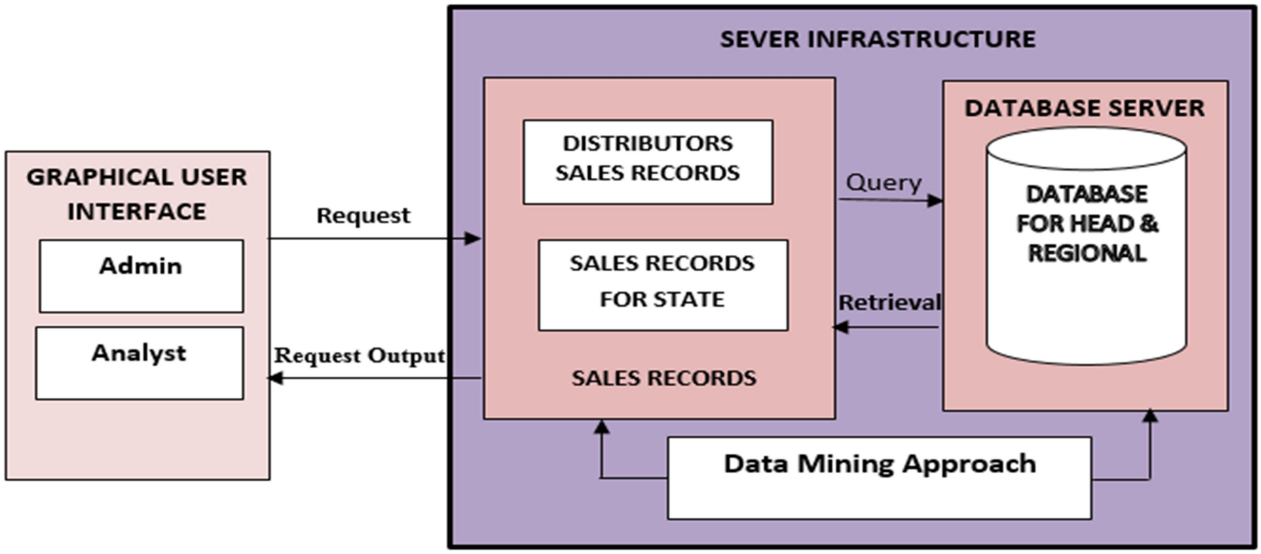


Figure 4.9: Framework of the proposed system.

4.2.1.6 Input/output Design

The Input module is the interface through which the user communicate with the system by supplying an input. The Output module provides scripts (documents, results based on the operations performed in the system.

Input Design

The system uses this module to capture information from external environment.

(a) New User Registration: The template captures the details of an individual that want to access the system. This is shown in Figure 4.10

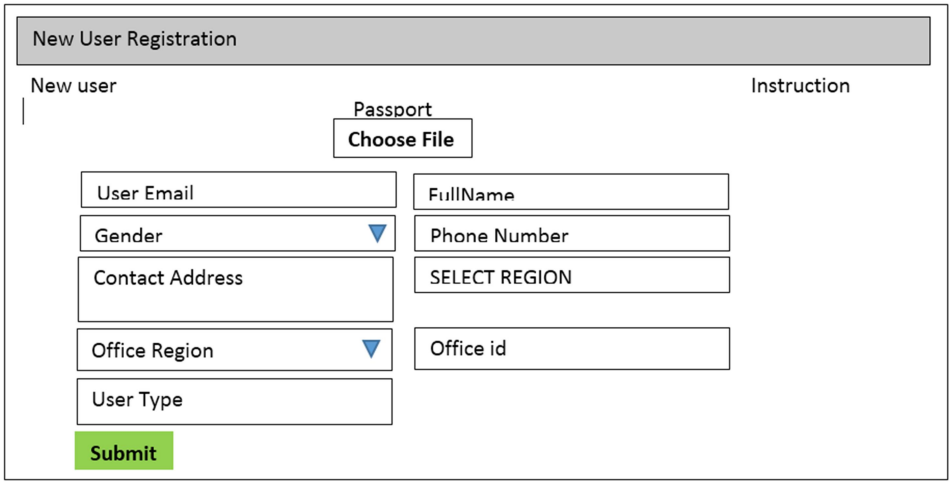


Figure 4.10: New User Registration Template

(b) Airtime Sales Registration: The templates captures the details of airtime sales made in the region and how to mine historical sales record as shown in figure 4.11.

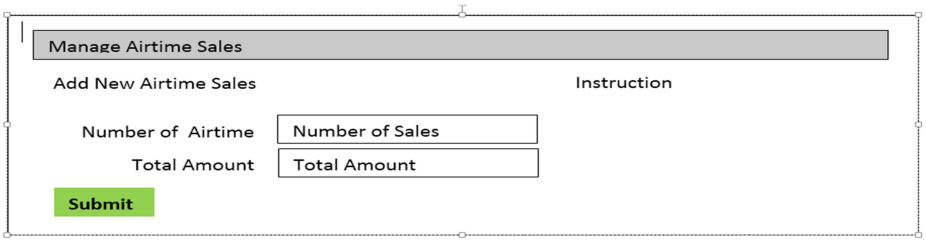


Figure 4.11: Airtime Sales Registration Template

Input: manage Airtime sales to mine: The templates captures details of airtime sales to be mine. This is shown in Figure 4:12.

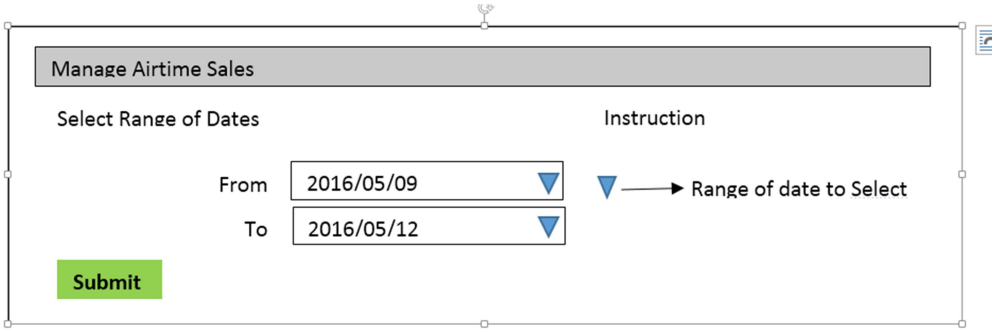


Figure 4.12: Mine out template for Airtime

(c) News E-top Details: The templates captures the details of e-top up sales made in the regions and how to mine historical e-top up sales records as shown in Figure 4.13.

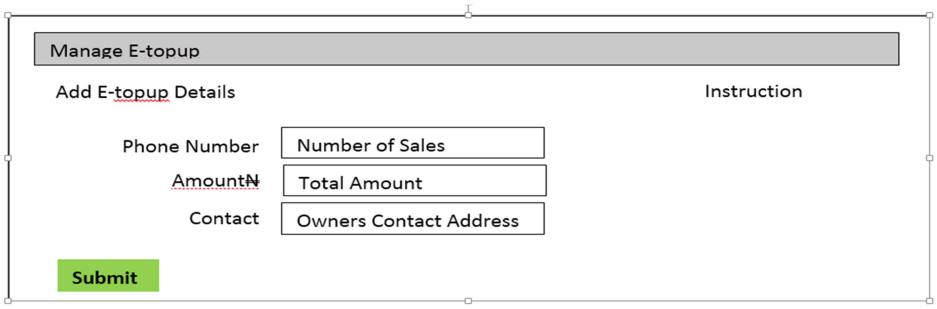


Figure 4.13: E-Top Up Registration sales Template

Input: manage E-topup to mine: The templates provides means in which the user can mine historical sales record of e-top up. This is shown in Figure 4:14.

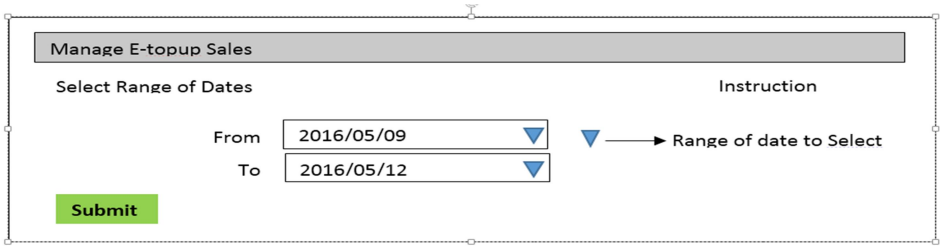


Figure 4.14: Mine out e-top up template

(d) New Sim Card Registration Template: This template captures details of New Sim Card to be purchase as shown in Figure 4.15.

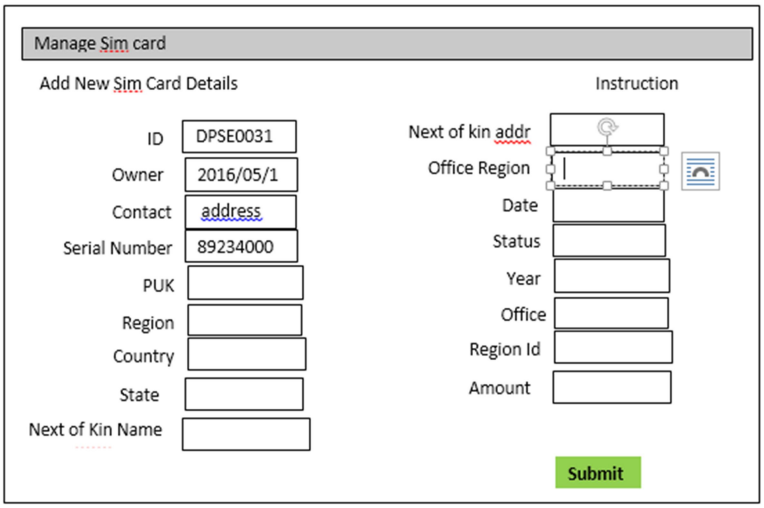


Figure 4.15: Create New Sim Card Template

Output Design

The system uses this module to convey information such as results, acknowledgement receipt to people. Thus, the output templates are presented as follows.

(a) Sample Output for New User Registration: Figure 4.16 display the output of the new user registration.

Sample Output User Registration

S/N	User name	User level	Status	Action
1	Emeka Francis	Director		
2	Njoku John	Director		

Print

Figure 4.16: Sample output for User Registration.

(b) Output Airtime Sales Mine out

The template displays Airtime sales mine out of a specific period. Figure 4.17 display the sample output within the selected date.

Airtime Sales from 2008 to 2010

Analysis Table

#	Region	Value Gotten	Target	Difference
1	Lagos North			
2	Lagos South			
3	South West			
4	South East			
5	South South			
6	North 1			
7	North 2			

Figure 4.17: Sample Output Airtime Sales Mine out

(c) Output E-top up Sales Mine out: The template displays the E-top up sales mine out of a specific period. Figure 4.18 display the sample output within the selected date.

E-top up Sales from 2008 to 2010

Analysis Table

#	Region	Value Gotten	Target	Difference
1	Lagos North			
2	Lagos South			
3	South West			
4	South East			
5	South South			
6	North 1			
7	North 2			

Figure 4.18: Sample Output E-top up Sales Mine out

(d) Output Simcard Sales Mine out: The template displays simcard sales mine out of a specific period. Figure 4.19 display the sample output within the selected date.

Simcard Sales from 2008 to 2010

Analysis Table

#	Region	Value Gotten	Target	Difference
1	Lagos North			
2	Lagos South			
3	South West			
4	South East			
5	South South			
6	North 1			
7	North 2			

Figure 4.19: Sample Output Simcard Sales Mine out

4.2.1.7 Database Design

The database designs illustrate the classification details of the data records used for the data mining analysis. This is done in order to establish the relationships within the entities making up the data set in the data warehouse and data marts. The tables below show the operations required to perform analysis using the data mining techniques.

1. Admin.dbfs

This database file structure captures and stores the admin/users login details with respect to their levels of permission. Table 4.8 presents it.

Table 4.8: Admin.dbf Structure

S/N	FIELD NAME	DATA TYPE	SIZE
1	Username	Varchar	122
2	Password	Varchar	122
3	Level	Varchar	122
4	Creator	Varchar	122

2. Etopup.dbf

This database file structure captures the details of all E-Top-Up sales. The information from here will be used to perform analysis from based on this aspect. This is presented in Table 4.9.

Table 4.9: Etopup.dbf Structure

S/N	FIELD NAME	DATA TYPE	SIZE
1	Phone	Varchar	19
2	Region	Varchar	122
3	State	Varchar	122
4	Address	Text	122
5	Amount	Varchar	122
6	Date	Varchar	30
7	Year	Varchar	30

8	Status	Int	12
9	Date_format	Varchar	12

3. AirtimeSales.dbf

The database file structure captures information about individual airtime sales made with respect to the id and region and Table 4.10 represent it.

Table 4.10: AirtimeSales.dbf Structures

S/N	FIELD NAME	DATA TYPE	SIZE
1	User_id	Varchar	122
2	Region	Varchar	122
3	Year	Varchar	122
4	Date	Varchar	122
5	Office_id	Varchar	122
6	Numofsales	Varchar	122
7	Amount	Varchar	122
8	Type	Varchar	122
9	Date_format	Varchar	122

4. Sim_card.dbf

The database file structure captures the details of individual sim card registered as purchased in the system and Table 4.11 represent the structure.

Table 4.11: Sim_card.dbf Structure

S/N	**FIELD NAME**	**DATA TYPE**	**SIZE**
1	Owner	Varchar	122
2	Contact	Varchar	122
3	Phone	Varchar	122
4	Serial	Varchar	122
5	Puk	Varchar	122
6	Region	Varchar	122
7	Country	Varchar	122
8	State	Varchar	122
9	Nkin	Varchar	122
10	Nkin_address	Varchar	122
11	Nkin_phone	Varchar	122
12	Office_region	Varchar	122
13	Date	Varchar	122
14	Year	Varchar	122

5. Target.dbf

The database file captures shows the details of individual targets with respect to the product.

Table 4.12: Target.dbf

S/N	FIELD NAME	DATA TYPE	SIZE
1.	Year	Varchar	122
2.	Region	Varchar	122
3.	Status	Int	12
4.	Amount	Varchar	122
5.	Date	Varchar	122
6.	Id	Varchar	122
7.	Creator	Varchar	122
8.	Date_format	Varchar	122
9.	Item	Varchar	122

6. Users.dbf

The database file structure captures the full details of individual users as captured by the system. The illustration is shown in Table 4.13.

Table 4.13: Users.dbf Structure

S/N	FIELDNAME	DATATYPE	SIZE
1.	Name	Varchar	122
2.	Gender	Varchar	122
3.	State	Varchar	122
4.	Nationality	Varchar	122
5.	Contact	Varchar	122
6.	Phone	Varchar	122
7.	Email	Varchar	122
8.	Region	Varchar	122
9.	Status	Int	12
10.	Picture	Varchar	122
11.	Office_address	Varchar	122

4.2.1.8 Entity Relationship Diagram

In software engineering, an entity–relationship model (ER model) is a data model for describing the data or information aspects of a business domain or its process requirements, in an abstract way that lends itself to ultimately being implemented in a database such as a relational database. The main components of ER models are entities (things) and the relationships that can exist among them. However, entities

are the Etop-up, Airtime Sales, Sim_card and the admin (Rouse, 2000). Thus, this is illustrated in Figure 4.20

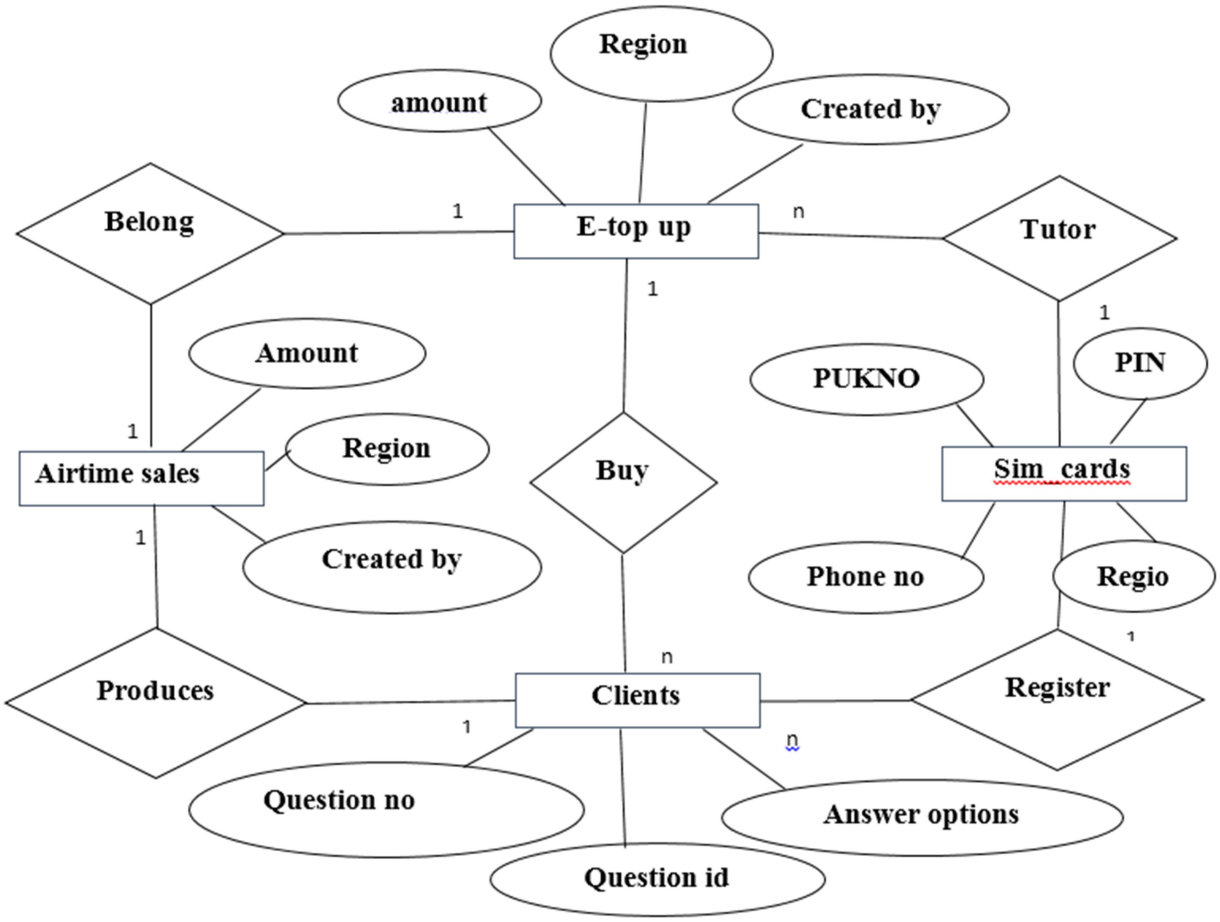


Figure 4.20: Entity-Relationship Diagram of the System

4.2.1.9 System Algorithm, Flowchart and Data Mining Method

Algorithm design is of utmost importance in software development; it simplifies the job of a programmer; inform him of his next step in the conception of a program and guide him towards the realization of the entire program.

(1) System Algorithm

An algorithm is a logical sequence of discrete steps used to solve a problem in a finite amount of time. It is therefore the breakdown of a particular task into several specific sub-tasks. It's important to a system analyst and a programmer is to understand the concept of algorithm in details so as to be able to come up with simple and efficient systems (Rouse, 2014).

(a) Simcard Sales Registration Algorithm

The algorithm controls the Sim Card registration processes.

```
start
        Open sim_card (E) for Output
        Open users.dbf(F) for input
           Powner = space(122)
                        Pcontact = space(122)
                        Pphone = space(122)
                        Pserial = space(122)
                        PPUK = space(122)
                        Pregion = space(122)
                              Pstate = space(122)
                              Pcountry = space(122)
                              Pnkin = space(122);
                              Popen_address = space
                              Pnkin_phone = space(122)
                              Poffice_region = space(122)
                              Pdate = space(122)
                              Pyear = space(122)

                        Gotoxy(8,5);Read Powner
                        Gotoxy(8,10);Read Pserial
                        Gotoxy(11,5);Read Pphone
```

```
                    Gotoxy(11,10);Read PPUK
                    Gotoxy(13,5;Read Pcontact
                    Gotoxy(13,10);Read Pnkin
                    Gotoxy(15,5);read Pnkin_phone
                      Append Blank record into {E}
                          Write('Is InputOk(Y/N)?');Read Response
                          if Response = 'Y' then
                                begin
                                Replace E.Owner with Powner
                                Replace E.Serial with Pserial
                                Replace E.Phone with Pphone
                                Replace E.Puk with PPUK
                                Replace E.Contact  with Pcontact
                                Replace E.Nkin with Pnkin
                                Replace E.Nkin_phone  with Pnkin_phone
                                Replace E.Year with Year
                                Replace  E.date with CURRENT_DATE
                                Replace E.Office_region with F.Office_region
                                EndProgram = 1
                                End do
            Close {E}
            Close {F}

End
```

(b) E-top up Sales Algorithm: This algorithm controls the E-top registration

```
start
        Open e-topup (E) for Output
        Open users.dbf(F) for input
          Pnumberof Sales = space(122)
                     Pitem = space(122)
                    PAmount = space(122)
                   Gotoxy(8,5);Read Pnumberof Sales
                    Gotoxy(8,10);Read Pitem
                    Gotoxy(11,5);Read PAmount

                      Append Blank record into {E}
                          Write('Is InputOk(Y/N)?');Read Response
                          if Response = 'Y' then
```

```
                                    begin
                                    Replace E.User_id with F.username
                                    Replace E.region with F.region
                                    Replace E.Year with CURRENT_YEAR
                                    Replace E.date with CURENT_DATE
                                    Replace E.Office_id  with E.Office_id
                                    Replace E.numofsales with Pnumofsales
                                    Replace E.amount  with Pamount
                                    Replace E.type with PItem
                                    EndProgram = 1
                                    End do
            Close {E}
            Close {F}

End
```

(c) Airtime sales Algorithm: This algorithm controls Airtime sales registration.

```
start
        Open airtime.dbf (E) for Output
        Open users.dbf(F) for input
          Pphonenumber Sales = space(122)
                        Pamount = space(122)
                       Pcontact = space(122)
                      Gotoxy(8,5);Read Pphonenumber Sales
                       Gotoxy(8,10);Read Pamount
                       Gotoxy(11,5);Read Pcontact

                          Append Blank record into {E}
                              Write('Is InputOk(Y/N)?');Read Response
                              if Response = 'Y' then
                                    begin
                                    Replace E.contact with F.contact
                                    Replace E.region with F.region
                                    Replace E.Year with CURRENT_YEAR
                                    Replace E.date with CURENT_DATE
                                    Replace E.Office_id  with F.Office_id
                                    Replace E.numofsales with Pnumofsales
                                    Replace E.phonenumber  with Pphonenumber
```

```
                    Replace E.contact with Pcontact
                    EndProgram = 1
                    End do
        Close {E}
        Close {F}

End
```

(d) Mine out Simcard, Airtime and Etop up Algorithm: The algorithm for mining Simcard, Airtime, E-top up are all the same and follows the classical software engineering algorithm. This is presented as follows:

```
Start
    Open airtime.dbf{F} for input
    Open simcard.dbf{E} for input
    Open etopup.dbf{H} for input
        Pentity = space(200)
        Pdate1 = space(date)
        Pdate2 = space(date2)
        Gotoxy(8,10);Read Pentity
        Gotoxy(10,10);Read Pdate1
        Gotoxy(10,18); Read Pdate2
        Recognise entity as follows

        if Pentity = simcard
          {M} = {E}
        end if
        if Pentity = etopup
          {M} = {H}
        end if

        if Pentity = airtime
          {M} = {F}
        end if

        search {M} Where date >= Pdate1 and date <= Pdate2
```

```
            if result found display result
            else
            display 'No result found'

            endif
        End {H}
        End {E}
        End {F}

Continue
End
```

(e) Login Algorithm for the proposed system: the algorithm that leads to

development of login process in the proposed system is as follows:

```
Start
        Open admin.dbf{A} for output
        PUsername = space(122)
                        PPassword = space(122)
                              Gotoxy(8,5);Read PUsername Sales
                        Gotoxy(8,10);Read PPassword

                        Search A where A.Password = PPassword and A.Username =
PUsername
                        if resultOk
                         Continue else
                         Start Over again
        Close A
End
```

4.2.1.10 Program Flowchart

Program flowchart is the diagrammatic representation of how the entire system works (BBC, 2014). However, the flowchart will be divide into two admin-end flowchart and students-end flowchart (BBC, 2014). The system flowcharts of the model is presented in Figure 4.21

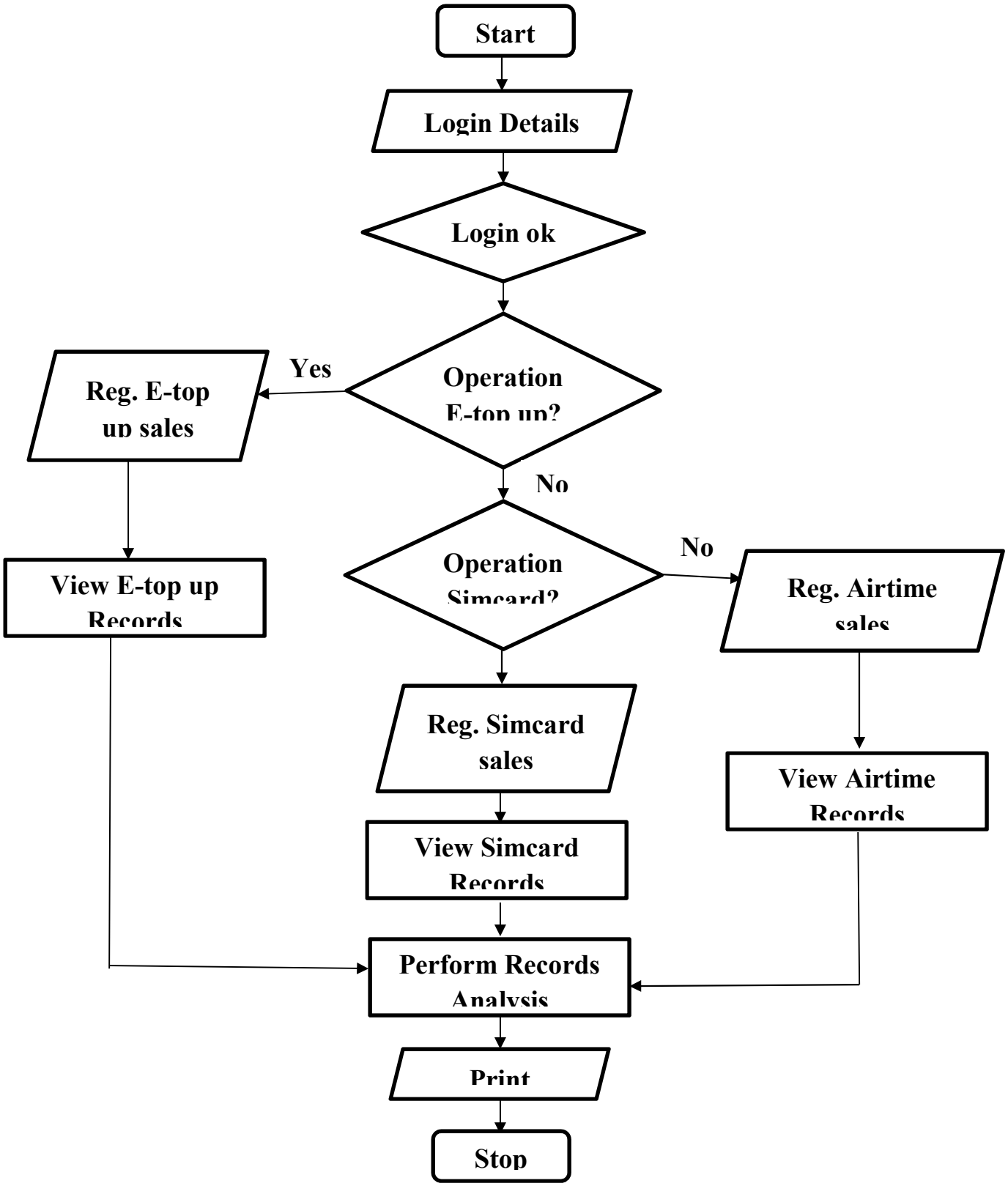


Figure 4.21: Flowchart of the program

Login Flowcharts: In order to launch the proposed model the flowcharts display the steps involved in the development of the login process:

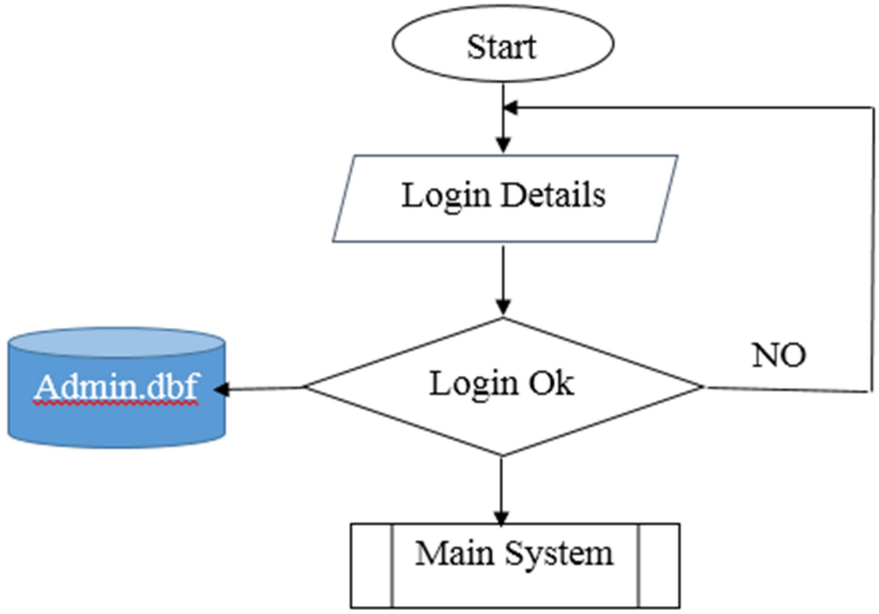


Figure 4.22: Flowchart for login Process

Adding of New sim card sales: The flowchart display the steps taken to achieve the development of this module. This is represented in Figure 4:23.

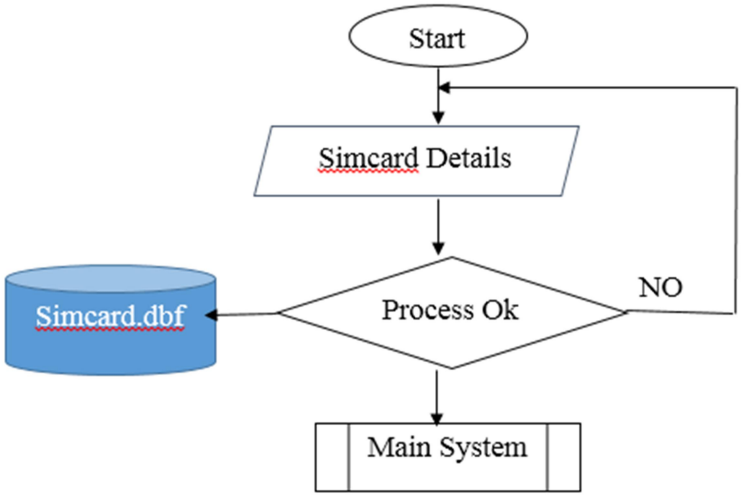


Figure 4:23: Flowchart for adding New Sim card sales.

Adding of New E-top up Sales: The flowchart display the steps taken to achieve the development of this module. This is represented in Figure 4:24.

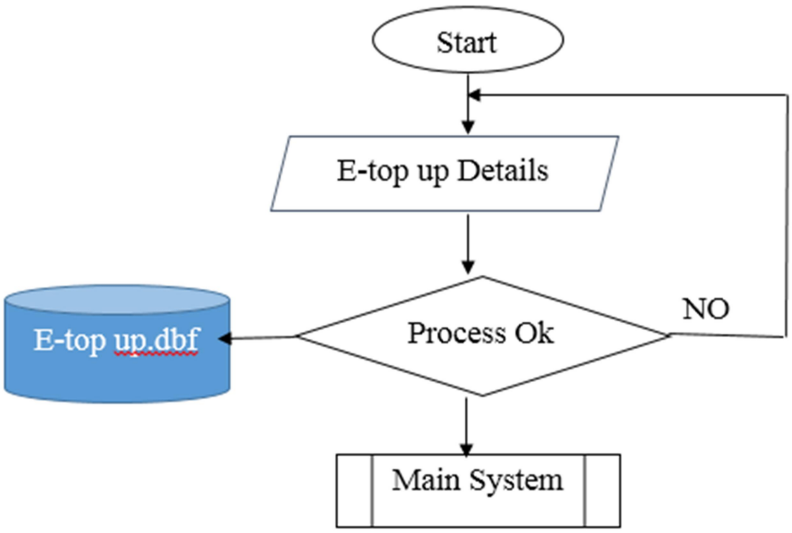


Figure 4:24: Flowchart for adding New E-top up sales.

Adding of New Airtime Sales: the flowchart display the steps taken to achieve the development of this module. This is represented in figure 4:25.

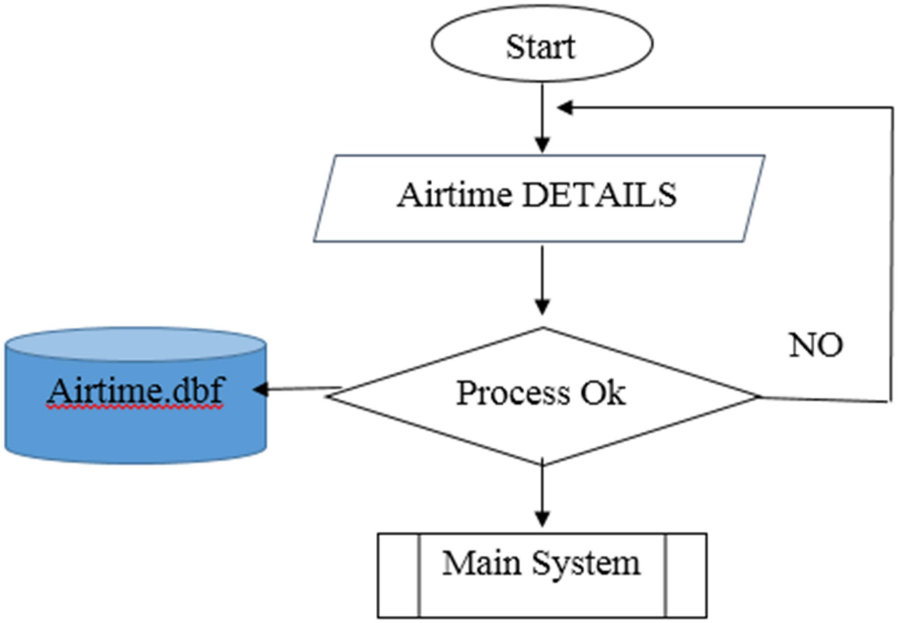


Figure 4: 25: Flowchart for adding Airtime sales.

Mine out Simcard, Airtime and E-top up: the flowchart display the steps taken to achieve the development of these modules. This is represented in figure 4:26.

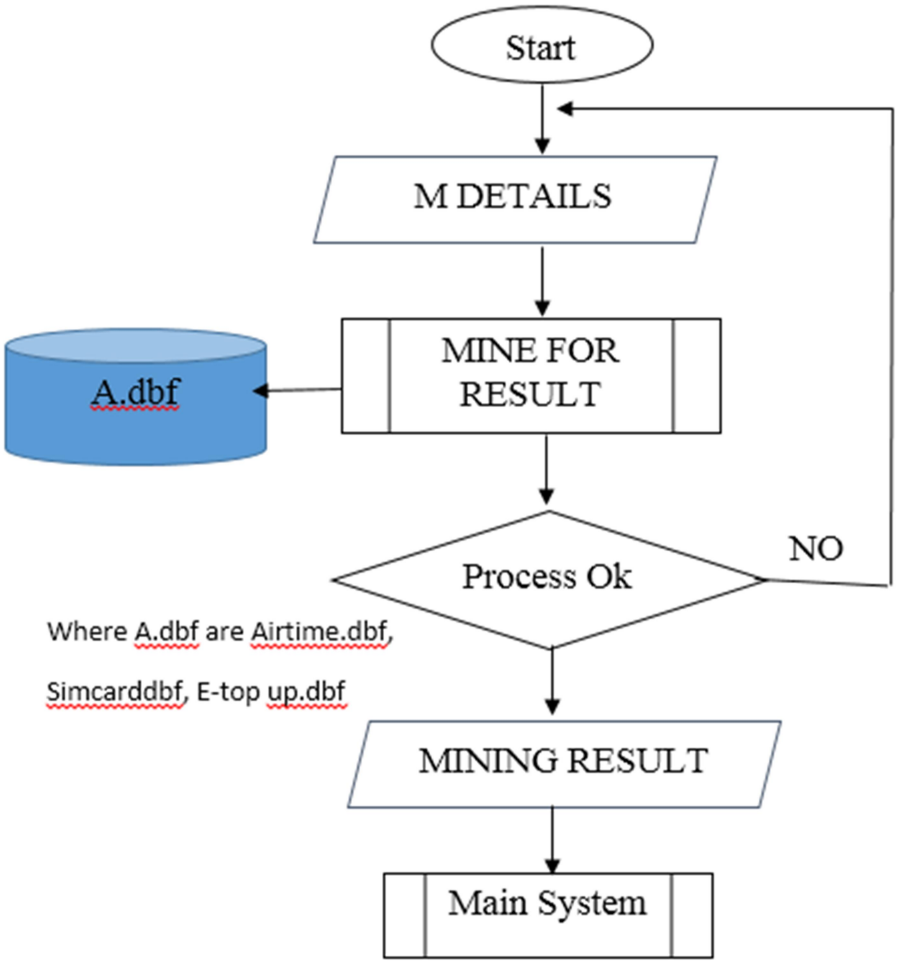


Figure 4:26: Flowchart for Mine Out

4.2.2 Physical System Design

The screenshot of the input/output design, database logical designs and exploratory analysis carried out with the new model were presented in this section.

4.2.2.1 Input Design screenshots.

Users: Create User: The template snapshot illustrate the registration of new users. It is shown in Figure 4.27.

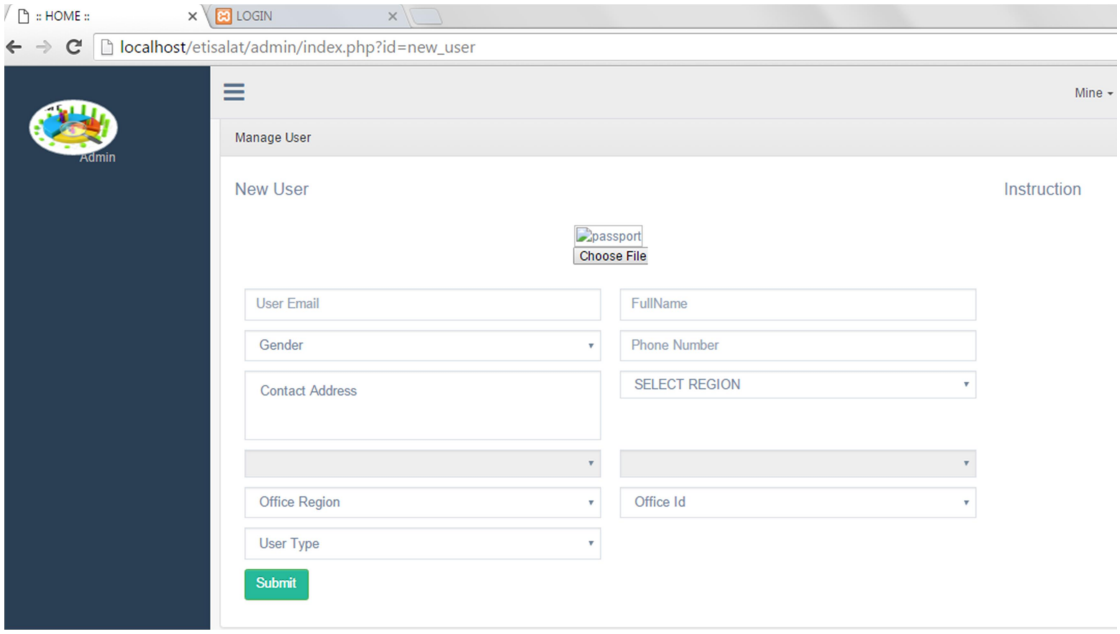


Figure 4.27: Snapshot for Registration of new users

New collation: This is made up of the following items

a) Add Airtime Sales: This is used to add Airtime sales realized from the business. Figure 4.28 is the screenshot.

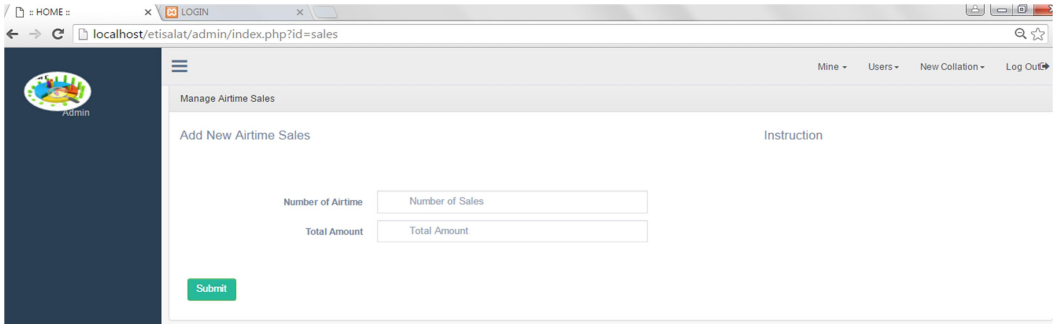


Figure 4.28: Screenshot for Adding New Airtime Sales

b) Mine Sales: This aspect is used to analyze airtime sales realized as shown in the Figure 4.29.

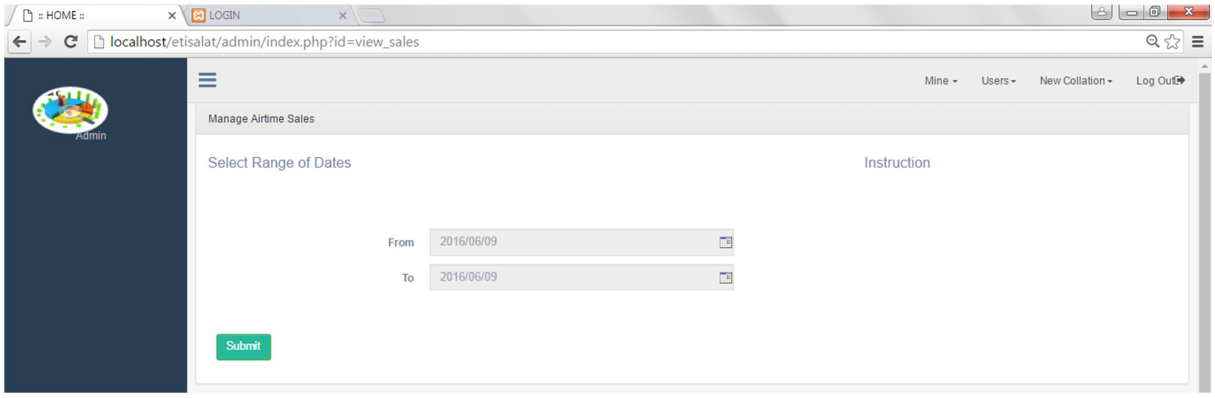


Figure 4.29: Snapshot for Managing and Analyzing Airtime Sales

c) New E-top up: Figure 4.30 and 4.31 show snapshot for adding new E-top up and analyzing E-top up records being added respectively.

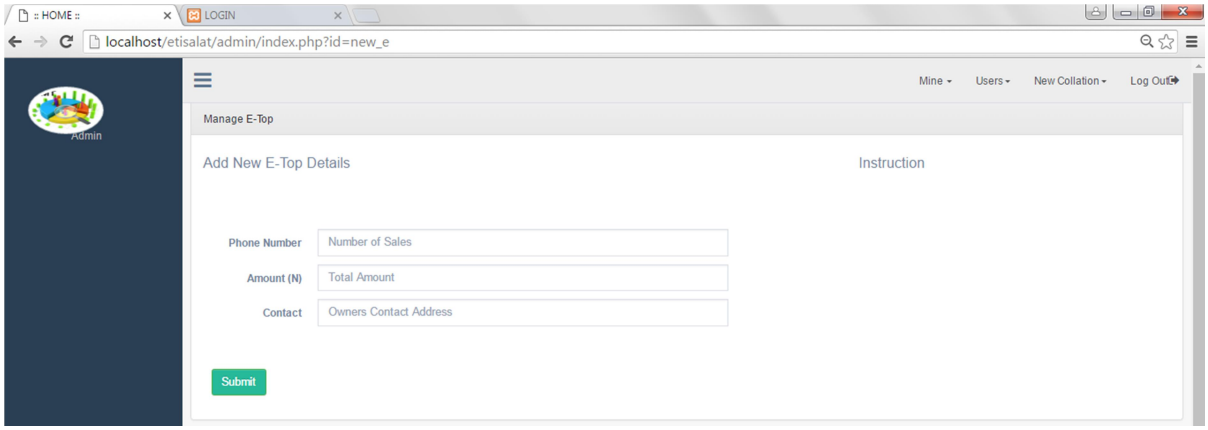


Figure 4.30: Snapshot for Adding New E-top up Sales

d) Mine E-top up: the template captures details of specification of records needed to mine out.

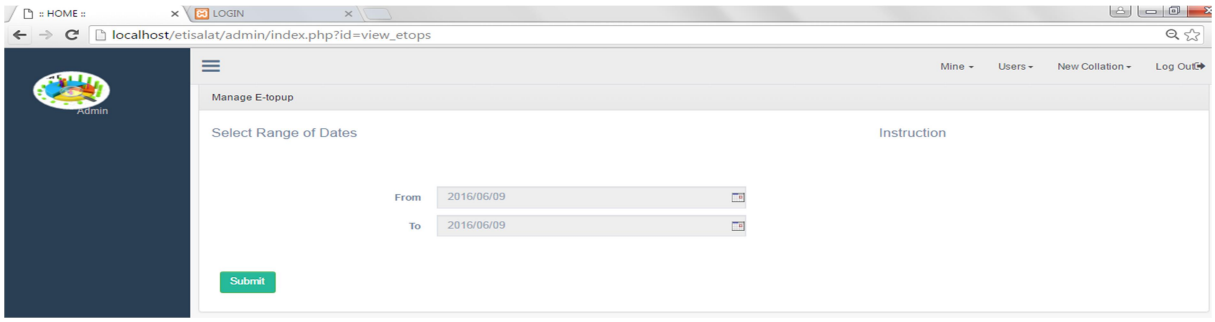


Figure 4.31: Snapshot for Mining E-top up Sales

e) New Sim card: The new sim card registered and amount realized are also important and should be registered. Figure 4.32 and 4.33 shows the screenshot for adding/registering new Sim card and Mining the sim card records being added respectively.

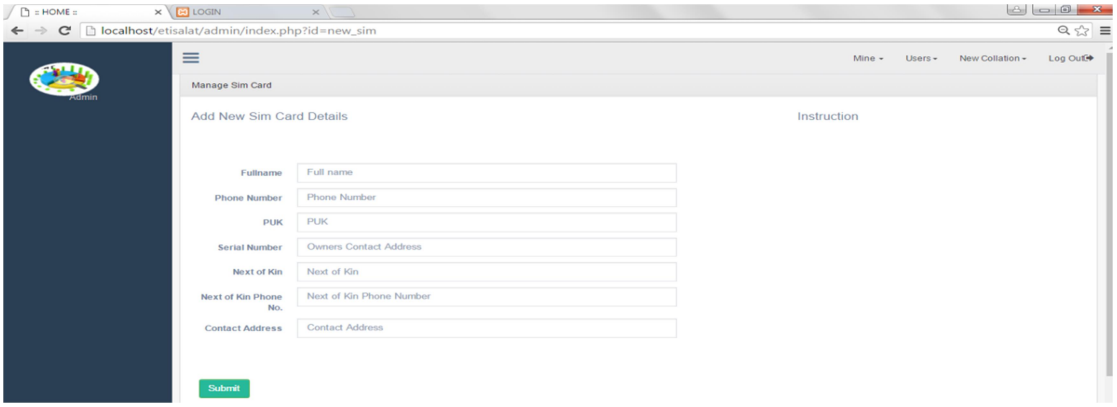


Figure 4.32: Snapshot for Adding/Registration of New Sim Card Sales

f) Mine out Sim Card: the templates captures details of sales on sim card sales record needed to mine out.

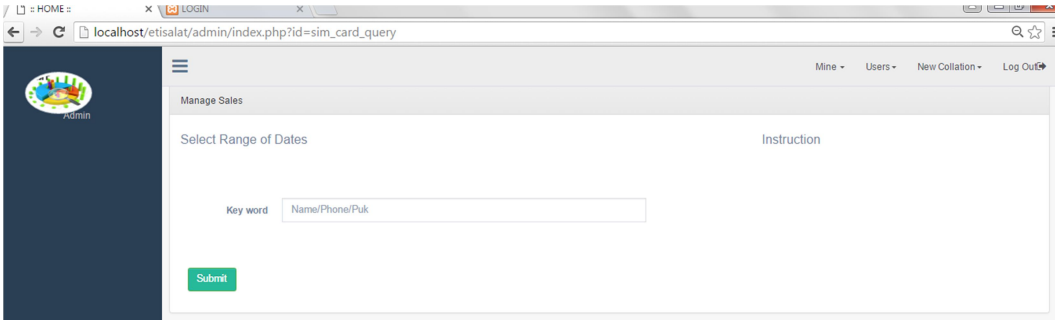


Figure 4.33: Snapshot for Mining registered Sim Card Sold.

4.2.2.2 Output design Screenshot.

The sample of the output displayed here is the screenshot of the mine out. The airtime sales report mine out was the sample used. The Figure 4.34 is the display.

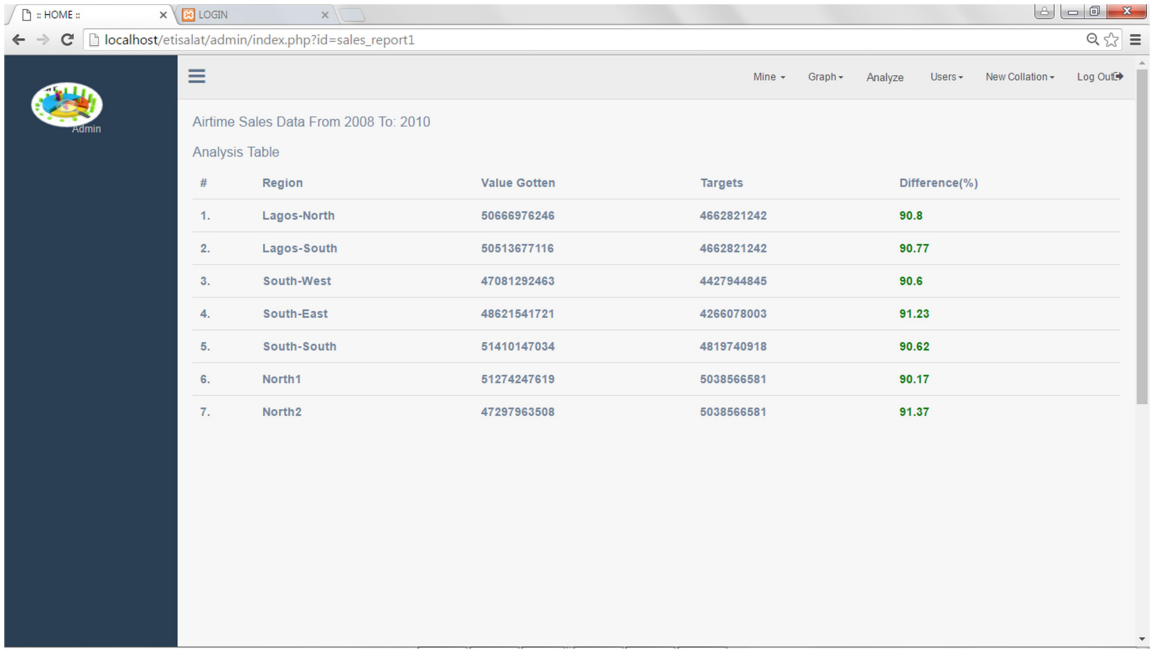


Figure 4.34: Mine out result of Airtime sales.

4.2.2.3 Database Logical designs Screenshots.

The database structure screenshots of the components identified in the input/output design is displayed as follows:

Admin Structure: This contains the database definitions of the variables used in programing and development of Admin section.

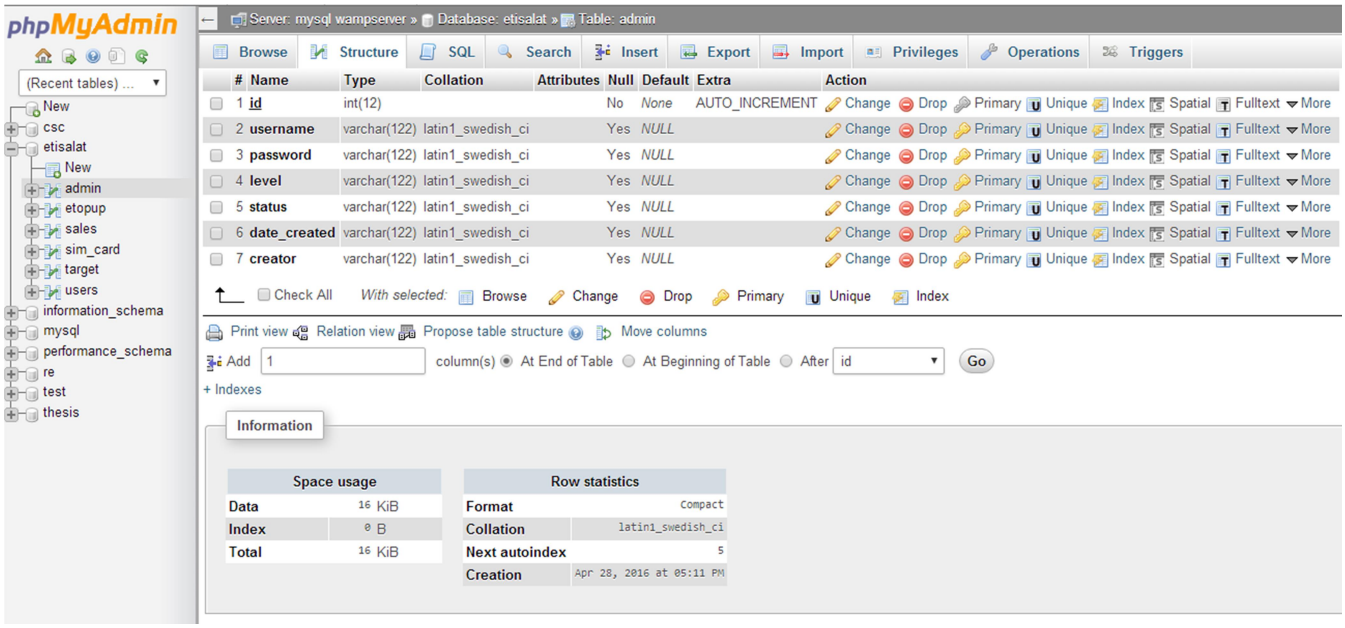


Figure 4.35: Admin Structure

E-top Up Structure: This contains the database definitions of the variables used in programing and development of e-top up sales record section.

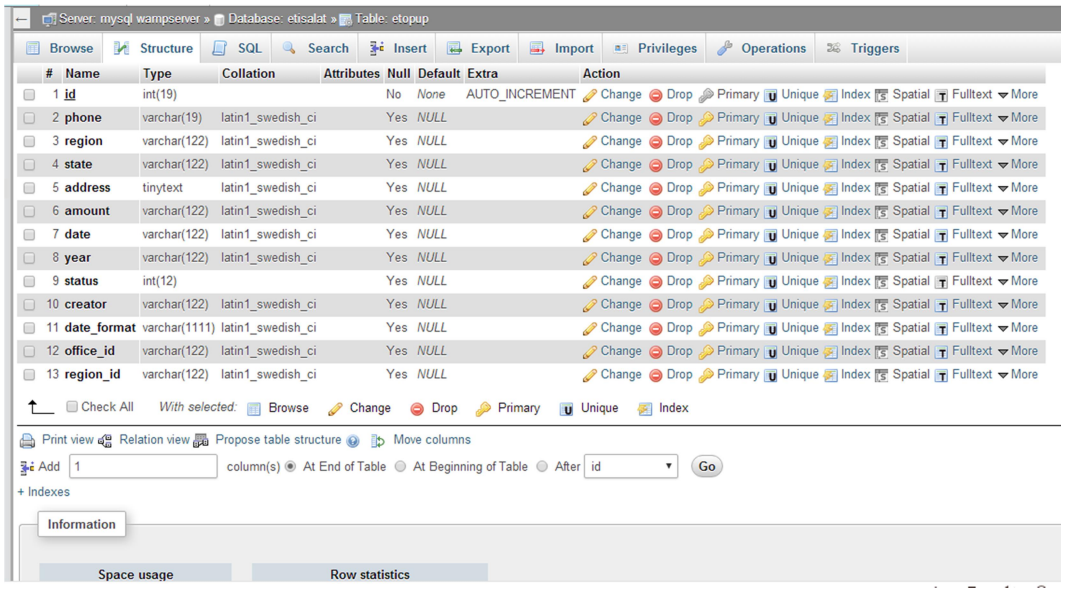


Figure 4.36: E-top Up Structure

Airtime Structure: This contains the database definitions of the variables used in programing and development of Airtime sales record section.

Server: mysql wampserver » Database: etisalat » Table: sales

Browse | Structure | SQL | Search | Insert | Export | Import | Privileges | Operations | Triggers

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
1	id	int(12)			No	None	AUTO_INCREMENT	Change Drop Primary Unique Index Spatial Fulltext More
2	user_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
3	region	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
4	year	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
5	date	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
6	office_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
7	numofsales	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
8	amount	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
9	status	int(12)			Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
10	date_format	varchar(1222)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
11	type	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
12	region_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More

Check All With selected: Browse Change Drop Primary Unique Index

Print view Relation view Propose table structure Move columns

Add 1 column(s) At End of Table At Beginning of Table After id Go

+ Indexes

Information

Space usage		Row statistics	
Data	16 KiB	Format	Compact

Figure 4.37: Airtime structure

Sim card Structure: This contains the database definitions of the variables used in programing and development of Sim card sales record section.

Server: mysql wampserver » Database: etisalat » Table: sim_card

Browse | Structure | SQL | Search | Insert | Export | Import | Privileges | Operations | Triggers

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
1	id	int(12)			No	None	AUTO_INCREMENT	Change Drop Primary Unique Index Spatial Fulltext More
2	owner	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
3	contact	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
4	phone	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
5	serial	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
6	puk	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
7	region	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
8	country	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
9	state	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
10	nkin	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
11	nkin_address	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
12	nkin_phone	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
13	office_region	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
14	date	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
15	status	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
16	year	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
17	office_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
18	region_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More
19	amount	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext More

Check All With selected: Browse Change Drop Primary Unique Index

Figure 4.38: Sim card Structure

User Structure: this contains the database definitions of the variables used in programing and development of Users section.

Server: mysql wampserver » Database: etisalat » Table: users

Browse | Structure | SQL | Search | Insert | Export | Import | Privileges | Operations | Triggers

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
1	id	int(12)			No	None	AUTO_INCREMENT	Change Drop Primary Unique Index Spatial Fulltext Distinct values
2	name	varchar(1111)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
3	gender	varchar(16)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
4	state	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
5	nationality	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
6	contact	varchar(1111)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
7	phone	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
8	email	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
9	region	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
10	status	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
11	picture	tinytext	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
12	office_address	varchar(1111)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values
13	office_id	varchar(122)	latin1_swedish_ci		Yes	NULL		Change Drop Primary Unique Index Spatial Fulltext Distinct values

Check All With selected: Browse Change Drop Primary Unique Index

Print view Relation view Propose table structure Move columns

Add 1 column(s) At End of Table At Beginning of Table After id Go

+ Indexes

Information

Space usage Row statistics

Figure 4.39: User Structure

4.2.2.4 Exploratory analysis on data mining model.

The outline are few illustration analysis that can be done with the proposed data mining model.

Mine: This consist of

a) **Mine Sales by components**: When clicked by selecting any components (either, airtime, E-top up, or sim card), selecting range in years such as (2008 – 2009, 2008 – 2010, 2009 – 2012, 2008-2015 etc) will display the records. When clicked on Graph (any of pie chart, bar chart, or line graph), the display

will be shown which is easier to interpret. Report displayed will be analyzed and prediction would be made.

The Figure 4.34 in subsection 4.2.2.2 shows the snapshot for mining records of Airtime (showing region of sales, value gotten and target given) from 2008 to 2010. The Figures 4.40 and 4.41 shows the interpretation using bar chart and analysis report (the airtime prediction for 2011) respectively.

Figure 4.40: Bar Chart for Figure 4.34

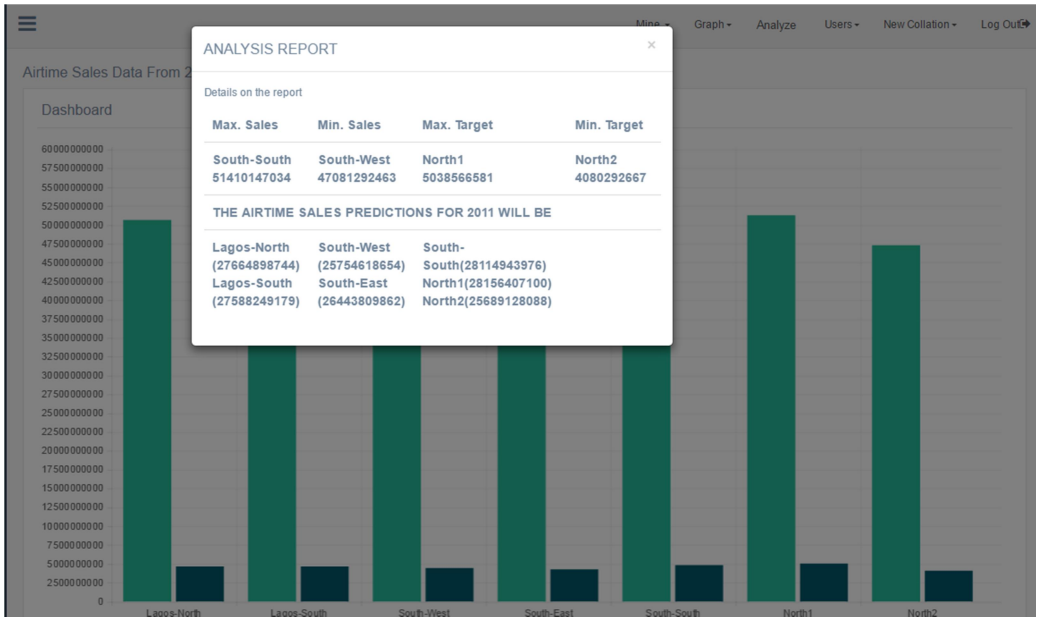
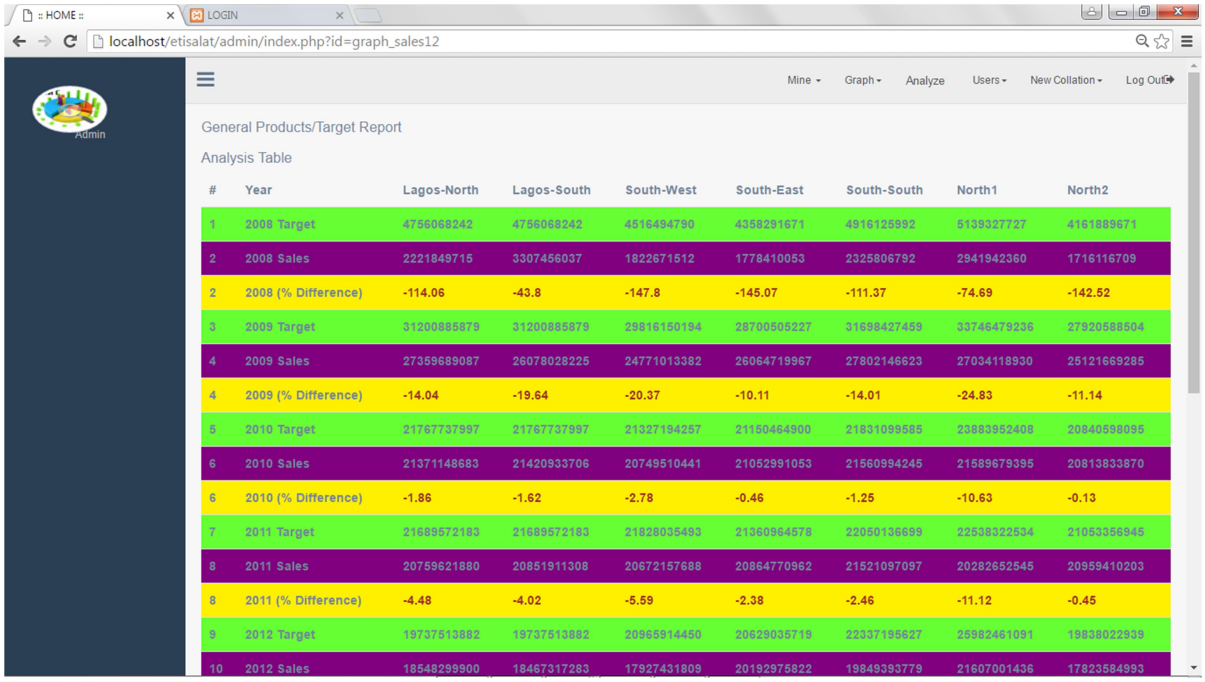


Figure 4.41: Analysis result

b) **All Product Data Target**: The figure 4.42 shows all the products target, sales made and difference – determines whether is actualized or not with respect to years of business undertaken. Figure 4.43 shows the Chart that interprets Figure 4.42



#	Year	Lagos-North	Lagos-South	South-West	South-East	South-South	North1	North2
1	2008 Target	4756068242	4756068242	4516494790	4358291671	4916125992	5139327727	4161889671
2	2008 Sales	2221849715	3307456037	1822671512	1778410053	2325806792	2941942360	1716116709
2	2008 (% Difference)	-114.06	-43.8	-147.8	-145.07	-111.37	-74.69	-142.52
3	2009 Target	31200885879	31200885879	29816150194	28700505227	31698427459	33746479236	27920588504
4	2009 Sales	27359689087	26078028225	24771013382	26064719967	27802146623	27034118930	25121669285
4	2009 (% Difference)	-14.04	-19.64	-20.37	-10.11	-14.01	-24.83	-11.14
5	2010 Target	21767737997	21767737997	21327194257	21150464900	21831099585	23883952408	20840598095
6	2010 Sales	21371148683	21420933706	20749510441	21052991053	21560994245	21589679395	20813833870
6	2010 (% Difference)	-1.86	-1.62	-2.78	-0.46	-1.25	-10.63	-0.13
7	2011 Target	21689572183	21689572183	21828035493	21360964578	22050136699	22538322534	21053356945
8	2011 Sales	20759621880	20851911308	20672157688	20864770962	21521097097	20282652545	20959410203
8	2011 (% Difference)	-4.48	-4.02	-5.59	-2.38	-2.46	-11.12	-0.45
9	2012 Target	19737513882	19737513882	20965914450	20629035719	22337195627	25982461091	19838022939
10	2012 Sales	18548299900	18467317283	17927431809	20192975822	19849393779	21607001436	17823584993

Figure 4.42: General Product/Target Report

The general product/target report can also be display on a bar chart. This is demonstratet on Figure 4.43.

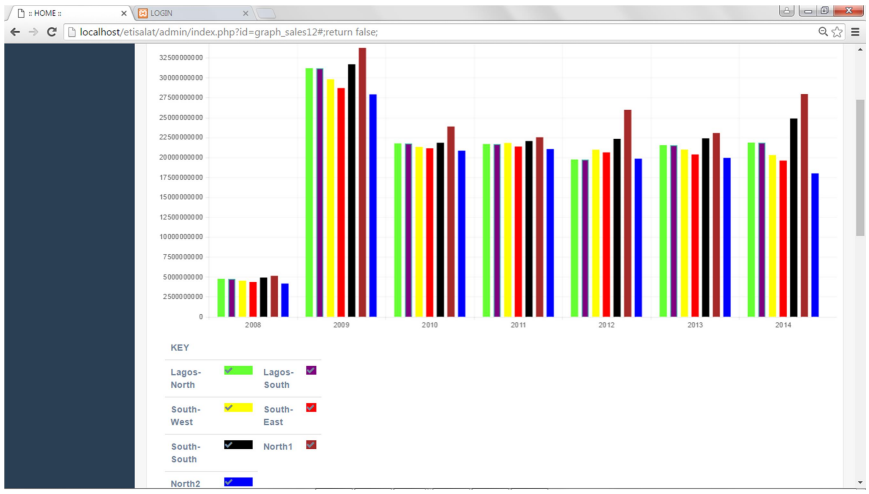


Figure 4.43: Bar chart illustrating figure 4.42

The Figure 4.44 shows the analysis report for figure 4.42 and 4.43 with respect to regions. From this analysis, it can be deduced that: (1) maximum sales was made from Lagos-North in 2009 at the amount of ₦27359689087.00, (2) Minimum sales was made in North2 in 2008 at the amount of ₦1716116709.00, (3) the Maximum and Minimum target can be seen in the analysis. The organization can make a prediction that in the next following years, according to the sales trends Lagos North sales will increase.

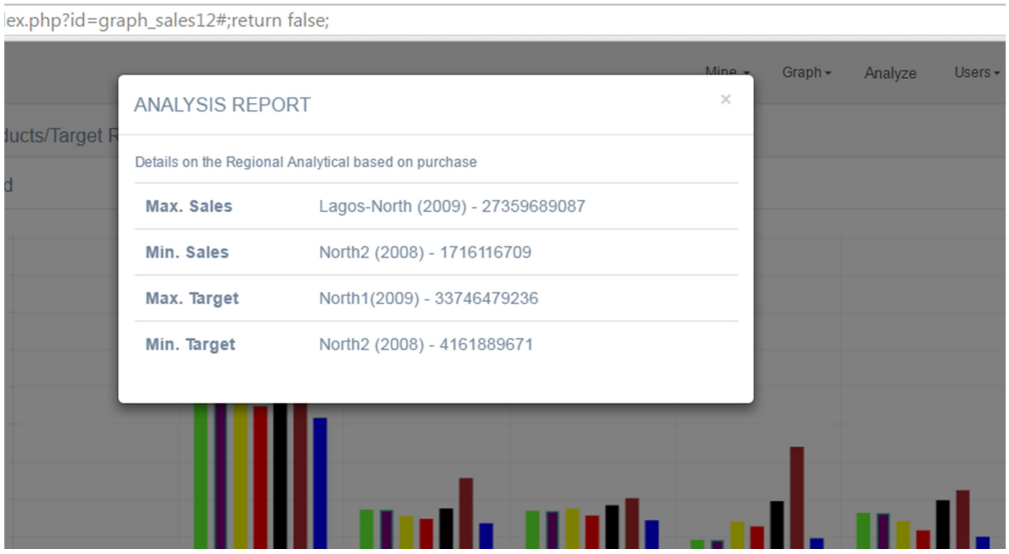


Figure 4.44: Analysis Report for figure 4.42 and 4.43.

4.3 System Requirements

The proposed system is a typical web-based application. Therefore, the basic system requirement consists of hardware, software and database component requirements. These are explained in section (4.3.1, 4.3.2, and 4.3.3) respectively.

4.3.1 Hardware Requirements

The new system is designed to be browse over the network with the following minimum hardware specification requirements:

1. Pentium II series processor
2. RAM (Random Access Memory) 312 MB(Mega Byte)
3. CD-ROM drive, 48x(speed)
4. 10 GB(Giga Byte)
5. 56kbps full duplex fax modem
6. Any external drive for backup
7. 15'' colour monitor, preferably a flat screen monitor for space management and to avoid excess radiation on the eye of the user.

4.3.2 Software Requirements

The new system is designed to be implemented with the following minimum software requirements:

1. Operating System: any version of Microsoft Windows operating system (windows NT/2000/ME/XP/Vista/7/8), any version of Linux, and any version of Macintosh will enable one to run this system comfortably.

2. Anti-virus/Anti-spyware Software: Any good anti-virus software consisting of Norton's, Avira, among others, that will protect the system against malicious attacks is useful.

3. System Browser Software: Internet Explorer (any version)/ Mozilla Firefox browser/Opera browser/Safari browser or any other system browser.

4.3.3 Database Requirements

The database requirements needed to implement the new system consist of the following:

1. Apache WAMP Server (which comprises PHP server, MySQL Server and local server), version of WAMP and PHP xampp 3.2.1, php myadmin 5.5.6 Database engine inno db mysql version 5.6.14, apache 2.4.7
2. Adobe Dreamweaver CS6 or any PHP IDE.

4.4 Implementation and Testing

4.4.1 Implementation

The proposed system choose of programming is PHP, it is well specified in the software requirement under database specification in the (section 4.3.2).

Coding: This is the actual programming for implementing the designs in this work. It consist of the major modules such as:

1. Database Module: This is to establish the connection between the front end and back end of the system development.
2. CSS module: These codes is used for interface layout design and arrangement.
3. Java script module: This is used for validation and events automation that links the functional programs.
4. Image module: This contents the images or pictures that displays on the system interface.
5. Login and logout module: It concerns with designing the login and logout interfaces.
6. Index page: These codes enables is used to realize the homepage of the system.
7. Program Listing: Program listing of the modules 1 to 6 above and other files can be seen at the appendix

4.4.2 Testing

Testing is an important part of every software development. This is done to verify if the system achieves the goal set for it. It involves the execution of a software component or system to evaluate one or more properties of interest (Eleven40 pro theme, 2016). There are various phases of testing done in this proposed system. However, the tests performed include (Eleven40 pro theme, 2016):

1. **Unit Testing**

This testing method checks individual units of source code for errors or repeated entities. It also used to test sets of one or more program modules together with associated control data, usage procedures, and operating procedures to determine if they are fit for use Kolawa (2007).

2. **Integration Testing**

Integration testing is a software testing in which individual software modules are combined and tested as a group Ould and Unwin (1986). It is usually done after unit testing and before validation testing. Integration testing takes modules that have been unit tested, groups them in larger aggregates and applies tests defined in an integration test plan to those aggregates. This serves as its input. The output is the integrated system ready for system testing. The purpose of integration testing is to detect any inconsistencies between the software units that are integrated together (called assemblages) or between any of the assemblages and the hardware.

3. **System Testing**

System Testing is testing conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. System testing is a more limited type of testing; it seeks to detect defects both within the "inter – assemblages" and also within the system as a whole IEEE Standard Computer Dictionary (1990). Here, the system testing was carried out and also achieved.

4.4.2.1 Assessment / Measurement of the New System.

The identified weak performance indicators are the major indicators considered in evaluation of the new system.

Evaluation/ Improvement.

The new system was evaluate and assess by Information professionals and Information Consumers. A third questionnaire was developed in respect to it. The control matrix of result from the evaluation is presented in Table 4.14.

Table 4.14: Control matrix for evaluation of the new system

Performance indicator	**Extremely low**	**Low**	**Neutral**	**Very high**	**High**	**Rank**	**Judgments**
Likert scale grade	**1**	**2**	**3**	**4**	**5**		
Presentation quality	0	0	0	2	25	**5**	**STRONG**
Accessibility	0	0	0	6	21	**5**	**STRONG**
Easy to use	0	0	0	2	25	**5**	**STRONG**
Precise	0	0	0	2	25	**5**	**STRONG**
Robust	0	0	0	1	26	**5**	**STRONG**
Grade in minutes	**1hr/more**	**45mis**	**30mins**	**20min less**	**5mins less**	**rank**	
Speed (response in time)	0	0	0	0	27	**5**	**STRONG**

In evaluation, the new system prove to have overcome the weak performance indicator of the existing system.

Benefits of the New System

1. The model presented in this project work is very robust, flexible, modular and easily understandable model and have a high accuracy given a large volume of record thereby making the tedious, depth data mining analysis a simple and quick task.
2. The quality of the data and data preparation issues, particularly relating to financial databases was well addressed and qualities of the rules or knowledge discovered are not cumbersome and the relationships obtained are not too complex to understand.
3. A user friendly system made for analyzing data, thus offers a great range of graphs, techniques and charts for easy description of relationships in data and knowledge acquisition.
4. The model gives a privileged to gain quick precise view and predictive insights into the organizational large and complex datasets and reveal relationships and trends hidden in the geospatial data.
5. The knowledge acquisition from this model provides a guideline for consultant in advising and directing an organization in their key business performance.

4.5 Result Discussion

The result are discussed here within the context of research questions.

1) **What are the key performance indicators for evaluation of data mining model?**

From the literature review as well as the results of analyses carried out, several performance indicators were identified for evaluation of data mining model. The performance indicators identified in this study are as follows: accuracy, interpretability, understanding, presentation quality (visualization and understanding), accessibility, consistency, easy to use, precision, concise, robustness, speed (response in time), reliability, and unambiguous.

2) **To what extent does the existing model not measure up with required standard of performance indicators?**

From the analysis carried out on the existing model using Delphi methods Berry (1994), it shows that the existing model has some weakness in terms of accessibility, presentation quality, easy to use, robustness and speed (response in time)

3) **To what extent have an enhanced system been designed based on established tasks?**

The basic tasks were to develop a data warehouse and data marts. According to Shaikh (2013) developing a data warehouse start with identifying and collecting requirements, design the dimensional model and execute T-SQL queries to create

and populate your dimension and fact tables. In this study, the data/information were collected from the existing system. Data collected are eight years historical sales records of emerging telecommunication markets (EMTs). The data were cleaned and transformed. The Delphi method were used to evaluate and measured the existing system and collect information with regards to data analysis. The dimension which is a master table/ fact table composed of individual non-overlapping data elements was identified and this was used to design the database. The implementation was done with a query language. The logical design of data ware house is defined by dimensional data modelling approach (Akintola et al., 2011).

4) **To what extent have the designed enhanced system performed to measure up with required standard of performance indicators?**

Through the measurement and assessment using Delphi methods Berry (1994), the enhanced system proves to have overcome the bottlenecks of the existing system. The performance indicators identified as weak indicator in existing system was improved in the enhanced system. During evaluation the enhance system proves to have accommodate the strong and weak indicators of the existing system, making it to measure up to the required standard of the performance indicators.

5) **What policy recommendations can be made?**

The enhance system was designed to support decision making in telecommunication providers case study Etisalat. In system analysis and financial analysis, the system proves to be efficient and the net present value shows that the system will be profitable and yields good returns that will have positive impact in the organization. The enhance system is quite beneficial to be use.

CHAPTER FIVE

5.0 SUMMARY, CONCLUSIONS AND RECOMMDATION

5.1 Summary of Findings

In this project work, several literature works were reviewed in other to identify the key performance indicators that makes a data mining techniques model a powerful tool to be used for carrying out data analytic tasks that will achieve a good result to support decision making in telecommunication providers. The performance indicators identified in this study are accuracy, interpretability, presentation quality, accessibility, consistency, easy to use, precise, concise, robustness, speed (response in time), reliability and unambiguous. Furthermore, in the study the identified performance indicators were assessed and measured against existing system method use for data analytic in telecommunication providers such as MTN, GLO, Airtel and EMTs (etisalat). The findings shows that out of the twelve(12) identified performance indicators six (6) performance indicator are weak with respect to existing system. They are accessibility, robustness, easy to use, precise and speed (response in time).

The enhanced system was developed which proves to be stronger than the existing system. The model developed is able to make a sales forecast of the year 2016

performance whereas the training data used for model exploratory analysis range from 2008 to 2015.

A financial feasibility were carried out in respect to proposed system and the result of the financial analysis produce a positive Net present Value and outstanding Return of Investment (ROI).

KNOWLEDGE CONTRIBUTION

Previous research has demonstrated that Data mining technique is finding increasing applications in expertise orientation and the development of applications for data mining technique is a problem-oriented domain (Liao et al, 2012). This present work design a model for data mining techniques for carrying out analytic task in order to improve an existing system and identify the several performance indicators that a data mining model should be evaluated with to ascertain its efficiency. The performance indicator are also the major characteristics if an existing system do not possess can make system is a weak one. The model developed in this present work is able to make a sales prediction of the year 2016 performance whereas the training data used for model exploratory analysis range from 2008 to 2015 and this can positively affect the growth of an organization.

5.2 Conclusions

Based on results of this study, we draw the following conclusions:

1. That most decision makers in telecommunication providers agree that the existing system have some weakness. This is not as a result of lack skills or staff to manage the existing system but the present level of technology in which the existing operates is the major cause of its defects. Telecommunication providers are well known to operate on high level of technology and high competitive market environment, this makes it imperatives to bridge the gap on the existing system, by improving or changing the existing data analytic system with a more powerful data mining techniques model proposed in this study.
2. That collectively, the identified performance indicators have significant effect on discovered hidden knowledge and quality of information use for decision making.
3. In this study, it can be justifiably concluded that data mining techniques which captures all the performance indicators enumerated in the section above should adopted by telecommunication providers in Nigeria and also the financial benefits are awesome.

5.3 Recommendations

Based on the summary of findings, the following recommendations were made:

1. Telecommunication providers must change or improve on their present system to accommodate certain dimensions and superiority such as accessibility, easy to use, presentation quality, precision and speed (response in time). Since they are found to be weak on the current system used for analytical of data use for decision making.
2. Telecommunication providers should embark on the project of change/modification of the existing system since the profit realized if the new system are adopted is outstanding.

REFERENCES

Aftarczuk, K. (2007). Evaluation of selected data mining algorithms implemented in Medical Decision Support Systems. Master thesis Software Engineering thesis number MSE-2007-21, School of Engineering at Blekinge Institute of Technology, pp 9-14.

Agrawal, R., Imielinski T., and Swami A. (1996). Mining Association Rules between Sets of Items in Large Databases, SIGMOD'93, pp. 207-216.

Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Conference, Washington DC, USA, pp. 1-10.

Agrawal, R., Mannila H., Srikant R., Toivonen H and Verkamo A.I. (1996). Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining, pp. 6-8

Agrawal, R., and Srikant, R. (1994). Fast algorithms for mining association rules. VLDB (Very Large Data Base Endowment), 20th VLDB Conference, Santiago, Chile, pp. 487-499

Ahaiwe and Ukwandu (2012). Management information Systems, pp. 163-176

Ahmed and Jiao (2004), Application of data mining techniques in customer relationship management and Retail industry. Expert Systems with

Application: International Journal vol 36, Issue 2, March 2009, pp. 2592-2602.

Ahn, K. (2012). Effective product assignment based on association rule mining in retail. Expert Systems with Applications, 39, 551-556.

Abhinab Choudhury (2011). Waterfall Model. Software development life cycle (SDLC) Tutorial.

Akintola K.G., Adetunmbi A.O. and Adeola O.S. (2011). Building Data Warehousing and Data Mining From Course Management Systems: A Case Study of Federal University of Technology (FUTA) Course Management Information Systems. Information Technology for People-Centred Development (Jorunal)

Alhassan, J.K., and Lawal S.A. (2015). Using Data Mining Technique for Scholarship Disbursement. World Academy of Science, Engineering and Technology: International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol: 9, No: 7, pp. 1-4

Alves, R., Ferreira, P., Belo, O., Lopes, J., Ribeiro, J., Cortesao, L and Martins, F. (2006). Discovering telecom fraud situations through mining anomalous behavior patterns. Proceedings of the ACM SIGKDD Workshop on Data Mining for business applications, New York: ACM Press, pp. 1-7

Airtel Nigeria Career Portal: http://careerslip.com/airtel-nigeria

Association of Modern Technologies Professionals (2016). Software Development Methodologies. Retrieved from: http://www.itinfo.am/eng/software-development-methodologies/home page: IT Knowledge portal.

BBC (2014).System Flowchart; GCSE Bitesize, pp.1-7. Retrieved from www.bbc.co.uk/school.

Balasubramanian, M., and Selvarani, M. (2014). International Journal of Scientific and Research Publications. Vol 4, issues 4, ISSN 2250-3153, pp. 1-5

Baritchi, A., Cook, D., and Holder, L. (2000). Discovering structural patterns in telecommunications data. Proceedings of the Thirteenth Annual Florida AI Research Symposium, pp. 82-85.

Basaltoa N., Bellottib R., De Carlob F., Facchib P., and Pascazio S. (2005).Clustering stock market companies via chaotic map synchronization, Physica A345, pp. 196

Bellazzi R., and Zupan, B. (2008), Predictive data mining in clinical medicine: Current issues and guidelines, International Journal of Medical Informatics, vol. 77, No. 2, pp. 81-97.

Berry M., and Linoff G., (2004), Data Mining Techniques: for marketing, sales, and customer relationship management (2nd. edition), Wiley, pp.540-598

Berry and Linoff (1999). Mastering data mining: The art and science of customer relationship management, John Wiley and Sons, Inc. pp. 512

Bharatheesh, T., and Iyengar, S., (2004). Predictive Data Mining for Delinquency Modeling: ESA/VLSI, pp. 99-105.

Blanchard, B.S., and Wolt Fabrycky (2010).Systems Engineering Analysis, third edition, pp. 1-274

Black, Rex; (2002). Managing the Testing Process (2nd ed.). Wiley Publishing. ISBN 0-471-22398-0

Boulicaut, J.F., Klemettinen, M., and Mannila, H. (1999). Modeling KDD Processes within the Inductive Database Framework. Data Warehousing and Knowledge Discovery: DaWaK, M.K. Mohania and A. M. Tjoa (eds), pp. 293-302

Boulicaut, J.F., Klemettinen, M., and Mannila, H. (1998). Querying inductive databases: a case study on the MINE RULE operator. 2nd European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD'98), Nantes, France, pp. 194-202.

Bramer, M. (2013). Principles of Data Mining. Undergraduate Topics in Computer Science. DOI 10.1007/978-1-4471-4884-5 2, Springer-Verlag London.

Berson A., Smith S., and Thearling K. (2000). Building data mining applications for CRM, McGraw-Hill.

Business dictionary (2016). WebfFnance,Inc.www.bussinessdictionary.com

Cai L., and Zhu Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal. 14, pp.2. DOI: http://doi.org/10.5334/dsj-2015-002.

Carrier C. G., and Povel O., (2003), Characterising data mining software, Intelligent Data Analysis pp.7**,** pp.181 - 192.

Changchien, S. W., and Lu, T. (2001). Mining association rules procedure to support on-line recommendation by customers and products fragmentation. Expert Systems with Applications, Vol 20, pp 325-335.

Chang, C.C., and Keisler, H. J. (1990). Model Theory. North-Holland 1973.

Chen, M., Han, J., and Yu, P. (1996). Data mining: An overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering*, 8*(6), 866-883.

Cheesemen P., and John Stutz (1996). Clustering, By Rui Xu, Don Wunsc.

Chi-Lin Lu and Ta-Cheng Chen (2010). International Journal of Economics and International Finance.

Clemente M., Giner-Bosch V., and San Matias, S. (2005). Assessing classification methods for churn prediction by composite indicators. Department of Applied Statistics, Operations Research and Quality, Universitat Politecnica de Valencia, Camino de Vera s/n, 46022 Spain.

Cortes, C. and Pregibon, D. (2001).Signature-based methods for data streams. Data Mining and Knowledge Discovery, 5(3), 167-182.

Cox, K., Eick, S., and Wills, G. (1997). Visual data mining: Recognizing telephone calling fraud. Data Mining and Knowledge Discovery, 1(2), 225-231.

CMS (2008). Selecting a Development Approach**.**

Chukwuemerie Uduchukwu(2013).The NCC, Telecom Operators, and Poor Services in Nigeria.

Chris Rygielski A., Jyun-Cheng Wang B., David C., and Yen A.A. (2002). Data mining techniques for customer relationship management. Technology in Society.

Department of DSC & MIS, Miami University, Oxford, OH, USA. Department of Information Management, National Chung-Cheng University, Taiwan, ROC. pp. 483–502

David Bowman (2009). David Bowman's Information Management Checklist. Retrieved from: http://www.information-management-architect.com/prototyping-methodology.html

Dell Inc. (2016). Data Mining Techniques. Retrieved from: https://en.wikipedia.org/wiki/Design_for_Six_Sigma.

Devitt, A., Duffin, J., and Moloney, R. (2005). Topographical proximity for mining network alarm data. Proceedings of the 2005 ACM SIGCOMM Workshop on Mining Network Data (pp. 179-184), New York: ACM Press.

Edward Frazelle H. (2001). Supply Chain Strategy: the logistics of Supply Chain Management.

Eleven40 Pro Theme (2016). Genesis Framework · WordPress.

Elliot Bendoly (2003). Enterprise Resource Planning: Developments and Directions for Operations Management Research. Journal of operational research 146(2):233-240.

Ehsan Hajizadeh, Hamed Davari Ardakani and Jamal Shahrabi (2010). Application of data mining techniques in stock markets: A survey, Journal of Economics and International Finance. Industrial Engineering Department, Amirkabir University of Technology, Tehran, Iran. Vol. 2(7), pp. 109-118, ISSN 2006-9812. Retrieved from: http://www.academicjournals.org/JEIF

Etisalat Nigeria - Insider Information | Ngcareers: https://ngcareers.com/company/241/**etisalat-nigeria**-241

Frans Coenen (2011). Data mining: past, present and future. The Knowledge Engineering Review. Vol. 26:1, pp. 25–29.

Fathima A.S, Manimegalai, D., and Hundewale, N. (2011). A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the arbovirus-Dengue. International Journal of Computer Science (IJCSI), vol. 8, no. 6, pp. 322-328.

Fashoto, Owolabi, Sadiku and Gbadiyan, (2013). Application of Data Mining Technique for Fraud Detection in Health Insurance Scheme Using Knee-Point K-Means Algorithm. Australian Journal of Basic and Applied Sciences, Vol 7(8), pp 140-144, ISSN 1991-8178.

Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview. In Advances in Knowledge Discovery and Data Mining.

Fayyad, U.M., Georges Grinstein and Andres Wierse, Morgan Kaufman (2001). Information visualization in data mining and knowledge discovery.

Fayyad U.M. (1996). Data mining and knowledge discovery: making sense out of data, IEEE Expert **11**(5), pp. 20-25

Fayyad, U.M, Piatetsky-Shapiro G.,Smyth P., and Uthurusamy R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI Press, pp. 307-328.

Floyd S. (2007). Data Mining Techniques for Prognosis in Pancreatic Cancer.

Frans Coenen. (2004).The Knowledge Engineering Review. Vol. 00:0, pp.1–24.

Fonollera D.L. (2009). Feasibility Study, Project Proposal and Business plan.

Gelman, A., Carlin, J.B., Stern H.S and Rubin D.B. (1995). Bayesian Data Analysis. Chapman & Hall.

Gilks, W.R., Richardson S., and Spiegelhalter D.J (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall.

Georges D., Olle Olle and Shuqin Cai (2013), A Hybrid Churn Prediction Model in Mobile Telecommunication Industry, International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 4, No. 1.

Goldsmith R. (2006). Benefit Estimation. Retrieved from: www.projectmanagement.com.

Grover and Mehra (2008). The lure of statistics in data mining. Retrieved from: http://www.amstat.org/publications/jse/v16n1/grover.html (Accessed 31 01 2012).

Ha S., Bae, S and Park S. (2000). Web mining for distance education. In IEEE International conference on management of innovation and technology, pp. 715–719.

Han, J., and Fu, Y. (1995). Discovery of multiple-level association rules from large databases. 21st VLDB (Very Large Data Base Endowment) Conference, Zurich, Switzerland.

Han and Kamber (2001). Data mining: Concepts and Techniques :3rd Edition.

Heckerman, D. (1997). Bayesian Networks for Data Mining. Data Mining and Knowledge Discovery. Vol 1, pp 79-119.

Heikki Mannila (2000). Theoretical Frameworks for Data Mining SIGKDD Explorations. Vol 1, Issue 2, pp 30. Heikki.Mannila@nokia.com

Heikki Mannila (1997). Inductive databases and condensed representations: concepts for data mining. International Logic Programming Symposium 1997, pp. 21-30.

Hofer J. A., Prescott M.B and Heikki T. (2008). Modern Database Management, Pearson International Edition.

Hofer J. A., Prescott M.B and Heikki T. (2008). Modern Database Management, Pearson International Edition.

http://www.academia.edu/19836789/On_the_Use_of_Information_Systems_Research_Methods_in_Datamining.

https://www.the-modeling-agency.com/crisp-dm.pdf (CRISP-DM)

http://www.tiu.ac.jp/org/forum-01/08WuhanDocuments/WHU-ROS-017-Baba-DQA.pdf

http://www.businessdictionary.com/definition/researchmethodology.html#ixzz47REVfqa4

http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/introductory-concepts/basic-concepts/sample-and-population/

http://www.nairaland.com/1089097/mtn-subscriber-base-hits-183m

https://en.wikipedia.org/wiki/**Globacom**

http://careerslip.com/*airtel-nigeria*

Imielinski, T. (1995). A database view on data mining. Knowledge Discovery in Database.

Imielinski T and Mannila. H (1996). A database perspective on knowledge discovery. Communications of the ACM **39**. Vol 11, pp. 58-64.

Idowu, A.P., Akhigbe B.I., Adeosun, O.O., Kayode, A.A., and Osungbade, A.F. (2013). Data Mining Techniques for Predicting Immunize-able Diseases:

Nigeria as a Case Study. International Journal of Applied Information Systems (IJAIS). Foundation of Computer Science FCS, New York, USA Volume 5. Retrieved from www.ijais.org.

Investorword (2016), Cost Estimate.WebFinance, Inc.

Isreal Glenn.D. (1992). Determining Sample Size PD006.

IEEE Standard Computer Dictionary (1990) A Compilation of IEEE Standard Computer Glossaries; IEEE; New York.

Jain and Dubes, (1998), Algorithms for Clustering Data, mcconnell Internetworking Computer Systems: Interconnecting Networks and Systems. *homepages.inf.ed.ac.uk/rbf/BOOKS/JAIN/Clustering_Jain_Dubes.pdf*

Jar-Long, W., and Shu-Hui C. (2006). Stock market trading rule discovery using two-layer bias decision tree, Expert Systems with Applications.

Jie, S., and Hui L., (2008). Data mining method for listed companies' financial distress prediction, Knowledge-Based Systems.

Katty Schwalbe (2012). Information Technology Project Management: 7th Edition.

Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods, and Algorithms. IEEE Press & John Wiley.

Kleinberg J., Papadimitriou and Raghavan P. (1998). A Microeconomic View of Data Mining, Data Mining and Knowledge Discovery. Vol 2, 4, pp. 311-324

Klemettinen, M., Mannila, H., & Toivonen, H. (1999), Rule discovery in telecommunication alarm data. Journal of Network and Systems Management, 7(4), 395-423.

Khalid Raza (2012). Application of Data Mining In Bioinformatics. Indian Journal of Computer Science and Engineering. Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi-110025, India. Vol 1 No 2, pp. 114-118

Kolawa et al., (2007) Automated Defect Prevention: Best Practices in Software Management. *Wiley-IEEE Computer Society Press, pp. 75.* ISBN 0-470-04212-5.

Knatterud G.L., Rockhold F.W., George S.L., Barton F.B., Davis C.E., Fairweather W.R., Honohan T., Mowery R., O'Neill R. (1998). Guidelines for quality assurance in multicenter trials: a position paper. *Controlled Clinical Trials* 19, pp. 477-493.

Kumari, M., and Godara, S. (2011). Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction. International Journal of Computer Science and Technology (IJCST), vol. 2, no. 2, pp. 304-308.

Kurt Thearling and Alex Berson, Stephen Smith (2005). An Overview of Data Mining Techniques.Data Mining Applications for CRM.

Kwetishe Danjuma and Adenike Osofisan (2015). Evaluation of Predictive Data Mining Algorithm in Erythemato-Squamous Disease Diagnosis. Modibbo Adama University of Technology, Yola, Adamawa State, Nigeria.

Larman Craig (2003). Iterative and Incremental Development: A Brief History" *(PDF). Computer 36 (6), pp. 47–56. doi:10.1109/MC.2003.1204375. ISSN 0018-9162.*

Liebowitz, J. (1988). Expert System Applications to Telecommunications. New York, NY: John Wiley and Sons.

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating classification and association rule mining. Knowledge Discovery and Data Mining, pp. 80-86.

Lustigi et.at. (2010). Operational research and Analytics. Retrieved from: www.whatisanalytics.co.uk/jm/index.php/articles/or/102-or-analytics

Margaret Rouse (2005). Guide to managing a data quality assurance program.

Mark Brown and John Brocklebank (2003).Data Mining. SAS Institute Inc.

Mark van Rijmenam (2014).Founder of Datafloq. Think Bigger - Developing a Successful Big Data Strategy for Your Business.

Martin Brown (2012). Data mining as a process. Retrieved from http://www.ibm.com/developerworks/library/ba-data-mining-techniques/

Martin James (1991). Rapid Application Development. Macmillan. ISBN 0-02-376775-8.

Michalski, R.S., and Wnek, J. (1993) Constructive induction an automated design of knowledge representation spaces for machine learning. Reports of the Machine Learning and Inference Laboratory, MLI 93-11, School of Information Technology and Engineering, George Mason University, Fairfax, VA, November.

Milovic, B., and Milovic, M. (2012). Prediction and Decision Making in Health Care using Data Mining, International Journal of Public Health Science (IJPHS), vol. 1, no. 2, pp. 69-78.

Michalski, R.S. (1997). Seeking knowledge in the deluge of facts, Fundamenta Informaticae **30**, pp. 283-297.

Mehta M., Rissanen J., and Agrawal R. (1995). MDL-based decision tree pruning. In U.M. Fayyad, R. Uthurusamy (Eds.) Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD 1995), AAAI Press, Montreal, Canada, 216-221.

Mishra, D., Das A.K and Mishra, S. (2010). Predictive Data Mining: Promising Future and Applications. International Journal of Computer and Communication Technology`, vol. 2, no. 1, pp. 20-28.

Mohammed Zaki (2001). What is *sequence database and sequential* pattern *mining*; Methods for SPADE (*Sequential pattern Discovery* using Equivalent Class).

Mubin M. Shaikh (2013). Create First Data WareHouse. Retrieved from:

http://www.codeproject.com/Articles/652108/Create-First-Data-WareHous

Muh-Cherng, W., Sheng-Yu, L., and Chia-Hsin, L. (2006). An effective application of decision tree to stock trading, Expert Systems with Applications.

Mutlu Yüksel Avcilar and Emre Yakut (2014). Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store. Canadian Social Science Vol. 10, No. 3, pp. 75-83, DOI: 10.3968/4541

Mykola Pechenizkiy, Seppo Puuronen and Alexey Tsymbal (1998). On The Use of Information Systems Research Methods In Data Mining. Retrived from http://www.academia.edu/19836789/On_the_Use_of_Information_Systems_Research_Methods_in_Datamining.

NG Data. (2014). Predicting and Preventing Banking Customer Churn by Unlocking Big Data. NG Data. Retrieved from http://www.ngdata.com/

Nyce, C. (2007). Predictive Analytics. AICPCU-IIA, Pennsylvania.

Optimization group (2016) Data Modeling and Data Mining. Retrieved from: http://www.optimizationgroup.com/methods/data-modeling-data-mining/

Osmanbegović, E., and Suljić, M. (2012). Data Mining Approach For Predicting Student Performance, Economic Review. Journal of Economics and Business, vol. X, no. 1, pp. 3-12.

Oyeniyi, A.O., and Adeyemo, A.B. (2015). Customer Churn Analysis in Banking Sector Using Data Mining Techniques: African Journal of Computing & ICT. University of Ibadan, Nigeria. Department of Computer Science. Vol 8. No. 3.

Piprant and Ernst (2008). A Model for Data Quality Assessment. Retrieved from: babap@attglobal.net, denise.mcconnell@rogers.com

Pete Chapman (NCR), Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler). (2000). CRISP-DM 1.0. Step-by-step data mining guide Retrieved from: www.spss.com/worldwide.

Qasem and Eman (2012). International Journal of Advanced Computer Science and Applications.

Ridha Derrouiche and Pongsak Holimchayachotikul and Komgrit Leksakul. (2011).Predictive Performance Model in Collaborative Supply Chain Using Decision Tree and Clustering Technique.

Roetzheim (2003). Estimating Software Costs. Cost Xpert Group, Inc.

Rouse M. (2000). Algorithm; Whatls.com. Retrieved from: whatis.target.com/definition.

Rouse, M. (2016). Data Preparation.

Rouse M. (2014). Predictive modelling. Retrieved from http://searchdatamanagement.techtarget.com/definition/predictive-modeling

Ruth Dilly (1995). Parallel Computer Centre. Queens University Belfast. Version 2.0. Retrieved from: http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_2.html#HEADING8.

Stan Gibilisco (2012). Business intelligence. Retrieved from: http://whatis.techtarget.com/definition/descriptive-modeling

Shaw et al. (2001). The relationship between pay and job satisfaction: A meta-analysis of the literature.

Shu-Hsien Liao, Pei-Hui Chu and Pei-Yuan Hsiao (2012). Data mining techniques and applications: Expert Systems with Applications. Tamkang University, Department of Management Science.

Skalak, D.B. (2002). Speeding Up Mining or Why Is Data Mining Iterative.

Sterritt, R., Adamson, K., Shapcott, C., and Curran, E. (2000). Parallel data mining of Bayesian networks from telecommunication network data.

Sundar, N.A., Latha, P and Chandra, M.R. (2012). Performance Analysis of Classification Data Mining Techniques over Heart Disease Data base. International journal of Engineering Science & Advanced Technology, Vol. 2, No. 3, pp. 470 – 478.

Srikant, R., and Agrawal, R. (1995). Mining generalized association rules. 21st VLDB (Very Large Data Base Endowment) Conference, Zurich, Switzerland.

SQL Server (2016). Data Mining Concepts. Retrieved from: https://msdn.microsoft.com/en-us/library/ms174949.aspx

SQL Server (2016). Data Mining Algorithms: Analysis Services - Data Mining.

Takim, R., and Akintoye, A. (2002). Performance indicators for successful construction project performance. *In:* Greenwood, D (Ed.), 18th Annual

ARCOM Conference. University of Northumbria. Association of Researchers in Construction Management, Vol. 2, pp. 545-55.

Thearling et al. (2011). Visualizing Data Mining Models.

Tsai, P. S. M., and Chen, C. (2004). Mining interesting association rules from customer databases and transaction databases. Information Systems, 29, pp. 685-696.

Tang, K., Chen, Y., and Hu, H. (2008). Context-based market basket analysis in a multiple-store environment. Decision Support Systems, 45, pp.150-163.

Ulrich, K.T., and Eppinger S. D. *(2000). Product and Development: fifth Edition*

Venkatalakshmi, B., and Shivsankar, M. (2014). Heart Disease Diagnosis Using Predictive Data mining. IEEE International Conference on Innovations in Engineering and Technology (ICIET'14), Tamil Nadu, India.

Viaene, S., Derrig, R., and Dedene, G. (2004). A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis. IEEE Transactions on Knowledge and Data Engineering.

Vincent Meens (2012). Introducing a data mining process framework to enable consultants to determine effective data analytics tasks. Delft University of Technology. Master Thesis – Sytems Engineering, Policy Analysis and Management.

Weiss, G., and Hirsh, H. (1998). Learning to predict rare events in event sequences. In R. Agrawal & P. Stolorz (Eds.). Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp.359-363, Menlo Park, CA: AAAI Press.

Weiss, G.M. (2009), Data Mining in the Telecommunications Industry.

Weiss, G. M. (2009). Data Mining in the Telecommunications Industry. Fordham University, USA.

Weiss, S. H., and Indurkhya N. (1998). Predictive Data Mining: A Practical Guide. San Francisco CA: Morgan Kaufmann Publishers.

Weiss, G., Ros, J., and Singhal, A. (1998), ANSWER: Network monitoring using object-oriented rule. Proceedings of the Tenth Conference on Innovative Applications of Artificial Intelligence, Menlo Park: AAAI Press. Pp. 1087-1093

www.is.informatik.uni-duisburg.de/.../im.../MiningSequentialPatterns.ppt

Zentut (2011).Data mining tools and techniques .International Journal of Scientific. Retrieved from: www.ijser.org/paper/Data-Mining-0714.html

Zhou, L., and Yau, S. (2007). Efficient association rule mining among both frequent and infrequent items. Computers and Mathematics with Applications, 54, 737-749.

APPENDIX A

PROGRAM SOURCE CODE

Code listing

(1) db.php

```
<?php
require_once('config/database.php');
class DB {
private     $CONN;
protected   $table;
protected   $data;
function __construct() {
$this->CONN = new
PDO("mysql:host=".DB_HOST.";dbname=".DB_NAME,DB_USER,DB_PASS);
if(!$this->CONN) {die('database error');
}}public function db() {

return $this->CONN;
```

(2) de.php

```
<?php

require_once("lib/Template.php");

require_once("lib/Session.php");

$temp = new Template();

require_once("settings.php");

$set1 = new sessions();

if(!isset($_SESSION))

session_start();

if(isset($_POST['login_submit'])){

$user = $_POST['user'];

$pass = $_POST['pass'];
```

```
$num = $set1->getnumrow("SELECT * FROM admin WHERE
username='".$user."' && password='".$pass."'");

if($num > 0){$_SESSION['admin'] = $_POST['user'];

$_SESSION['user'] = $_POST['user'];

header("location: admin/?");

exit;

//echo $_SESSION['admin'];

}else{$_SESSION['Error']  = array("Invalid Username/Password")

}

}

?>

<!DOCTYPE html>

<html>

<head>

<title>:: ANALYSIS   ::</title>

<link href="css/bootstrap.css" rel='stylesheet' type='text/css' />

<!-- Custom Theme files -->

<link href="css/style.css" rel="stylesheet" type="text/css" media="all" />

<!-- Custom Theme files -->

<script src="js/jquery.min.js"></script>

<!-- Custom Theme files -->

<meta name="viewport" content="width=device-width, initial-scale=1">

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<meta name="keywords" content="My Secret Spot Ui Kit Responsive web
template, Bootstrap Web Templates, Flat Web Templates, Android  Compatible
web template,
```

```
Smartphone Compatible web template, free webdesigns for Nokia, Samsung, LG,
SonyErricsson, Motorola web design" />

<script type="application/x-javascript"> addEventListener("load", function() {
setTimeout(hideURLbar, 0); }, false); function hideURLbar(){
window.scrollTo(0,1); } </script>

<!--webfont-->

<link href='//fonts.googleapis.com/css?family=Play:400,700' rel='stylesheet'
type='text/css'>

<!-- chart -->

<script src="js/Chart.js"></script>

<!-- //chart -->

<!----Calender -------->

<link rel="stylesheet" href="css/clndr.css" type="text/css" />

<script src="js/underscore-min.js" type="text/javascript"></script>

<script src= "js/moment-2.2.1.js" type="text/javascript"></script>

<script src="js/clndr.js" type="text/javascript"></script>

<script src="js/site.js" type="text/javascript"></script>

</head>

<body>

<div class="header">

<div class="container">

<div class="header-bottom">

<div class="header-head text-center">

<h1>My Secret Spot</h1>

</div>

<div class="navigation-strip">
```

```
<nav class="navbar navbar-default" role="navigation">

<div class="navbar-header">

<button type="button" class="navbar-toggle" data-toggle="collapse" data-target="#bs-example-navbar-collapse-1">

<span class="sr-only">Toggle navigation</span>

<span class="icon-bar"></span>

<span class="icon-bar"></span>

<span class="icon-bar"></span>

</button>

</div>

<!--/.navbar-header-->

<div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">

<ul class="nav navbar-nav">

<li><a href="#">Home</a></li>

<li><a href="#">ABOUT US</a></li>

<li><a href="#">GALLERY</a></li>

<li class="dropdown">

<a href="#" class="dropdown-toggle" data-toggle="dropdown">PLACES <b class="caret"></b></a>

<ul class="dropdown-menu multi-column columns-3">

<div class="row">

<div class="col-sm-4">                                                <ul class="multi-column-dropdown">
        <h6>ALL</h6>
        <li><a href="#">Lorem ipsum</a></li>
                <li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>
```

```
<li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>

</ul>

</div>

<div class="col-sm-4"><ul class="multi-column-dropdown">

<h6>MOST POPULAR</h6><li><a href="#">Lorem ipsum</a></li>
                                <li><a href="#">Lorem ipsum</a></li>

<li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>

<li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>
        <li><a href="#">Lorem ipsum</a></li>

</ul>

</div><div class="col-sm-4">

<ul class="multi-column-dropdown">

<h6>NEAREST TO YOU</h6>

<li><a href="#">Lorem ipsum</a></li>

<li><a href="#">Lorem ipsum</a></li>

<li><a href="#">Lorem ipsum</a></li>

</ul>

</div>

<div class="clearfix"></div>

</div>

</ul>

</li>

</ul>
```

```
</div>

<!--/.navbar-collapse-->

</nav>

<div class="nav-right">

<div class="login-pop">

<div id="loginpop"><a href="#" id="loginButton"><span>Login</span></a>

<div id="loginBox">

<form id="loginForm">

<fieldset id="body"><fieldset> <label for="email">Email Address</label>
<input type="text" name="email" id="email">

</fieldset><fieldset>

<label for="password">Password</label>

<input type="password" name="password" id="password">

</fieldset>

<input type="submit" id="login" value="Sign in">

<label for="checkbox"><input type="checkbox" id="checkbox"> <i>Remember
me</i></label>

</fieldset>

<span><a href="#">Forgot your password?</a></span>

</form>

</div>

</div>

</div>

<script src="js/menu_jquery.js"></script>

<div class="search">
```

```
<form>

<input type="text" value="Search" onfocus="this.value = ";" onblur="if (this.value
== '') {this.value = 'Search';}">

<input type="submit" value="">

</form>

</div>

</div>

<div class="clearfix"></div>

</div>

<!--/.navbar-->

</div>

</div>

</div>

<div class="content">

<div class="container">

<div class="span_1_by_4">

<div class="col-md-3 spanfirst">

<div class="span1">

<h3 class="tlt">STATISTICS</h3>

<canvas id="doughnut" height="200" width="300" style="width: 300px; height:
200px;">                                    </canvas>

<ul class="mnth-view">

<li><span class="tw"></span></span>this week</li>

<li><span class="tm"></span></span>this month</li>

<li><span class="lm"></span></span>last month</li>
```

```
</ul>
<script>

var doughnutData = [
{
value: 30,
color:"#e7e7e7"
},
{
value : 50,
color : "#5f8b9e"
},
{
value : 40,
color : "#cbb25c"
},
];
new
Chart(document.getElementById("doughnut").getContext("2d")).Doughnut(dough
nutData);
</script>
</div>
<div class="profile">
<h3 class="tlt">Profile</h3>
<h4>Erica Example</h4>
```

```
<p>Here stands a little description about Erica Example. Lorem ipsum dolor
sunt.</p>

<img src="images/profile.jpg" alt="" />

<div class="followers">

<ul>

<li>

<h3>97</h3>

<p>Followers</p>

</li>

<li>

<h3>112</h3>

<p>Following</p>

</li>

<li>

<h3>43</h3>

<p>Uploaded</p>

</li>

<div class="clearfix"></div>

</ul>

</div>

<div class="follow">

<a class="f-left" href="#">Follow</a>

<a class="f-right" href="#">Unfollow</a>

<div class="clearfix"></div>

</div>
```

```
</div>

<div class="portrait">

<h3 class="tlt">Portrait Post</h3>

<img src="images/portrait.jpg" alt="" />

<h5>Have a look at my own little Happy Place</h5>

<p>Lorem ipsum dolor sit amet, consetetur sar————————————adipscing elitr, sed diam nonipsum dolor sit amet, consetetur sar adipscing elitr, sed diam nonumy eirmod esttempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. </p>

<a href="#">Read more</a>

<div class="postdate">

<p class="date">17 April 2015</p>

<p class="postedby">by <a href="#">Eric_Example123</a></p>

<div class="clearfix"></div>

</div>

</div>

</div>

<div class="col-md-3 spansecound">

<div class="span1">

<h3 class="tlt">STATISTICS</h3>

<canvas id="bar" height="300" width="400" style="width: 400px; height: 300px;"></canvas>

<script>

var barChartData = {

labels : ["Jan","Feb","Mar","Apr","May","Jun","jul"],

datasets : [
```

```
{
fillColor : "#cbb25c",
data : [65,59,90,81,56,55,40]
},
{
fillColor : "#5f8b9e",
data : [28,48,40,19,96,27,100]
}
]
};
new
Chart(document.getElementById("bar").getContext("2d")).Bar(barChartData);
</script>
</div>
<div class="calender">
<h3 class="tlt">CALENDAR</h3>
<div class="cal1 cal_2"> <div class="clndr"><div class="clndr-controls"><div
class="clndr-control-button"><p class="clndr-previous-
button">previous</p></div><div class="month">July 2015</div><div
class="clndr-control-button rightalign"><p class="clndr-next-
button">next</p></div></div><table class="clndr-table" border="0"
cellspacing="0" cellpadding="0"><thead><tr class="header-days"><td
class="header-day">S</td><td class="header-day">M</td><td class="header-
day">T</td><td class="header-day">W</td><td class="header-day">T</td><td
class="header-day">F</td><td class="header-
day">S</td></tr></thead><tbody><tr><td class="day adjacent-month last-month
calendar-day-2015-06-28"><div class="day-contents">28</div></td><td
class="day adjacent-month last-month calendar-day-2015-06-29"><div
class="day-contents">29</div></td><td class="day adjacent-month last-month
calendar-day-2015-06-30"><div class="day-contents">30</div></td><td
```

class="day calendar-day-2015-07-01"> 1 </td><td class="day calendar-day-2015-07-02"> 2 </td><td class="day calendar-day-2015-07-03"> 3 </td><td class="day calendar-day-2015-07-04"> 4 </td></tr><tr><td class="day calendar-day-2015-07-05"> 5 </td><td class="day calendar-day-2015-07-06"> 6 </td><td class="day calendar-day-2015-07-07"> 7 </td><td class="day calendar-day-2015-07-08"> 8 </td><td class="day calendar-day-2015-07-09"> 9 </td><td class="day calendar-day-2015-07-10"> 10 </td><td class="day calendar-day-2015-07-11"> 11 </td></tr><tr><td class="day calendar-day-2015-07-12"> 12 </td><td class="day calendar-day-2015-07-13"> 13 </td><td class="day calendar-day-2015-07-14"> 14 </td><td class="day calendar-day-2015-07-15"> 15 </td><td class="day calendar-day-2015-07-16"> 16 </td><td class="day calendar-day-2015-07-17"> 17 </td><td class="day calendar-day-2015-07-18"> 18 </td></tr><tr><td class="day calendar-day-2015-07-19"> 19 </td><td class="day calendar-day-2015-07-20"> 20 </td><td class="day calendar-day-2015-07-21"> 21 </td><td class="day calendar-day-2015-07-22"> 22 </td><td class="day calendar-day-2015-07-23"> 23 </td><td class="day calendar-day-2015-07-24"> 24 </td><td class="day calendar-day-2015-07-25"> 25 </td></tr><tr><td class="day calendar-day-2015-07-26"> 26 </td><td class="day calendar-day-2015-07-27"> 27 </td><td class="day calendar-day-2015-07-28"> 28 </td><td class="day calendar-day-2015-07-29"> 29 </td><td class="day calendar-day-2015-07-30"> 30 </td><td class="day calendar-day-2015-07-31"> 31 </td><td class="day adjacent-month next-month calendar-day-2015-08-01"> 1 </td></tr></tbody></table>

```
<div class="weather">
<h3 class="tlt">Weather</h3>
<div class="t-weather">
<h5>Today</h5>
<img src="images/weather.png" alt="weather" />
<h3>31 C°</h3>
<p>50% chance of rain</p>
<ul class="time">
<li>8am</li>
<li>10am</li>
<li>12am</li>
<li>2pm</li>
<li>4pm</li>
<li>6pm</li>
<li>8pm</li>
</ul>
</div>
<div class="weather_by_days">
<div class="weather_by_day">
<p class="w-day">FRIDAY</p>
<p class="w-num">22</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">SATURDAY</p>
```

```
<p class="w-num">24</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">SUNDAY</p>
<p class="w-num">24</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">MONDAY</p>
<p class="w-num">26</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">TUESDAY</p>
<p class="w-num">30</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">WEDNESDAY</p>
<p class="w-num">31</p>
<div class="clearfix"></div>
</div>
</div>
</div>
```

```
</div>

<div class="col-md-3 spanthird">

<div class="form-group"><?php

Template::message();


?> </div>

<div class="user-login">

<h3 class="tlt">User Login</h3>

<div class="login-form">

<form method="post">

<div>

<span>Username</span>

<input type="text"  value="Username or E Mail address" onfocus="this.value = ";"
onblur="if (this.value == '') {this.value = 'Username or E Mail address';}"
name="user">

</div>

<div>

<span>Password</span>

<input type="password" class="text" value="enter your Password"
onfocus="this.value = ";" onblur="if (this.value == '') {this.value = 'enter your
Password';}" name="pass">

</div>

<!--<a href="#">Forgot Password</a>  -->

<a class="news-letter" href="#">

<label class="checkbox"><input type="checkbox" name="checkbox"
checked=""><i> </i>Sign me up for Newsletter</label>

</a>
```

```
<input type="submit" name="login_submit" value="send">

</form>

</div>

</div>

<div class="upload">

<h3 class="tlt">Image Upload</h3>

<div class="login-form">

<form>

<div>

<span>Title</span>

<input type="text" class="text" value="How do you want to name it?" onfocus="this.value = ";" onblur="if (this.value == '') {this.value = 'How do you want to name it?';}">

</div>

<div>

<span>Short Description</span>

<input type="text" class="text" value="Where, When, Why ..." onfocus="this.value = ";" onblur="if (this.value == '') {this.value = 'Where, When, Why ...';}">

</div>

</form>

</div>

<form id="upload" method="post" action="upload.php" enctype="multipart/form-data">

<div id="drop">

<a>Upload</a>

<input type="file" name="upl" multiple />
```

```
</div>
<ul>
<!-- The file uploads will be shown here -->
</ul>
</form>
<!-- JavaScript Includes -->
<script src="js/jquery.knob.js"></script>
<!-- jQuery File Upload Dependencies -->
<script src="js/jquery.ui.widget.js"></script>
<script src="js/jquery.iframe-transport.js"></script>
<script src="js/jquery.fileupload.js"></script>
<!-- Our main JS file -->
<script src="js/script.js"></script>
</div>
<div class="settings">
<h3 class="tlt">settings</h3>
<div class="setting-one">
<h5>Settings1</h5>
<div class="demo5">
<div class="switch demo3">
<input type="checkbox">
<label><i></i></label>
</div>
</div>
<div class="clearfix"></div>
```

```
</div>
<div class="setting-two">
<h5>Settings2</h5>
<div class="demo6">
<div class="switch demo3">
<input type="checkbox" checked="">
<label><i></i></label>
</div>
</div>
<div class="clearfix"></div>
</div>
<a class="advancedsettings" href="#">Advanced Settings</a>
</div>
</div>
<div class="col-md-3 spanfourth">
<div class="featured">
<h3 class="tlt">Featured</h3>
<div class="feature">
<h5>Algier</h5>
<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam
nonumy.</p>
<a class="lightbox" href="#goofy1">
<img src="images/f1.jpg" alt="" />
<div class="zoom-img">
<img src="images/zom.png" alt="" />
```

```
</div>
</a>
<div class="lightbox-target" id="goofy1">
<img src="images/pic1.jpg" alt="" />
<a class="lightbox-close" href="#"> </a>
<div class="clearfix"> </div>
</div>
</div>
<div class="feature">
<h5>San Diego</h5>
<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam
nonumy.</p>
<a class="lightbox" href="#goofy2">
  <img src="images/f2.jpg" alt="" />
<div class="zoom-img">
<img src="images/zom.png" alt="" />
</div>
</a>
<div class="lightbox-target" id="goofy2"><img src="images/pic2.jpg" alt="" />
<a class="lightbox-close" href="#"> </a>
<div class="clearfix"> </div>
</div>
</div>
<div class="seemore">
<a class="sm" href="#">seemore</a>
```

```
<a class="smi" href="#"><img src="images/arrow-right.png" alt="" /></a>
<div class="clearfix"></div>
</div>
</div>
<div class="weather">
<h3 class="tlt">Weather</h3>
<div class="t-weather cloud">
<h5>Today</h5>
<img src="images/cloud.png" alt="weather" />
<h3>24 C°</h3>
<p>50% chance of rain</p>
<ul class="cloudtime">
<li>8am</li>
<li>10am</li>
<li>12am</li>
<li>2pm</li>
<li>4pm</li>
<li>6pm</li>
<li>8pm</li>
</ul>
</div>
<div class="weather_by_days">
<div class="weather_by_day">
<p class="w-day">FRIDAY</p>
<p class="w-num">22</p>
```

```
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">SATURDAY</p>
<p class="w-num">24</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">SUNDAY</p>
<p class="w-num">24</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">MONDAY</p>
<p class="w-num">26</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">TUESDAY</p>
<p class="w-num">30</p>
<div class="clearfix"></div>
</div>
<div class="weather_by_day">
<p class="w-day">WEDNESDAY</p>
<p class="w-num">31</p>
```

```
<div class="clearfix"></div>

</div>

</div>

</div>

</div>

<div class="clearfix"></div>

</div>

<div class="span_1_by_2">

<div class="col-md-6 onehalf">

<div class="submit-form">

<h3 class="tlt">Submit Form</h3>

<form>

<div id="lefthalf">

<span>Name</span>

<input type="text" class="text" value="enter your first and last name"
onfocus="this.value = ";" onblur="if (this.value == '') {this.value = 'enter your first
and last name';}">

</div>

<div id="righthalf">

<span>Email</span>

<input type="text" class="text" value="enter your email adress here"
onfocus="this.value = ";" onblur="if (this.value == '') {this.value = 'enter your
email adress here';}">

</div>

<div class="clearfix"></div>

<div>
```

```
<span>Your message</span>

<textarea onfocus="this.value = ';' onblur="if (this.value == '') {this.value = 'Type your message here ...';}">Type your message here ...</textarea>

</div>

<input type="submit" value="send">

</form>

</div>

<div class="long-text-post">

<img src="images/lpost.jpg" alt="" />

<div class="long-text-post-text">

<div class="ltptl">

<h3 class="tlt">Long Text Post</h3>

<h4>Have a look at my own little Happy Place</h4>

<p>Lorem ipsum dolor sit amet, consetetur sar————————————————adipscing elitr, sed diam nonumy eirmod esttempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

Lorem ipsum dolor sit amet, consetetur sura————————————————adipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo. </p>

</div>

<div class="ltptr">

<p>itr, sed diam nonumy eirmod esttempor aret invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

Lorem ipsum dolor sit amet, consetetur satur————————————————adipscing elitr, sed diam nonumy eirmod est tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero su eos et accusam et justo duo dolores et.</p>

<a href="#">read more</a>
```

```
</div>
<div class="clearfix"></div>
<div class="postdate1">
<p class="date">17 April 2015</p>
<p class="postedby">by <a href="#">Eric_Example123</a></p>
<div class="clearfix"></div>
</div>
</div>
</div>
<div class="short-text-post">
<h3 class="tlt">Short Text Post</h3>
<div class="short-post">
<div class="spli">
<img src="images/lpost.jpg" alt="" />
</div>
<div class="sprt">
<h4>Have a look at my own little Happy Place</h4>
<p>Lorem ipsum dolor sit amet, consetetur sar————————————————adipscing elitr, sed diam nonumy eirmod esttempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. </p>
<a href="#">read more</a>
</div>
<div class="clearfix"></div>
<div class="postdate">
<p class="date">17 April 2015</p>
```

```
<p class="postedby">by <a href="#">Eric_Example123</a></p>

<div class="clearfix"></div>

</div>

</div>

</div>

<div class="treaser-post">

<h3 class="tlt">Teaser Post</h3>

<div class="treaser">

<div class="treaser-left">

<img src="images/tpost.jpg" alt="" />

</div>

<div class="treaser-right">

<h4>Have a look at my own little Happy Place</h4>

<p>Lorem ipsum dolor sit amet, consetetur sar————————————————adipscing
elitr, sed diam nonumy.</p>

<a href="#">read more</a>

</div>

<div class="clearfix"></div>

<div class="postdate1">

<p class="date">17 April 2015</p>

<p class="postedby">by <a href="#">Eric_Example123</a></p>

<div class="clearfix"></div>

</div>

</div>

</div>
```

```
</div>
<div class="col-md-6 secoundhalf">
<div class="world_map">
<h3 class="tlt">World map</h3>
<div class="map_container"><div id="vmap" style="width: 100%; height:
400px;"></div></div>
</div>
<!-- map -->
<link href="css/jqvmap.css" rel='stylesheet' type='text/css' />
<script src="js/jquery.vmap.js"></script>
<script src="js/jquery.vmap.sampledata.js" type="text/javascript"></script>
<script src="js/jquery.vmap.world.js" type="text/javascript"></script>
<script type="text/javascript">
jQuery(document).ready(function() {
jQuery('#vmap').vectorMap({
map: 'world_en',
backgroundColor: '#333333',
color: '#ffffff',
hoverOpacity: 0.7,
selectedColor: '#666666',
enableZoom: true,
showTooltip: true,
values: sample_data,
scaleColors: ['#C8EEFF', '#006491'],
normalizeFunction: 'polynomial'
```

```
});
});
</script>
<!-- //map -->
<div class="video">
<iframe
src="https://player.vimeo.com/video/38562069?color=ffffff&title=0&byline=0&p
ortrait=0" frameborder="0" webkitallowfullscreen mozallowfullscreen
allowfullscreen></iframe>
</div>
<div class="comments">
<h3 class="tlt">Comments</h3>
<div class="activity_box">
<div class="scrollbar" id="style-2">
<div class="commentlady">
<div class="user">
<img src="images/user2.jpg" alt="" />
</div>
<div class="user-comment">
<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua ...</p>
</div>
<div class="clearfix"></div>
</div>
<div class="comment-gent">
<div class="user-comment1">
```

```
<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua ...</p>

</div>

<div class="user1">

<img src="images/user1.jpg" alt="" />

</div>

<div class="clearfix"></div>

</div>

<div class="clearfix"></div>

<div class="commentlady">

<div class="user">

<img src="images/user2.jpg" alt="" />

</div>

<div class="user-comment">

<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua ...</p>

</div>

<div class="clearfix"></div>

</div>

<div class="comment-gent">

<div class="user-comment1">

<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua ...</p>

</div>
```

```
<div class="user1">
<img src="images/user1.jpg" alt="" />
</div>
<div class="clearfix"></div>
</div>
<div class="clearfix"></div>
<div class="commentlady">
<div class="user">
<img src="images/user2.jpg" alt="" />
</div>
<div class="user-comment">
<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua ...</p>
</div>
<div class="clearfix"></div>
</div>
<div class="comment-gent">
<div class="user-comment1">
<p>Lorem ipsum dolor sit amet, consetetur ilias sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua ...</p>
</div>
<div class="user1">
<img src="images/user1.jpg" alt="" />
</div>
```

```
<div class="clearfix"></div>
</div>
</div>
</div>
</div>
<script src="js/responsiveslides.min.js"></script>
<script>
$(function () {
$("#slider").responsiveSlides({
auto: true,
nav: true,
speed: 500,
namespace: "callbacks",
pager: true,
});
});
</script>
<div class="slider">
<div class="callbacks_container">
<ul class="rslides" id="slider">
<li>
<img src="images/pic1.jpg" alt="">
</li>
<li>
<img src="images/pic2.jpg" alt="">
```

```
</li>
<li>
<img src="images/pic1.jpg" alt="">
</li>
<li>
<img src="images/pic2.jpg" alt="">
</li>
<li>
<img src="images/pic1.jpg" alt="">
</li>
<li>
<img src="images/pic2.jpg" alt="">
</li>
</ul>
</div>
</div>
<!---->
</div>
<div class="clearfix"></div>
</div>
</div>
</div>
<div class="container">
<div class="footer">
<div class="footer-top">
```

```
<div class="footer-grids">

<div class="col-md-3 footer-grid">

<h3>SHARE</h3>

<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua.</p>

<div class="social-icons">

<ul>

<li><a class="facebook" href="#"></a></li>

<li><a class="twitter" href="#"></a></li>

<li><a class="instragram" href="#"></a></li>

<li><a class="pinterest" href="#"></a></li>

</ul>

</div>

</div>

<div class="col-md-3 footer-grid">

<h3>blog</h3>

<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua.</p>

</div>

<div class="col-md-3 footer-grid">

<h3>lorem</h3>

<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut.</p>

<p>Dolor sunt amet consectuer ilias erat prologe.</p>

</div>
```

```
<div class="col-md-3 footer-grid">

<h3>ipsum dolor</h3>

<p>Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut.</p>

<p>Consectuer psum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
ares eirmod tempor.</p>

</div>

<div class="clearfix"></div>

</div>

</div>

<div class="col-md-12 footer-bottom">

<div class="copyrights text-center">

<p>© 2015 My Secret Spot Ui Kit. All Rights Reserved | Template by  <a
href="http://w3layouts.com">  W3layouts</a></p>

</div>

</div>

<div class="clearfix"></div>

</div>

</div>

<!-- for bootstrap working -->

<script type="text/javascript" src="js/bootstrap.js"></script>

<!-- //for bootstrap working -->

</body>

</html>
```

(3) Forget password.php

```
<?php

require_once("aut.php");

require_once("settings.php");

$msg1 = 0;

$logi =  new sessions();

if((isset($_POST['username']))){

//echo md5("admin");

$logi->username = filter_var($_POST['username'],
FILTER_SANITIZE_STRING);

$num = $logi->getnumrow("SELECT * FROM $logi->admin WHERE
username='".$logi->username."'");

if($num > 0){

$password1 = randomString(7);

$logi->exec_code("UPDATE admin SET password='".md5($password1)."'
WHERE username='".$logi->username."'");

$title = "ICAPS Admission Application Signup";

$to = $query1['email'];

$headers = "From: Admissions@icaps.org.ng\r\n";

$headers .= "Reply-To: admissions@icaps.org.ng";

$headers .= "Organization: ICAPS\r\n";

$headers .= "MIME-Version: 1.0\r\n";

$headers .= "Content-type: text/html; charset=iso-8859-1\r\n";

$headers .= "X-Priority: 3\r\n";

$headers .= "X-Mailer: PHP". phpversion() ."\r\n" ;

$message = "<html>
```

```
<head>
<meta http-equiv='Content-Type' content='text/html; charset=iso-8859-1'>
<title>PASSWORD RECOVERY</title>
</head>
<body>
<div align='center'>
Imo College of Advance Professionals Studies  </div><br><br>
Your New Login Credentials is shown below <br><br>
Username : ".$logi->username." <br><br>
Password : ".$password1."<br><br><br><br>
Kindly login to www.icaps.org.ng with it to complete the admission processes
<br><br><br><br>
Signed<br><br> Management
</body>
</html>
";
$email = $logi->username;
// edit below
$from = "ICAPS";
$fromemail = "admissions@icaps.org.ng";
$reply = "admissions@icaps.org.ng";
$subject = "Password Recover";
$body = "Imo College of Advance Professionals Studies
\r\n\r\n
Username : ".$logi->username."\r\n
```

```
Password : ".$password1."\n\r\n
Signed\n\r Management
";
// send code, do not edit unless you know what your doing
$header1 .= "Reply-To: Support <$reply>\r\n";
$header1 .= "Return-Path: Support <$reply>\r\n";
$header1 .= "From: $from <$fromemail>\r\n";
$header1 .= "Organization: ICAPS\r\n";
$header1 .= "Content-type: text/plain \r\n";
mail("$email", "$subject", "$body", $header1);
//mail($logi->username,$title, $message, $headers);;
echo "1";
//$_SESSION['Success']  = array(" An Email Containing your Login Credentials
has been sent to your Email");
}else{
echo  0;
}
}
```

(4) Index.php

```
<?php
if(!is_dir("installation/install")){
header("location: installation/install.php");
exit;
}
require_once("lib/Template.php");
```

```
require_once("lib/Session.php");

$temp = new Template();

require_once("settings.php");

$set1 = new sessions();

if(!isset($_SESSION))

session_start();

if(isset($_POST['login_submit'])){

$user = $_POST['user'];

$pass = $_POST['pass'];

$num = $set1->getnumrow("SELECT * FROM admin WHERE
username='".$user."' && password='".$pass."'");

if($num > 0){

$_SESSION['admin'] = $_POST['user'];

$_SESSION['user'] = $_POST['user'];

header("location: admin/?");

exit;

//echo $_SESSION['admin'];

}else{

$_SESSION['Error']  = array("Invalid Username/Password");

}

}

?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
```

```
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
<title>LOGIN</title>
<link href="css/layout.css" media="all" rel="stylesheet" type="text/css">
<link href="css/bg.css" media="all" rel="stylesheet" type="text/css">
<script type="text/javascript" src="js/jquery.js"></script>
<script type="text/javascript" src="js/validation.js"></script>
</head>
<body class="bodysettingw" >
<table align="center" class=""  height="150" width="900">
<tr>
<td valign="top"  align="right">
</td>
</tr>
</table>
<table align="center" width="900"  height="300" >
<tr>
<td align="right" width="300" valign="middle" >
<img src="images/stat1.png" width="108" height="115" />
</td>
<td>
<form name="userForm1"  method="post" action=""  enctype="multipart/form-
data">
<table  align="left"  width="290"    class="pwd" >
<div class="form-group"> </div>
```

```
<tr><td colspan="4" align="right"><div style="color:#344F5F; text-
decoration:blink;">Control Panel Login</div>        </td></tr>

<tr><td><?php

Template::message();

?></td></tr>

<tr><td>Username</td></tr>

<tr><td><input name="user" type="text" id="user"></td></tr>

<tr><td>Password</td></tr>

<tr><td><input name="pass" type="password" id="pwd"></td></tr>

<tr><td colspan="4" align="left">

<input type="submit"  value="Login"  class="btnS" name="login_submit">

</td></tr>

</table>

</form>

</td>

</tr>

</table></body>

</html>
```

(5) Login.php

```
<?php

require_once("aut.php");

require_once("../config/autoload.php");

require_once("../models/login.php");

require_once("../models/aut.php");

$logi = new login();
```

```
$msg = "";
if((isset($_POST['submit']))){
//echo md5("admin");
$logi->username = $_POST['user'];
$logi->password = md5($_POST['password']);
$dis = $logi->authenti();
if($dis == 1){
$msg = array("");
$_SESSION['user']  = $logi->username;
$_SESSION['current_page'] = 'home';
echo "2";
}else{
$msg = array("Invalid Username / Password");
$_SESSION['Error']  = $msg;
echo "1";
}
}else{
echo "1";
}
```

(6) Settings.php

```
<?php
require_once('db.php');
class sessions extends DB
{
public $username;
```

```
public $password;
public $table1;
public $table;
public $level;
public $status;
public $admin                    = 'admin';
//Populate the user object when it's created
public function __construct()
{
parent::__construct();
//$this->db = parent::db();
$this->table = 'session';
$this->table1 = 'admin';
}
/**
* Fun to check validation
* @param: clas name
* @return:
*
*/
public function authenti(){
$this->db = $this->db();
$sql= "SELECT * FROM admin WHERE username= '".$this->username."' &&
password = '".$this->password."'";
$q  = $this->db->prepare($sql);
```

```
$q->execute();
if($q->errorCode() == 0) {
if($q->rowCount() > 0 ){
return 1;
}else{
return 0;
}
}else{
print_r($q->errorInfo()); die;
}
}
public function exec_code1($sql){
$this->db = $this->db();
$q  = $this->db->prepare($sql);
$q->execute();
if($q->errorCode() == 0) {
return $q;
}else{
return $q->errorInfo();
}
}
public function getnumrow($sql){
$this->db = $this->db();
$q  = $this->db->prepare($sql);
$q->execute();
```

```
return $q->rowCount();
}
public function exec_code($sql){
$this->db = $this->db();
$q  = $this->db->prepare($sql);
$q->execute();
if($q->errorCode() == 0){
return 0;
}else{
return 1;
};
}
public function sections_details($rt){
$q =  sessions::exec_code1("SELECT * FROM section WHERE name = '".$rt."'");
$r = $q->fetch(PDO::FETCH_ASSOC);
return $r['content'];
}
}
function randomString( $length )
{
$seed =
'ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijqlmnopqrtsuvwxyz0123456
789';
$max = strlen( $seed ) - 1;
$string = '';
```

```
for ( $i = 0; $i < $length; ++$i )
$string .= $seed{intval( mt_rand( 0.0, $max ) )};
return $string;
}
```

(7) Signout.php

```
<?php
require_once('../icaps12/config/constants.php');
require_once('../icaps12/lib/Session.php');
$session = new Session();
$session->destroy_sess();
$_SESSION['Error']  = array("Logged Out Successfully");
header("location: index.php");
exit;
?>
```

(8) Upload.php

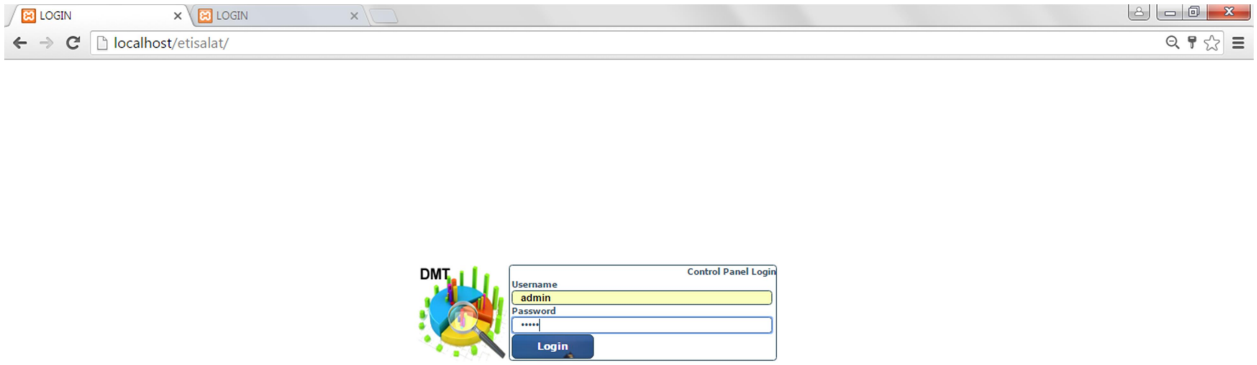
```
<?php
// A list of permitted file extensions
$allowed = array('png', 'jpg', 'gif','zip');
if(isset($_FILES['upl']) && $_FILES['upl']['error'] == 0){
$extension = pathinfo($_FILES['upl']['name'], PATHINFO_EXTENSION);
if(!in_array(strtolower($extension), $allowed)){
echo '{"status":"error"}';
exit;
}
```

```
if(move_uploaded_file($_FILES['upl']['tmp_name'],
'uploads/'.$_FILES['upl']['name'])){

echo '{"status":"success"}';

exit;

}

}

echo '{"status":"error"}';

exit;
```

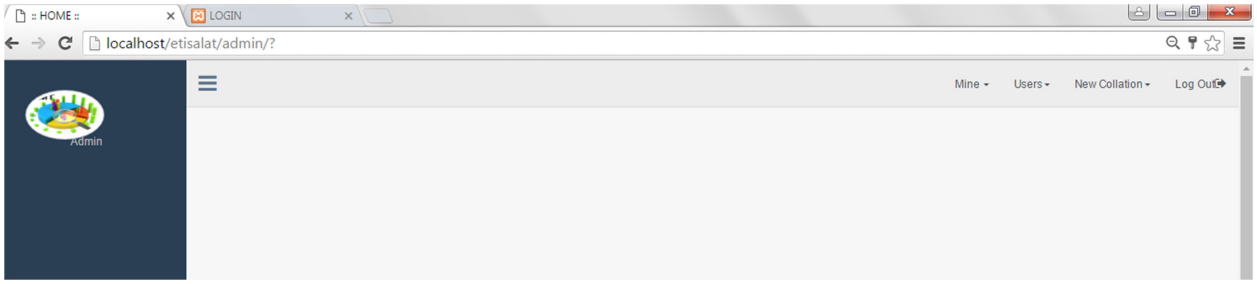
APPENDIX B

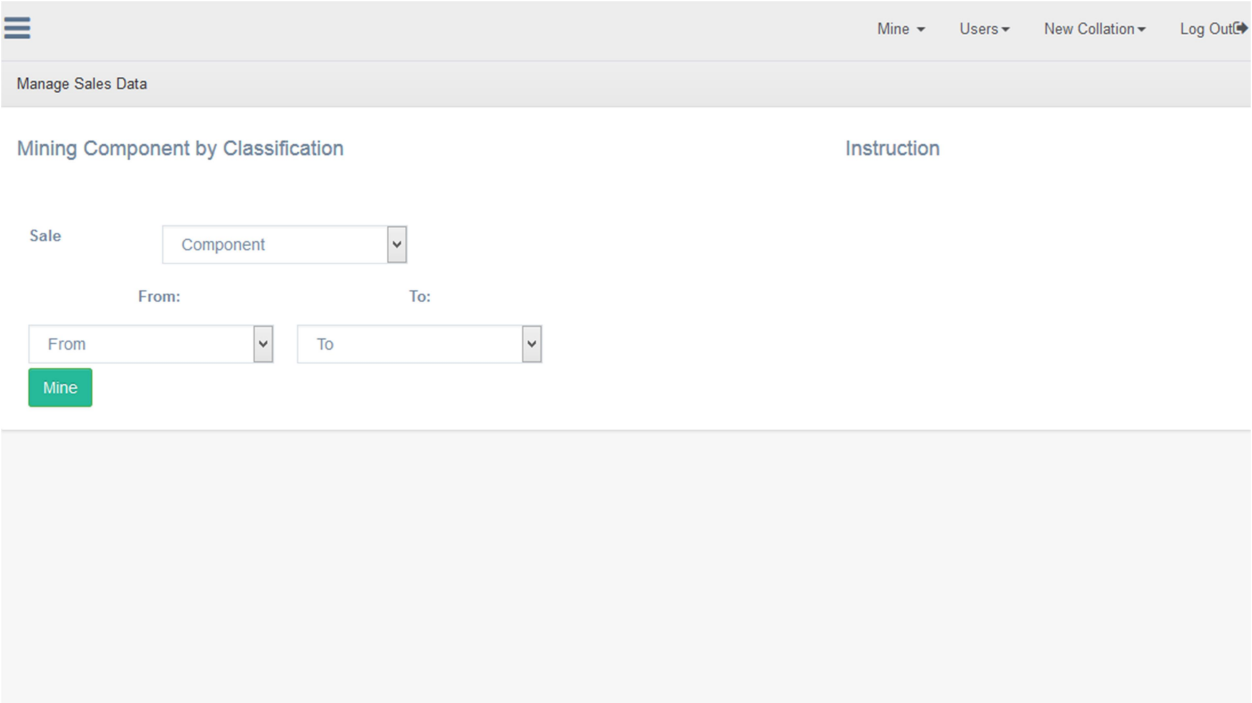
SAMPLES PROGRAM INPUT/OUTPUT INTERFACES

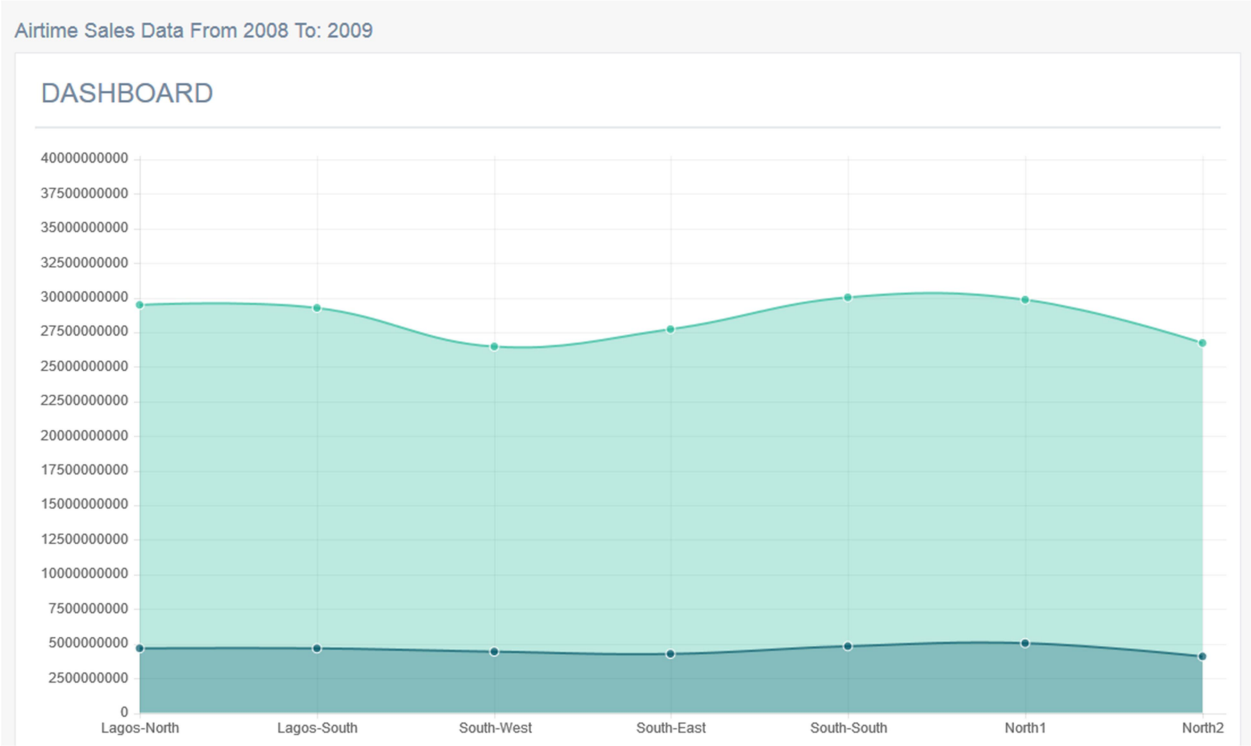
Login Interface



Page Immediately after login




Mine
Users
New Collation
Log Out
Manage Sales Data
Mining Component by Classification
Instruction
Sale
Component
From:
To:
From
To
Mine

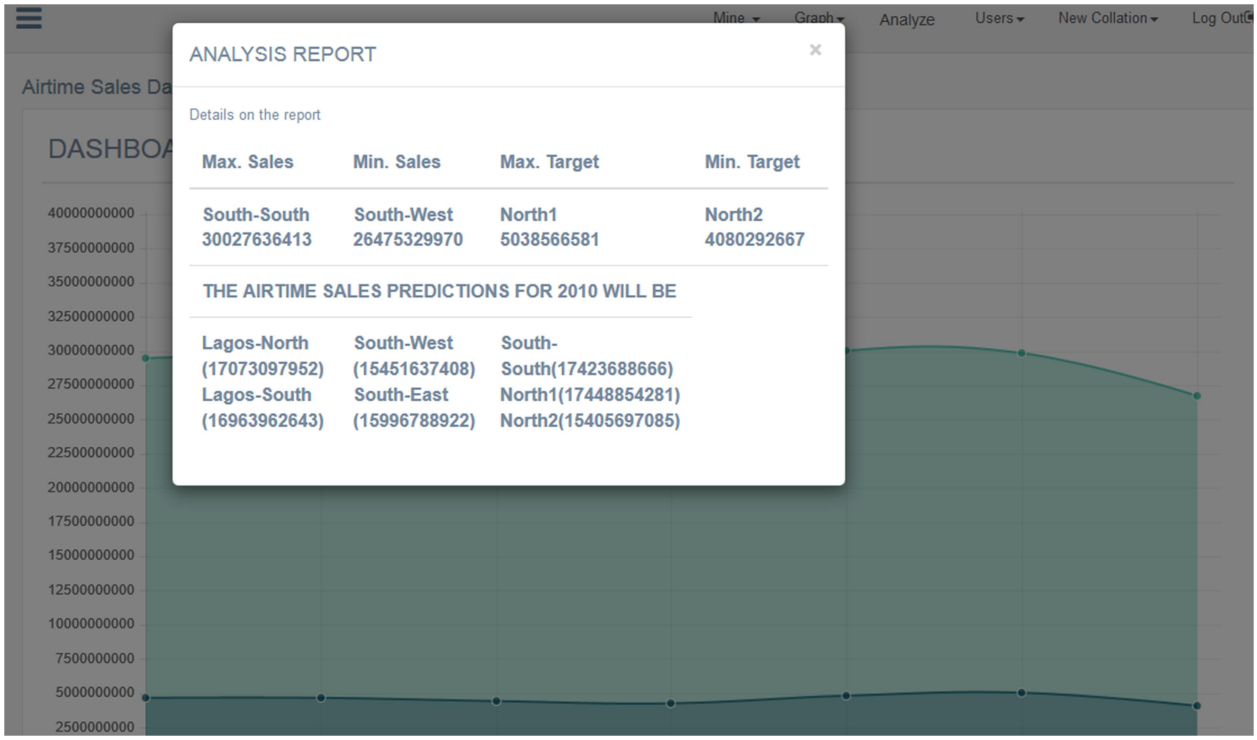

Airtime Sales Data From 2008 To: 2009
DASHBOARD
40000000000
37500000000
35000000000
32500000000
30000000000
27500000000
25000000000
22500000000
20000000000
17500000000
15000000000
12500000000
10000000000
7500000000
5000000000
2500000000
0
Lagos-North
Lagos-South
South-West
South-East
South-South
North1
North2

Mine Graph Analyze Users New Collation Log Out

Airtime Sales Data From 2008 To: 2009

Analysis Table

#	Region	Value Gotten	Targets	Difference(%)
1.	Lagos-North	29483374662	4662821242	84.18
2.	Lagos-South	29265104043	4662821242	84.07
3.	South-West	26475329970	4427944845	83.28
4.	South-East	27727499840	4266078003	84.61
5.	South-South	30027636413	4819740918	83.95
6.	North1	29859141980	5038566581	83.13
7.	North2	26731101502	5038566581	84.74



Mine Graph Analyze Users New Collation Log Out

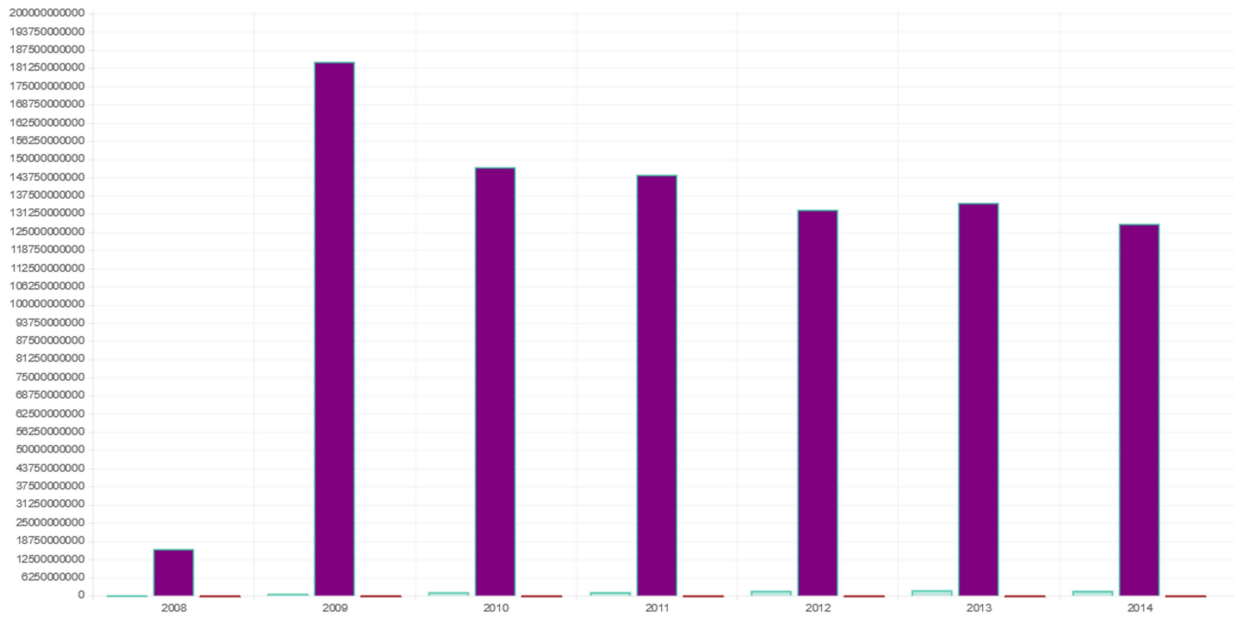
General Products Report

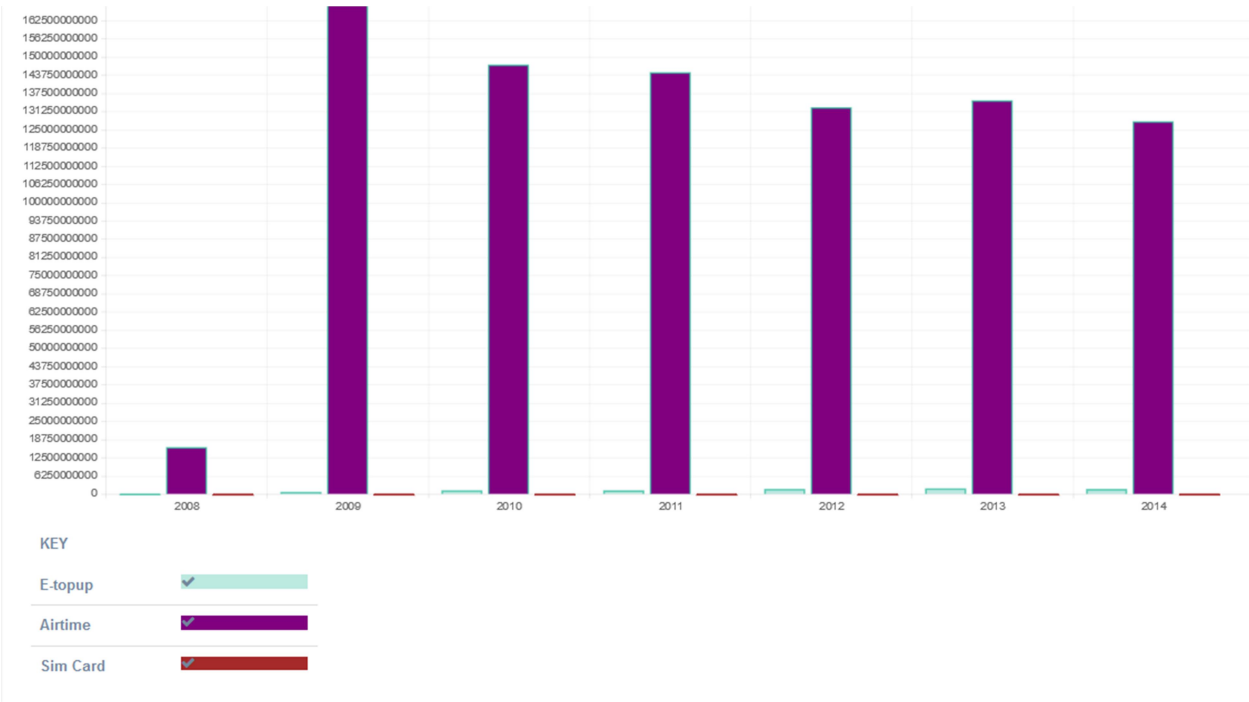
Analysis Table

#	Year	E-topup	Airtime	Sim Card	Total
1	2008	65685002	16046513130	2055046	16114253178
2	2009	690375207	183522675280	18335012	184231385499
3	2010	1249250452	147296657297	13183644	148559091393
4	2011	1216136005	144682355324	13130354	145911621683
5	2012	1725292146	132678801560	11911316	134416005022
6	2013	1915339220	135036553796	11829908	136963722924
7	2014	1699564026	127847573312	14204082	129561341420

General Products Report

Dashboard

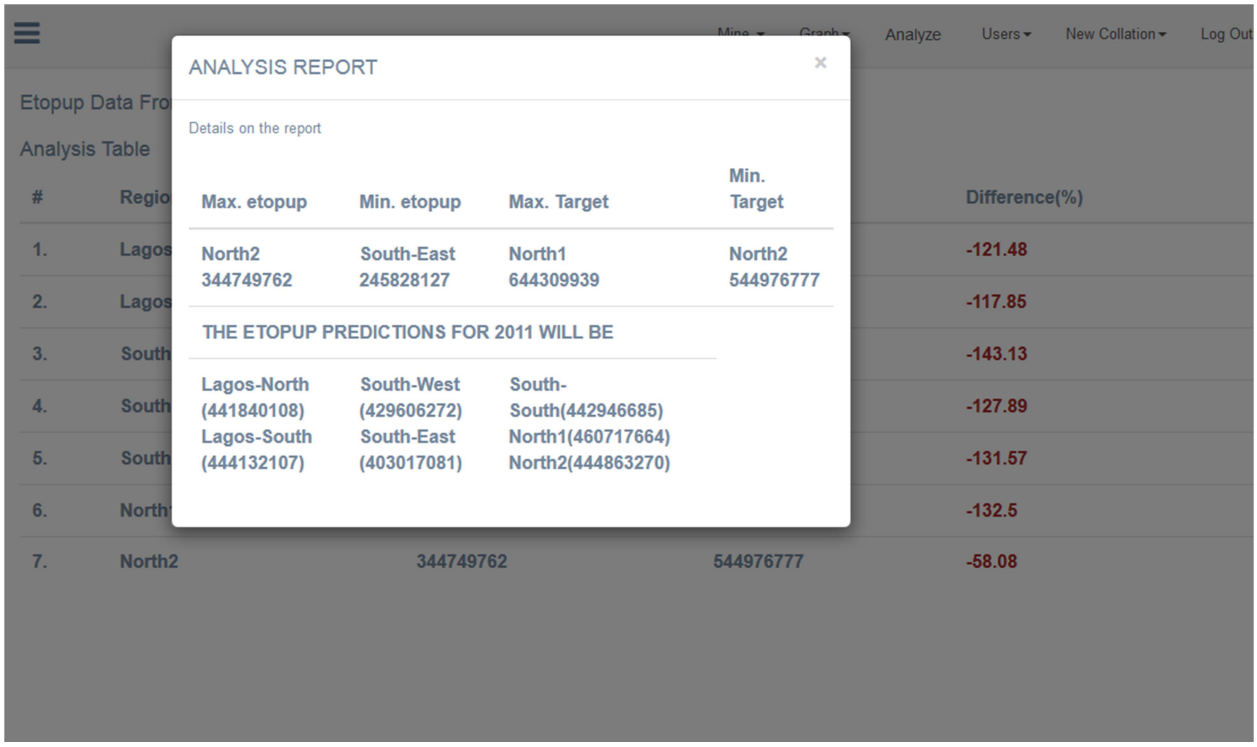


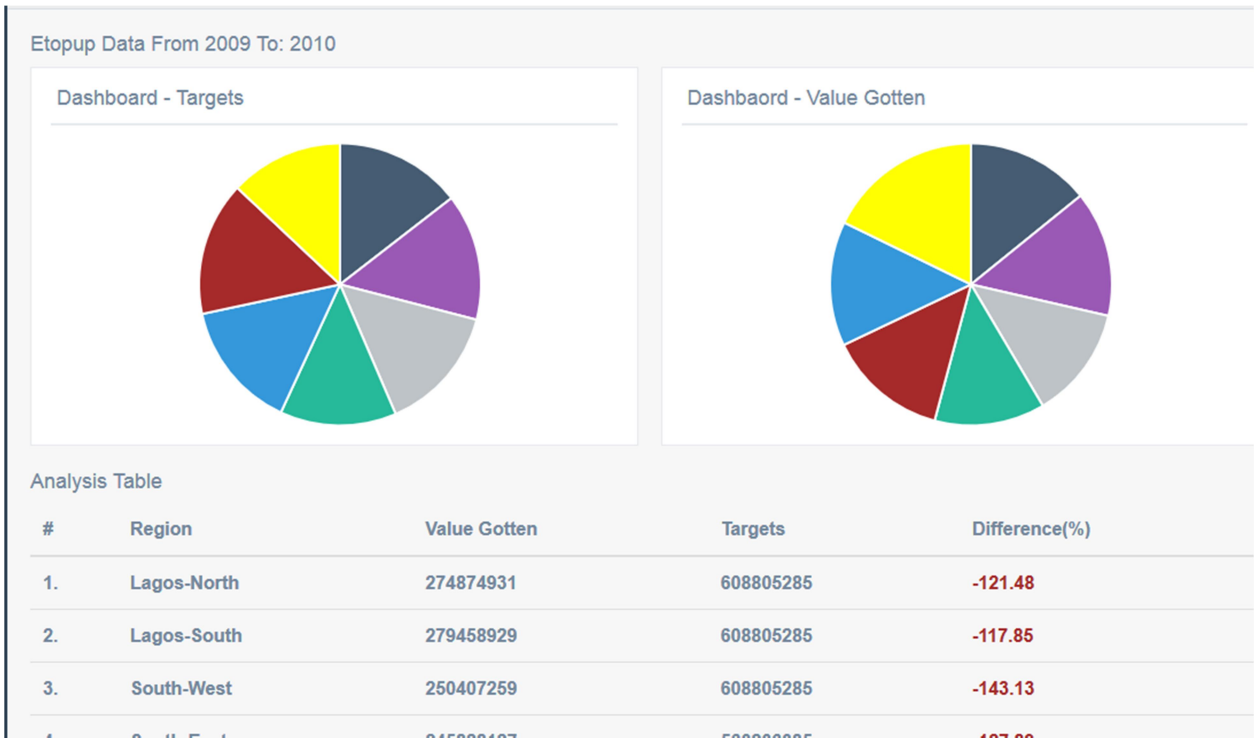


Etopup Data From 2009 To: 2010

Analysis Table

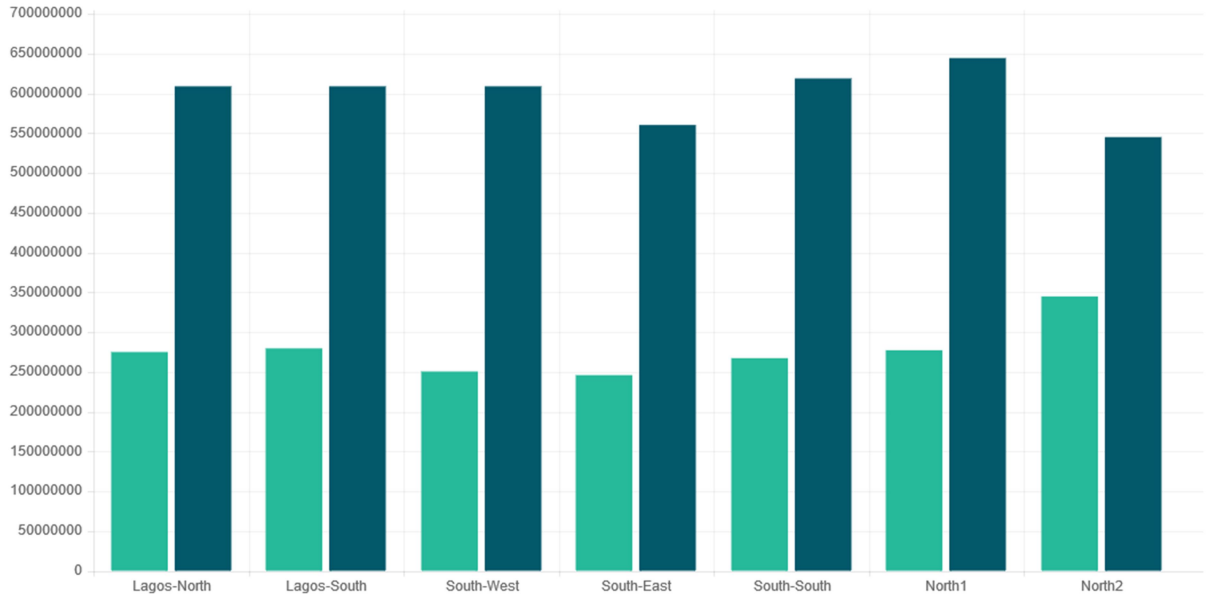
#	Region	Value Gotten	Targets	Difference(%)
1.	Lagos-North	274874931	608805285	-121.48
2.	Lagos-South	279458929	608805285	-117.85
3.	South-West	250407259	608805285	-143.13
4.	South-East	245828127	560206035	-127.89
5.	South-South	267181263	618712106	-131.57
6.	North1	277125388	644309939	-132.5
7.	North2	344749762	544976777	-58.08

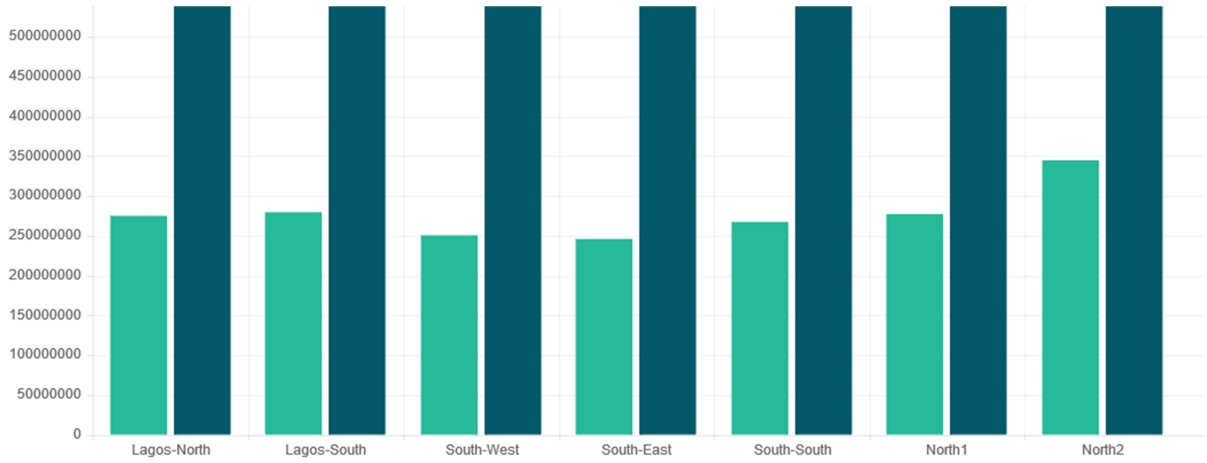

Analyze
Users
New Collation
Log Out
ANALYSIS REPORT
Details on the report
Max. etopup
Min. etopup
Max. Target
Min. Target
North2 344749762
South-East 245828127
North1 644309939
North2 544976777
THE ETOPUP PREDICTIONS FOR 2011 WILL BE
Lagos-North (441840108)
South-West (429606272)
South-South(442946685)
Lagos-South (444132107)
South-East (403017081)
North1(460717664)
North2(444863270)
Etopup Data Fro
Analysis Table
Difference(%)
1. Lagos -121.48
2. Lagos -117.85
3. South -143.13
4. South -127.89
5. South -131.57
6. North -132.5
7. North2 344749762 544976777 -58.08


Etopup Data From 2009 To: 2010
Dashboard - Targets
Dashbaord - Value Gotten
Analysis Table
Region Value Gotten Targets Difference(%)
1. Lagos-North 274874931 608805285 -121.48
2. Lagos-South 279458929 608805285 -117.85
3. South-West 250407259 608805285 -143.13

Etopup Data From 2009 To: 2010

Dashboard





KEY

Amount Gotten

Targets

APPENDIX C

FIRST ASSESSMENT QUESTIONAIRE

QUESTIONNAIRE FOR EVUALUATION OF EFFECTS OF PERFORMANCE INDICATORS IN HISTORICAL DATA ANALYSIS FOR BUSINESS PERFORMANCE AND EVALUATION IN TELECOMMUNICATION PROVIDERS.

Note: Select all that applies where necessary.

SECTION A

1. First Name…….

2. Surname……

3. Highest Level of Education:

WASC/GCE ☐ OND ☐ NCE ☐ HND ☐ B.Sc ☐ MSc ☐ Ph.D ☐

4. What is the Name of your Organization?

5. What is your Job Tittle?

6. Years of Experience?

7. Roles considered are (a) **Information Professionals (IP):** these are people who collect, analyses and maintain the information of the Organization.

(b) Information Consumers (IC): these are people who use the information in the organization for decision marking.

Independent Experts (IE): these are experts that have appropriate amount of practical or academic experience in the practices of the organization being evaluate.

Kindly tick* [X] *in any of the box below to identify the category of your role

IP ☐ IC ☐ IE ☐

SECTION B

With to respect to the method (either Manual or Automated) been used by your organization in carrying out data analysis on historical sales records, kindly rate the following performance indicators below.

Please rate the following performance indicators. Kindly mark* [X] *in box below to identify your answer.

Using a scale ranging from 1 to 5 where 1= extremely low , 2= low , 3= Neutral , 4= High and 5= Extremely High.

8. Does telecommunication providers always have an accurate historical data for business performance analysis; such as historical data of past six years that can be used for decision making for market promotion, regional performance and market segmentation etc.?

1	**2**	**3**	**4**	**5**

9. Does result generate from data analysis of historical data of telecommunication provider easy to interpret?

1	**2**	**3**	**4**	**5**

10. Does presentation quality such as the use of pie chart, line graph and histogram etc affect information use for decision making in your organization and rate this in its level of perfection?

1	**2**	**3**	**4**	**5**

11. Is telecommunication provider historical sales transaction data such as past six year's data easy to obtain or access for business performance evaluation?

1	**2**	**3**	**4**	**5**

12. Can telecommunication provider boast of consistent past six years historical sales data that can be used for decision making for market promotion, market segmentation and price allocation to a certain specific product?

1	**2**	**3**	**4**	**5**

13. Does the present method that your organization used for evaluation of business performance based on analysis of historical data easy to use?

1	**2**	**3**	**4**	**5**

14. Can the present method provide a precise view of a target area such as a request for information of a region that made the highest sales of a particular product in the past six years?

1	**2**	**3**	**4**	**5**

15. Does your present method be it manual or automated used by organization, provides a concise information that can be used for a prompt decision making by decision makers?

1	**2**	**3**	**4**	**5**

16. Does your present system very robust in the sense that it can adjust to any change such as change in figure and still produce an accurate result?

1	**2**	**3**	**4**	**5**

17. How long can it take the present method used by your organization for historical data analysis to respond to requested information such as a request of five years sales records?

Less than 5mins ☐ greater than 5mins less/equal to 20mins ☐ 30 mins time ☐

45mins time ☐ One hour or more ☐

18. Can your management always boast of reliable information based on the present method been used for historical data analysis that generate result used by decision maker?

1	**2**	**3**	**4**	**5**

19. Does your present method for historical data analysis for business evaluation unambiguous?

1	**2**	**3**	**4**	**5**

APPENDIX D

SECOND ASSEMENT QUESTIONNAIRE

QUESTIONNAIRE FOR EVUALUATION OF EFFECTS OF PERFORMANCE INDICATORS IN HISTORICAL DATA ANALYSIS FOR BUSINESS PERFORMANCE IN TELECOMMUNICATION PROVIDERS.

SECTION A

1. First Name…….

2. Surname……

3. Highest Level of Education:

WASC/GCE ☐ OND ☐ NCE ☐ HND ☐ B.Sc ☐ MSc ☐ Ph.D ☐

4. What is the Name of your Organization?

5. What is your Job Tittle?

6. Years of Experience?

7. Roles considered are (a) **Information Professionals (IP):** these are people who collect and maintain the information of the Organization.

(b) Information Consumers (IC): these are people who use the information in the organization

Kindly tick* [X] *in any of the box below to identify the category of your role

IP ☐ IC ☐

SECTION B

Using a scale ranging from 1 to 5 where 1= extremely unimportant, 2= Unimportant, 3= Somewhat Important, 4=Important and 5= extremely important.

Please rate the following performance indicators with respect to its performance in this new model. Kindly mark* [X] *in box below to identify your answer.

PERFORMANCE INDICATORS	1	2	3	4	5
Presentation quality(visualization/understanding)					
Accessibility					
Easy to Use					
Precision					
Robustness					
	1hr	**45mins**	**30mins**	**20mis**	**5mins**
Speed (Response in time)					

APPENDIX E

INTERVIEW QUESTIONS

1. How does telecommunication providers perform their business performance evaluation?
2. How many regions does EMTs, GLO, MTN and Airtel have and the states classified in the region are the same with the sates in region division in Nigeria?
3. How many data analysts does purchase and sales department have?
4. Who are the decision maker in telecommunication providers?
5. What percentage of discount rate do your organization use in carrying out NPV and ROI financial analysis of a proposed project.