



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Data Mining Techniques of Telecommunication Service Providers in Nigeria

¹Bello Musa Ibrahim, ²Kabir Ismail Umar, ³Suleiman Sa'ad, ⁴Abdulwahab Shehu

Department of Information Technology, Modibbo Adama University, Yola - Nigeria^{1,2,3,4}

ABSTRACT: The study analyzes the various data mining techniques of telecommunication companies (MTN, Airtel, Glo and 9Mobile) in Adamawa State, Nigeria. Census sampling technique was adopted to successfully complete the study. Data for the study was collected from the respondents through the use of questionnaires. The results of the findings revealed that Data mining is utmost important to telecommunication industry, and it improves telecommunication services. The study recommends that telecommunication industry should create data warehouse for easy storage and mining. It is also recommended that the stakeholders should encourage the Telecommunication Companies to delivered quality services so as to tackle the issues of fraud and other internet related crimes.

KEYWORDS: Data, Data Mining, Techniques, Telecommunication Service Providers,

I. INTRODUCTION

The telecommunications industry generates and stores a tremendous amount of data (Han et al, 2015). These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers (Roset et al, 2014). The amount of data is so great that manual analysis of the data is difficult, if not impossible. The need to handle such large volumes of data led to the development of knowledge-based expert systems. These automated systems performed important functions such as identifying fraudulent phone calls and identifying network faults. The problem with this approach is that it is time consuming to obtain the knowledge from human experts (the “knowledge acquisition bottleneck”) and, in many cases; the experts do not have the requisite knowledge. The advent of data mining technology promised solutions to these problems and for this reason the telecommunications industry was an early adopter of data mining technology (Roset et al, 2014).

Telecommunication data pose several interesting issues for data mining. The first concerns scale, since telecommunication databases may contain billions of records and are amongst the largest in the world. A second issue is that the raw data is often not suitable for data mining. For example, both call detail and network data are time-series data that represent individual events. Before this data can be effectively mined, useful “summary” features must be identified and then the data must be summarized using these features. Because many data mining applications in the telecommunications industry involve predicting very rare events, such as the failure of a network element or an instance of telephone fraud, rarity is another issue that must be dealt with. The fourth and final data mining issue concerns real-time performance because many data mining applications, such as fraud detection, require that any learned model/rules be applied in real-time (Ezawa and Norton, 2013). Several techniques have also been applied is tackling all these issues in telecommunication companies.

Telecommunication networks are extremely complex configurations of equipment, comprised of thousands of interconnected components. Each network element is capable of generating error and status messages, which leads to a tremendous amount of network data. This data must be stored and analyzed in order to support network management functions, such as fault isolation. This data will minimally include a time stamp, a string that uniquely identifies the hardware or software component generating the message and a code that explains why the message is being generated. For example, such a message might indicate that “controller 7 experienced a loss of power for 30 seconds starting at 10:03 pm on Monday, May 12.” Due to the enormous number of network messages generated, technicians cannot possibly handle every message. For this reason expert systems have been developed to automatically analyze these

messages and take appropriate action, only involving a technician when a problem cannot be automatically resolved (Weiss et al, 2012). This study is focused on MTN Nigeria, Airtel Nigeria, GLO Mobile Nigeria and 9Mobile Nigeria.

Internet usage in Nigeria has grown rapidly; this has had several positive impacts on the socio-economic and educational developments in the country (Akinola, 2013). Unfortunately, the country's image has also suffered as a result of the nefarious activities of some cyber criminals, that has now turn the internet channel for the perpetration of cybercrime. Consequently because of the rapidly growing rate of cyber crimes in Nigeria, an average Nigerian youth today do not fancy hard work or education as they now value cheap or easy money making activities of fraud.

Attracting new customers and customer churn is another issue in telecommunication industry. The possibility of gaining new customer to patronize your service and maintaining the already existing one not to leave your service is a tedious issue. However, this study will examine various data mining techniques of telecommunication companies in Nigeria.

II. RELATED WORK

Sumathi and Sivanandam (2016), are of the view that research and development in this area, often referred to as data mining and knowledge discovery has experienced a tremendous growth in the last couple of years. The goal of these methods and algorithms is to extract useful regularities from large data archives, either directly in the form of "knowledge" characterizing the relations between the variables of interest, Introduction to Data Mining Principles or indirectly as functions that allow to predict, classify, or represent regularities in the distribution of the data.

An enormous proliferation of databases in almost every area of human endeavor has created a great demand for new, powerful tools for turning data into useful, task-oriented knowledge. In the efforts to satisfy this need, researchers have been exploring ideas and methods developed in machine learning, pattern recognition, statistical data analysis, data visualization, neural nets, etc. These efforts have led to the emergence of a new research area, frequently called data mining and knowledge discovery. The current Information Age is characterized by an extraordinary growth of data that are being generated and stored about all kinds of human endeavors (Eze et al, 2016).

An increasing proportion of these data is recorded in the form of computer databases, so that the computer technology may easily access it. The availability of very large volumes of such data has created a problem of how to extract form useful, task-oriented knowledge. Data analysis techniques that have been traditionally used for such tasks include regression analysis, cluster analysis, numerical taxonomy, multidimensional analysis, other multivariate statistical methods, stochastic models, time series analysis, nonlinear estimation techniques, and others. These techniques have been widely used for solving many practical problems. They are, however, primarily oriented toward the extraction of quantitative and statistical data characteristics, and as such have inherent limitations (Sumathi and Sivanandam, 2013).

For example, a statistical analysis can determine covariances and correlations between variables in data. It cannot, however, characterize the dependencies at an abstract, conceptual level and procedure, a casual explanation of reasons why these dependencies exist. Nor can it develop a justification of these relationships in the form of higher-level logic-style descriptions and laws. A statistical data analysis can determine the central tendency and variance of given factors, and regression analyses can fit a curve to a set of data points. These techniques cannot, however, produce a qualitative description of the regularities and determine their dependence of factors not explicitly provided in the data, nor can they draw an analogy between the discovered regularity and regularity in another domain (Zentut fall, 2011). 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).

According to Eze et al, (2016), the enterprise of knowledge discovery is aimed at the automation of the Millennium-old effort of humans to gain information and build models and theories about phenomena in the world around us. Data miners and knowledge discoverers can learn a lot and, i.e., sharpen their awareness, by looking at the scientific method of experimentation, modeling, and validation/falsification in the natural sciences, engineering sciences, social sciences, economics, as well as philosophy.

The next decade of research in network-based information services promises to deliver widely available access to unprecedented amounts of constantly expanding data. Users of many commercial, government, and private information services will benefit from new machine learning technologies that mine impact new knowledge by

integrating and analyzing very large amounts of widely distributed data to uncover and report upon subtle relationships and patterns of events that are not immediately discernible by direct human inspection. (Kolajo, and Adeyemo, 2012).

Data Mining Techniques according to Bharati (2017) comprises of the following:

Classification: Is the processing of finding a set of models (or function) which 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 label. In general, the class labels are not present in the training data simply because they are not known to begin with.

Association Rule: Is a rule which implies certain association relationships among a set of objects (such as “occur together” or “one implies the other”) in a database.

Sequence Discovery: Is a data mining problem closely related to that of association rules. The key difference is in the input dataset. Here a data sample called a sequence is split across multiple consecutive rows in the input table.

Prediction: Any prediction can be thought of as classification or estimation. In prediction the records are classified according to some predicted future behavior or estimated future value.

Visualization: Data visualization is a powerful form of descriptive data mining. It is not always easy to come up with meaningful visualization, but the right picture really can be worth a thousands association rules since the human beings are extremely practiced at extracting meaning from visual scenes.

Online Analytical Processing (OLAP): This idea of dimensionality can be expanded: a table with n-independent attributes can be seen as an n-dimensional space. We need to explore the relationship between these dimensions as standard relational database is not very good at this. OLAP tools were developed to solve this problem. These tools store their data in a special multidimensional format.

Case-Based Learning (K-Nearest Neighbor): Is a prediction technique that is quite similar to clustering-its essence is that in order to predict what a prediction value is in one record look for the records with similar predictor values in the historical database and use the prediction value from the record that it “nearest” to unclassified record.

Decision Tree: Is a predictive model viewed as a tree. Specifically each branch of the tree is a classification question and the leaves of the tree are partition of the dataset with their classification.

Neural Networks: Is a set of connection input/output units and each connection has a weight present with it. Artificial neural networks are computational models composed of many non linear processing elements arranged in a pattern similar to biological neuron networks.

Genetic Algorithm: Are search algorithms based on mechanic of natural selection and natural genetics. They combine survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search.

Data Mining Task

Nora Galambos (2014) viewed that the actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD (Knowledge Discovery in Database) process as additional steps.



Uses of Data Mining Techniques

According to Zentut, (2011) and Rijmenam, (2014), there are several major data mining techniques that have been developed and used in data mining projects recently such as classification, clustering, regression, association rule learning, sequence discovery, prediction and visualization.

Data Mining Applications

Weiss et al, (2012) opined that the two main factors on which Data Mining and Business Intelligence (BI) applications interact on include the availability of the problem that has to be approached and solved by the Data Mining and BI technologies and the availability of Data for implementing the technologies. The main reason behind the significance of Data Mining and Business Intelligence Applications in the Telecommunications industry is the availability of tremendously large volume of data.

Oyeniya and Adeyemo, (2015), Different Data Mining techniques are used to model customer life time value for telecommunication customers. The key element of modeling the life time value for a telecommunication customer is to estimate how long he/she will remain with their current network. It will help the company to predict when a customer is likely leave and to take proactive steps to retain the customer. One of the serious issues that the telecommunications industries face is the customer churn. The process that a customer leaving a company is referred to as churn and churn analysis can be done through numerous systems and methods.

III. METHODOLOGY

The research design used for this study is descriptive research design. Since data characteristics were described using frequencies and percentages, and no manipulations of data or variables were necessary. We discarded other alternatives such as the causal and explanatory research designs, because accurate findings and data analysis may not be achieved.

The population for this study were employees of MTN, Airtel, Glo and 9mobile Yola, Nigeria. The population figure for the study were 100 respondents. Data for this study was collected from the respondents through the use of questionnaires, and field surveys will serve as the main source of primary data for the study. Various analytical tools such as pie charts, bar charts, and tables were used in analyzing data for the study.

IV. EXPERIMENTAL RESULTS

Is Data Mining Used to examine or Explore the Data using Queries?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid SA	28	28.0	28.0	28.0
A	32	32.0	32.0	60.0
N	16	16.0	16.0	48.0
D	14	14.0	14.0	30.0
SD	10	10.0	10.0	24.0
Total	100	100.0	100.0	

DISCUSSION OF FINDINGS

This paper analyzes the various data mining techniques of telecommunication service providers in Nigeria. The aim of the study is to extract implicit, previously unknown and potentially useful patterns from data. The objectives of the study is to evaluate what data mining is, to examine the various data mining techniques of telecommunication company in Nigeria and to identify the challenges of data mining faced by telecommunication company in Nigeria.

In doing so, questionnaires were administered and analyzed on collection. Upon the analysis the following were arrived at:

- i. Data mining is important to telecommunication industry
- ii. Data mining improve telecommunication services
- iii. There are different data mining techniques used in telecommunication services
- iv. Telecommunication company experience challenges in data mining
- v. Data mining significantly impact the performance of telecommunication industry.

From the analysis it can be deduced that data mining is used to examine or explore the data using queries. Data mining is used to determine the pattern and relationship in a sample data. This is because the responses from most of the respondents were in the affirmative. Data mining is all about solving problems by analyzing data already present in database. This is because the responses from most of the respondents were in the affirmative. It can also be deduced that the purpose or effort of data mining normally either to create a descriptive model or predictive model. This is because the responses from most of the respondents were in the affirmative.

V. CONCLUSION AND RECOMMENDATION

It can be seen in the study that telecommunication companies utilize data mining to improve their sales and marketing operation strategies. All the Telecommunication Service Providers in Nigeria offer similar products and services but package it in different ways. The operations within the companies are almost the same. Major sources of revenue to this TSP are sales made from their distribution partners located in all their regions. In sales and marketing department of TI, three products were considered such as airtime, electronic recharge (e-top up) and SIM card sales. The training data analyzed was collected for eight years (2011 to 2019). From the model results, one can mine to get such information as; a precise view of sales in a particular region of the country at a specified period of time, total sales in a region at a specified period of time, a region with highest sales in any of the three product considered, year with highest number of sales etc. Appropriate application of this research work will also assist the organization in forecasting and predicting future outcomes such as; which region will likely make highest sales or profit in the nearest future say two or three years?, which product will likely have highest sales in future? Data Mining as a term used for some set of six activities or tasks as follows: Classification, Estimation, Prediction, Affinity grouping or Association rules, Clustering, Description and Visualization.

The first three tasks-classification, estimation and prediction are all examples of directed data mining or supervised learning. In directed data mining, the goal is to use the available data to build a model that describes one or more particular attribute(s) of interest (target attributes or class attributes) in terms of the rest of the available attributes. The next three tasks-association rule, clustering, and description are examples of undirected data mining that is no attribute is singled out as the target; the goal is to establish some relationship among all the attributes.

There are several major data mining techniques that have been developed and used in data mining projects recently such as classification, clustering, regression, association rule learning, sequence discovery, prediction and visualization etc.

During the last 15 years, various techniques have been proposed to improve the generalization properties of neural estimators. The basic mechanism is to control the richness of the class of possible functions that can be obtained through training.

From the above analysis we concluded that Data mining is important to telecommunication industry, Data mining improve telecommunication services, there are different data mining techniques used in telecommunication services, Telecommunication company experience challenges in data mining, Data mining significantly impact the performance of telecommunication industry.

- i. Telecommunication industry should create data warehouse. Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into data warehouses. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access these data freely. The data analysis software is what supports data mining. Hence, data warehousing provides the enterprise with a memory. Data mining provides the enterprise with intelligence. Any organization or a system in general is faced with a wealth of

data that is maintained and stored, but the inability to discover valuable, often previously unknown information hidden in the data, prevents it from transferring these data into knowledge or wisdom.

To satisfy these requirements, these steps are to be followed.

1. Capture and integrate both the internal and external data into a comprehensive view “Mine” for the integrated data information
2. Organize and present the information and knowledge in ways that expedite complex decision making.
- i. Innovations are also required in the area of database technology to support interactive data mining and knowledge discovery. The user with his knowledge and intuition about the application domain should be able to participate in the search for new structures in data, e.g., to introduce a priori knowledge and to guide search strategies.
- ii. It is also recommended that the stakeholder which is Nigerian Communications Commission (NCC) will mount pressure on Telecommunication Company to delivered quality services, to tackle the issues of fraud and other internet crimes.

REFERENCES

1. Alves R, Ferreira P, Belo O, Lopes J, Ribeiro J, Cortesao L, Martins F (2013). 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
2. Bharati,M.Ramageri. (2017). An Overview of Data Mining Techniques. Indian Journal of Computer Science and Engineering, Vol.4, PP301-305.<http://www.theartling.com/text/dmtechniques/dmtechniques.htm>
3. Cortes and Pregibon. (2013). Signature based methods for data streams. Data Mining and Knowledge Discovery , Vol. 5(3), PP 167-182.
4. Cortes and Pregibon.(2011). Giga Mining. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining,PP,27-31; New York, NY: AAAI Press.
5. Cox K, Eick S, Wills G (2012). Visual data mining: Recognizing telephone calling fraud. Data Min. Knowl.Discov. 1(2): PP 225-231.
6. Ezawa, K. Norton, S. (2013). Knowledge Discovery in Telecommunication Services Data, using Bayesian Network models. Proceedings of the First International Conference on Knowledge Discovery and Data Mining, Montreal Canada. AAAI Press: Menlo Park, CA, PP 20-21.
7. Eze, U.F,Onwuegbuchulam, C.J. and Ugwuishiwu, C.H. (2016). Application of Data Mining in Telecommunication Industry, International Journal of Physical Sciences. Vol.12, (6), PP 75-78.
8. Fashoto SG, Olumide O, Sadiku, J, Gbadeyan JA (2013). Application of data mining technique for fraud detection in health insurance scheme using knee-point k-means algorithm. Aust. J. Basic Appl. Sci. 7(8): PP 140-144.
9. Fawcett,T.Provost,F. (2012). Activity monitoring: Noticing interesting Changes in Behavior. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego. ACM Press: New York, NY. PP,53-62.
10. Fawcett,T.Provost,F. (2012). Adaptive Fraud Detection.Data Mining and Knowledge Discovery. Vol. 1(3), PP,291-316.
11. Han,J.Altman,R.B. Kumar,V. Mannila, H. andPregibon, D. (2015). Emerging Scientific Applications in Data Mining. Communications of the ACM .Vol.(8),PP,54-58.
12. Han,J.Altman,R.B. Kumar,V.Mannila,H.Pregibon,D. (2015). Emerging Scientific Applications in Data Mining. Communications of the ACM, Vol. 45(8), PP, 54-58.
13. Kantardzic M (2011). Data Mining: Concepts, Models, Methods, and Algorithms. IEEE Press and John Wiley.
14. Kaplan,H.Strauss,M. and Szegedy,M. (2012). Just the Fax differentiating Voice and Fax Phone lines using call billing data.Proceedings of the Tenth Annual ACM-SIAM Symposium.
15. KolajoTaiwo and Adeyemo B Adesesan.(2012). Data Mining Technique for Predicting Telecommunication Industry Customer Churn using both Descriptive and Predictive Algorithms. Vol.3 No.2, PP,29-33.
16. Meens V (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.
17. Nora Galambos, PhD. (2014). Data Mining Methods: Applications for Institutional Research NEAIR Annual Conference Philadelphia.



18. Oyeniyi and Adeyemo.(2015). Best Practice in Churn Prediction: SAS Institute White paper; Information Systems and Development Informatics Journals. PP 32-35.
19. Oyeniyi AO, Adeyemo AB (2015). Customer Churn Analysis in Banking Sector Using Data Mining Techniques: African Journal of Computing & ICT. University of Ibadan, Nigeria. Department of Computer Science 8(3).
20. Rijmenam MV (2014). Founder of Datafloq. Think Bigger - Developing a Successful Big Data Strategy for Your Business.
21. Roset,S.Murad,U. Neumann,E. Idan,Y. Pinkas,G. (2014). Discovery of Fraud rules for Telecommunications Challenges and Solutions. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; PP,409-413, San Diego CA. New York: ACM Press.
22. Steve Lohr. (2014). Data Wrangling: for big data scientist, The New York Times.
23. Sumathi, S. and Sivanandam, S.N. (2013). Introduction to Data Mining Principles and its Application; Studies in Computational Intelligence (SCI) 29,PP15-20.
24. Udachukwu C (2013). The NCC, Telecom Operators, and Poor Services in Nigeria.
25. Weiss and Hirsh. (2013). Mining with Rarity: A unifying Framework SIGKDD Explorations.
26. Weiss, G.M., Ros, J. and Singhal, A. (2012). ANSWER: Network Monitoring using Object-Oriented rule. Proceedings of the Tenth Conference on Innovative Applications of Artificial Intelligence; 1087-1093. AAAI Press, Menlo Park, CA,
27. Zentut (2011).Data mining tools and techniques .International Journal of Scientific. Available at: www.ijser.org/paper/Data-Mining-0714.html
28. Zetut Fall. (2011). Data Analysis Project.Academics Journals; math 130 section 01.



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details