academicJournals

Vol. 12(6), pp. 74-88, 30 March, 2017
DOI: 10.5897/IJPS2016.4587
Article Number: B6B3DA363627
ISSN 1992 - 1950

http://www.academicjournals.org/IJPS

International Journal of Physical Sciences

Full Length Research Paper

Application of data mining in telecommunication industry

U. F. Eze[1], C. J. Onwuegbuchulam[2], C. H. Ugwuishiwu[3*] and S. Diala[4]

[1]Department of Information Management Technology, Federal University of Technology, Owerri, Nigeria.
[2]Department of Consumer and Sales, Etisalat Communication Company, Nigeria.
[3]Department of Computer Science, University of Nigeria, Nsukka, Nigeria.
[4]Department of Computer Science, Federal University of Technology, Owerri, Nigeria.

Received 6 December, 2016; Accepted 8 March, 2017

This paper applied a data mining model in sales and marketing department of Telecommunication Industry (TI) in Nigeria. The motivation behind the paper is as a result of competitive challenges facing most TI sales and marketing departments globally such as inability in gaining precise view of targeted data, inability to translate and formulate business question correctly and Problem of addressing data quality. The aim of this research work is to develop and implement a model that would be used to retain existing customers, attract new ones, effectively manage and allocate resources, goods and services in TI. The data mining techniques used were classification, association, sequence discovery, visualization and prediction. The tools used to implement the model were PHP, JavaScript, CSS and HTML. Telecommunication Service Providers (TSP) considered were Mobile Telephone Networks (MTN), GlobaCom (GLO), Airtel and emerging telecommunication markets (EMTs) also known as Etisalat. Three products on sales and marketing department of TI such as Airtime, Electronic Recharge (e-top up) and SIM card sales were considered. The training data used for model exploratory analysis range from 2008 to 2015 (eight years) and was collected from historical sales records of EMTs. The data were cleaned and transformed. The enhanced system was achieved through the implementation of the model which proves to be more efficient than the existing system. The model implemented was able to extract relevance information from database of TI and makes sales forecast for subsequent year. Therefore the system is recommended to be used by the TI to enhance their productivity.

Key words: Telecommunication service provider, data mining, data mining technique, model, implementation.

INTRODUCTION

Information and Communication Technology (ICT) has made business so easy that many now describe the world as a global village. According to Uduchukwu (2013), it is a widely known fact that life in the world today has been made easier through ICT. Therefore, adoption of ICT by organizations and industries will extensively improve the standard of their operations. Generally, organizations including Telecommunication industries (TI)

*Corresponding author. E-mail: chikodili.ugwuishiwu@unn.edu.ng. Tel: 08037723003.

face several challenges in analyzing their large volume of data sets in order to extract meaningful information that will enhance their decision making. Based on the aforementioned problems, it is imperative to adopt an appropriate modeling tool (data mining) to support the operations of these organizations, specifically the TI to enable them achieve their statutory goal which is 'profit making'.

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 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 (Oyeniyi and Adeyemo, 2015). The data mining techniques used in this research are classification, association, sequence discovery and prediction. Telecommunication companies utilize data mining to improve their sales and marketing operation strategies.

The aim of the work is to develop and implement data mining model in the sales and marketing department of TI to enable them discover meaningful patterns and rules that will enhance their decision making. The specific objectives of the research is to: develop a data mining model for data analysis, implement the model developed and analyze the organizational sales data (data from customers) to extract some useful information.

This work considered four telecommunication service providers (TSP) in Nigeria, such as Etisalat, MTN, AIRTEL and GLO.

All the TSP 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 these 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 (2008 to 2015). 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? etc.

Conceptual framework of the study

Figure 1 shows summary of the research work in a block diagram.

REVIEW OF RELATED WORKS

Oyeniyi and Adeyemo (2015) developed a data mining model for customer churn analysis in banking sectors using data mining techniques. 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. The data was analyzed using Waikato Environment for Knowledge Analysis (WEKA). 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.

Also, Fashoto et al. (2013) researched on application of data mining on fraud detection in National Health Insurance Scheme (NHIS) in Nigeria. This researcher applied Knee-Point K-means clustering algorithm to detect fraudulent claims in (NHIS). The aim of the model is 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 which was applied to data streams of call detail records for fraud detection in TI. This work 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. 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. First, the signature-based method cannot support fraud incidences that did not follow the profiles. Second, these systems require upgrading to update them with current frauds methods. Customer level data such as price plan and credit rating information also help in fraud analysis (Alves et al., 2006). More recent works using signatures has adopted dynamic clustering as well as deviation detection to detect fraud (Kantardzic, 2011). Clustering has problem of not producing a single output variable that leads to easy conclusions, but instead requires that you observe the output and attempt to draw your own conclusions (Cox et al., 1997).

Another method for detecting fraud exploits human pattern recognition skills. Research on Cortes and Pregibon (2001) built a suite of tool 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.

METHODOLOGY

The software methodology used in the analysis and design of the

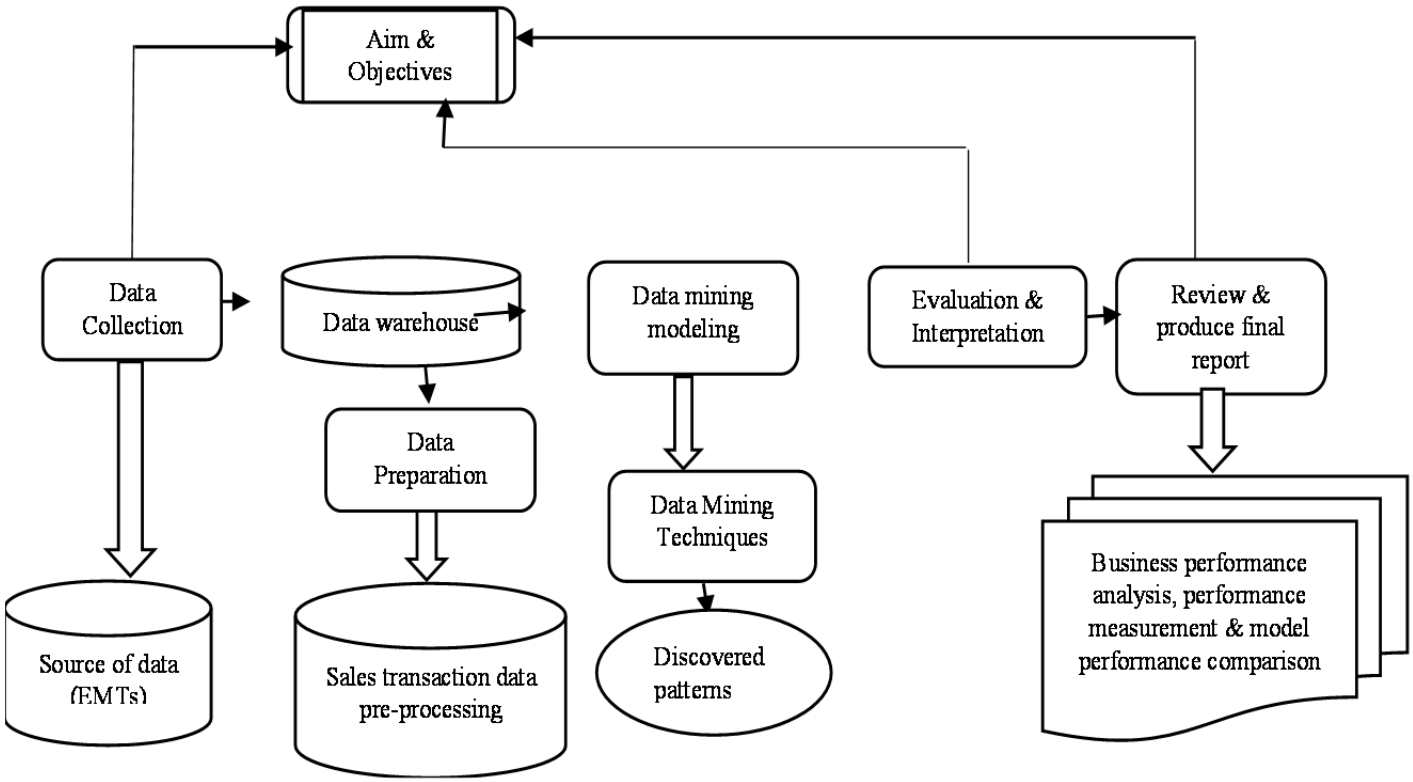


Figure 1. Conceptual framework of the study.

proposed system were Object Oriented Analysis and design (OOAD) and structural system analysis and design. Classification, association, sequence discovery, and prediction were data mining tools used in developing the data mining model. Programming tools such as PHP, HTML, CSS, and JavaScript were used to implement the model. TSPs considered in the work were MTN, GLO, EMTs. Three products considered were airtime, electronic recharge (e-top up) and SIM card sales. The training data used for data analysis ranged from 2008 to 2015 and was collected from historical sales records of EMTs.

SYSTEM ANALYSIS OF DESIGN

Here is a presentation of the system analysis and design of the existing and proposed systems, and the data mining model in form of algorithm.

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.
2) Difficulty in information retrieval and data analysis.

Data flow diagram (DFD)

Figure 2 shows the business operations of the proposed system. The conventional way of accessing and analyzing the record in repository is being replaced by proposed data mining approach.

Proposed system

Use case diagram of the system

Figure 3 presents the relationship between the main entities (actors) and the system, that is, actors and activities they perform.

System architecture

Figure 4 illustrates the operational 3–tier framework of the proposed system. This diagram shows the communication existing between the users (the interface), sales records, database storage, and data mining approach with respect to the server infrastructure that controls all the entities. Graphical user interface/presentation tier is responsible for the interaction between the system and humans. The users are the top management and the Analyst. The programming tools used to design the interface were HTML and CSS.

Server Infrastructure enables privilege users to communicate and access the dynamic web document. It includes middle tier and data tier. Middle tier connects the data tier and the presentation tier together via programming language(s). PHP and Java script were used to achieve this. Data tier stores data for the application and MySQL was used to query the database. Data mining approach involves the application of data mining model on the sales data in order to carry out analysis.

Database design, input, output

This deals with the designs of the software implementation.

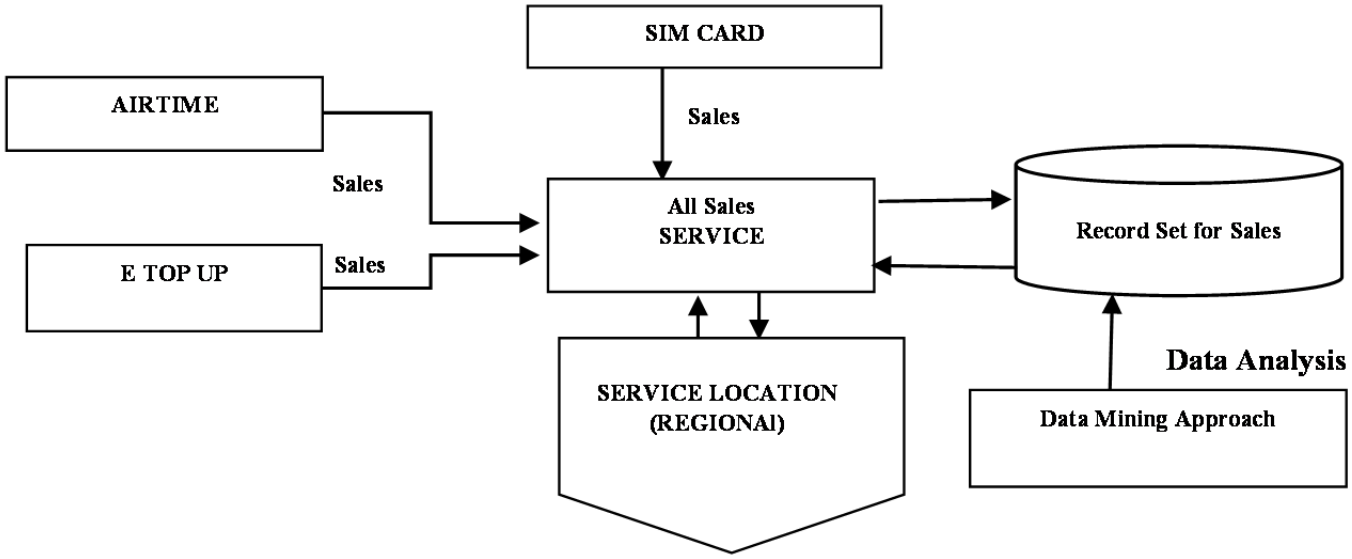


Figure 2. Business operation in the proposed system.

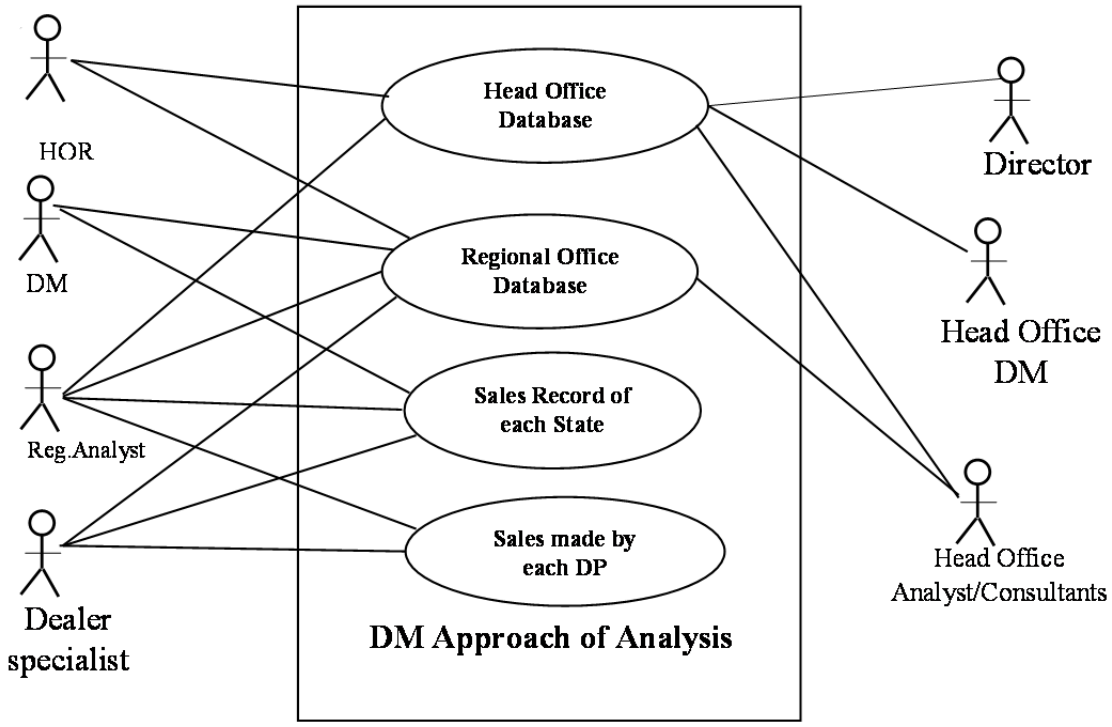


Figure 3. Use case of the proposed system.

Database design

Here shows database tables that store the data that was used in the work. Some of the tables are as shown. Table 1 stores the admin/users login details with respect to their levels of permission.

SIM card sales are stored in Table 2. Airtime and e-top-up sales are being captured by table with slight difference in field name.

Input design

Input design is used by the system to capture information from external environment e.g. New user registration template, airtime sales registration, mine airtime sales, new SIM card registration template, E-Top Up sales etc. Figure 5 is used to capture a new user. Figure 6 captures Airtime sales. E-Top Up sales has almost

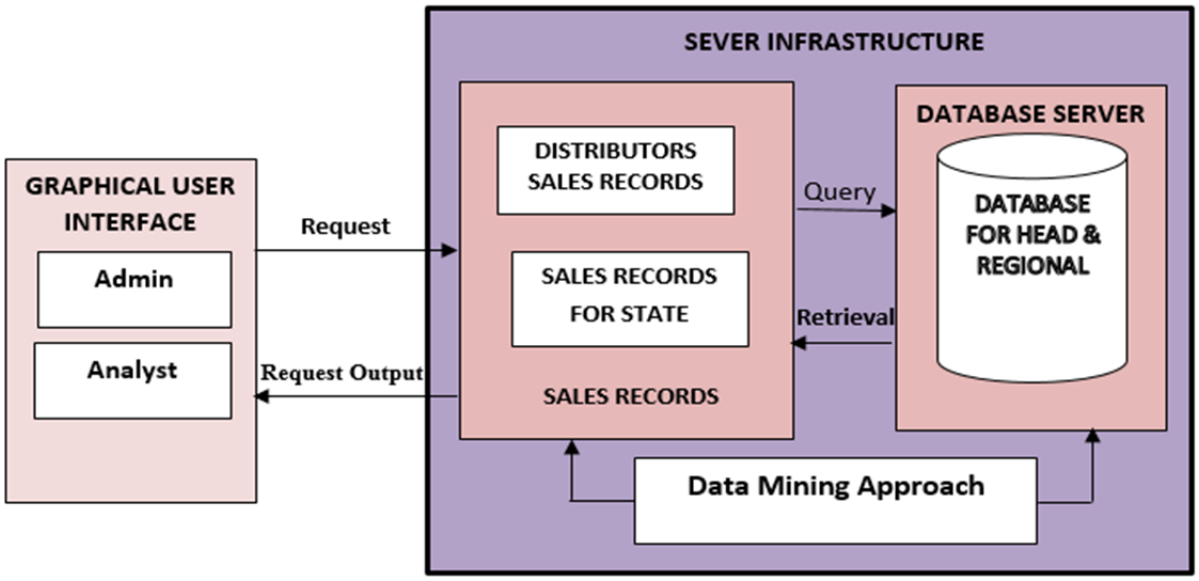


Figure 4. The proposed system architecture.

Table 1. Admin.dbf table.

S/N	Field name	Data type	Size
1	Username	Varchar	122
2	Password	Varchar	122
3	Level	Varchar	122
4	Creator	Varchar	122

Table 2. SIM_card.dbf.

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

the same template with this. Airtime sales data is analyzed using Figure 7. SIM card and e-top up have similar template. New SIM Card Registration is captured in Figure 8.

Output design

The system uses this module to convey information such as results,

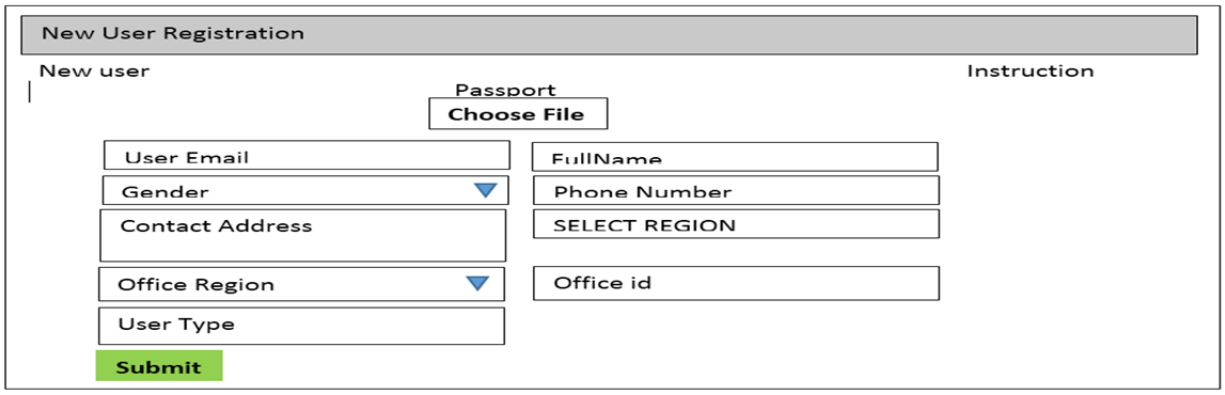


Figure 5. New user registration template.

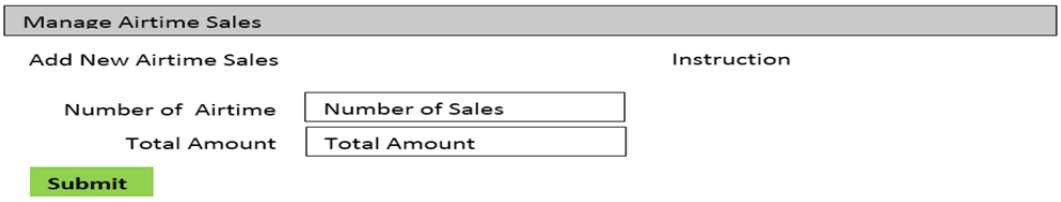


Figure 6. Airtime sales registration template.

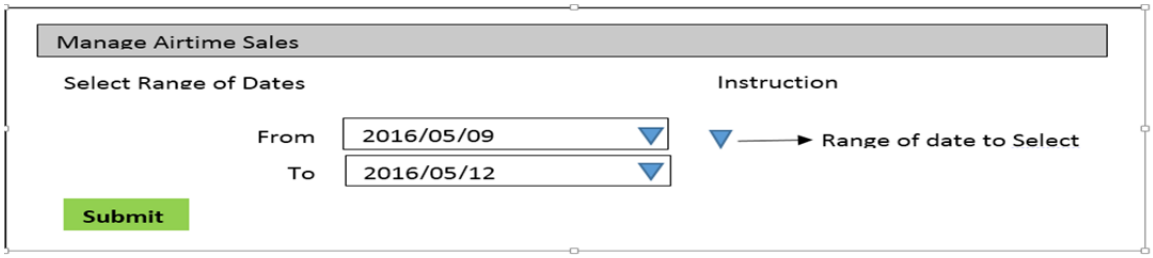


Figure 7. Mine airtime sales template.

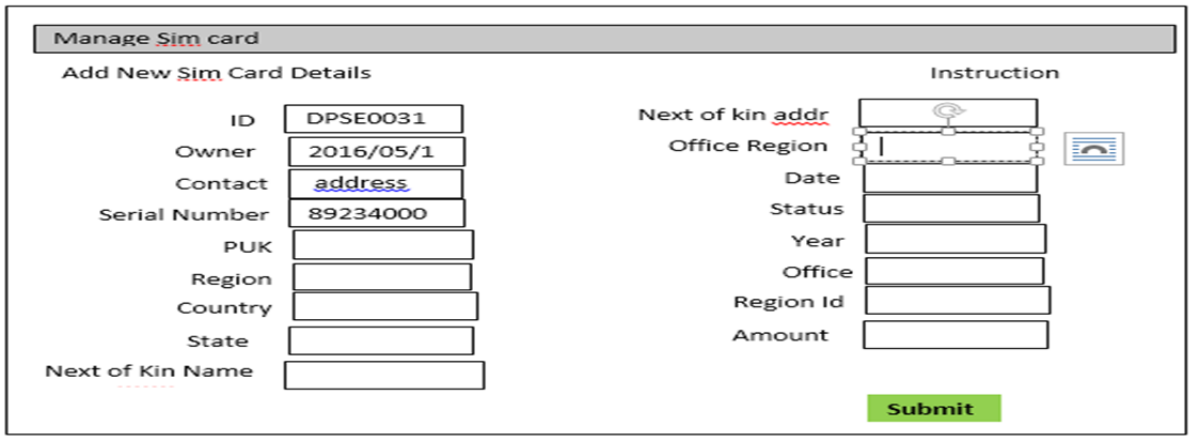
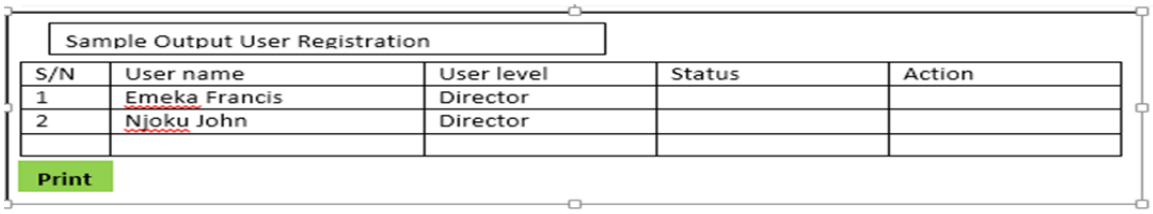


Figure 8. New SIM Card Registration Template.

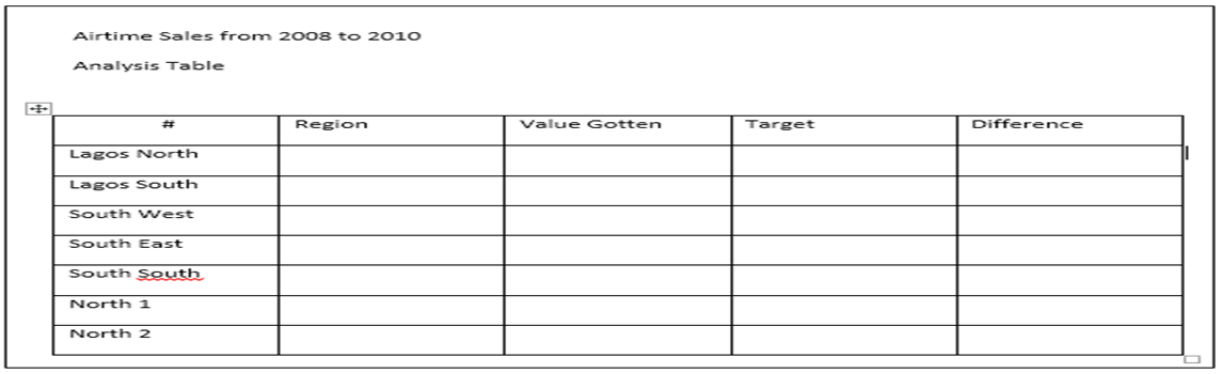


Sample Output User Registration

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

Print

Figure 9. Sample output for User Registration.

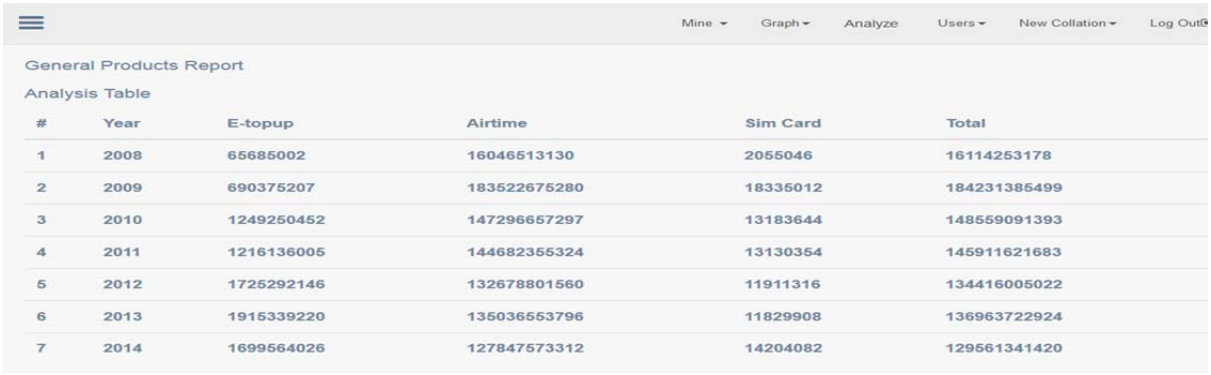


Airtime Sales from 2008 to 2010

Analysis Table

#	Region	Value Gotten	Target	Difference
Lagos North				
Lagos South				
South West				
South East				
South South				
North 1				
North 2				

Figure 10. Sample output of airtime sales mining.



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

Figure 11. General products report analysis table.

acknowledgement receipt to people etc. Figure 9 displays the output of the new user registration.

Figure 10 displays airtime sales data analysis result in a specified period. This figure has the same template with E-top up and SIM card Sales mining template.

System algorithm, flowchart and data mining model

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. These algorithms were developed as a result of knowledge gained in reviewing data mining techniques applied in this study. This forms the data mining model developed for the data analysis (Appendix).

RESULTS AND DISCUSSION

System implementation

Here shows some screenshots of the research work output and exploratory analysis of the system implementation. It shows a few illustrative analyses that can be done with the data mining model.

Figure 11 shows the general products report analysis table. This table contains analysis of all the sales made from year 2008 to 2014 in all the products considered. Figure 12 shows the Bar chart interpretation of the analysis of all the sales products shown in Figure 11.

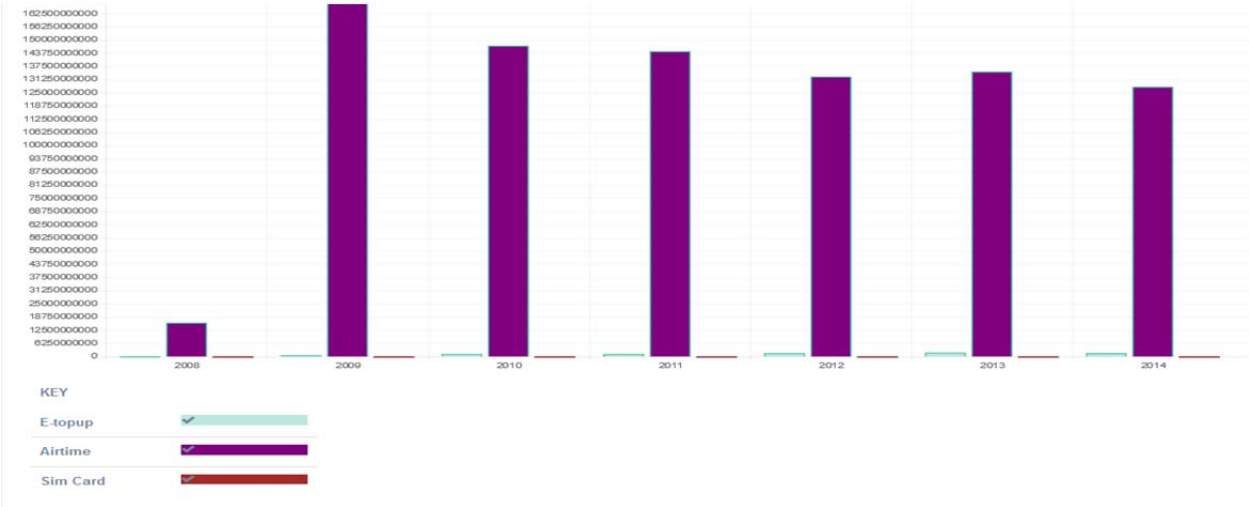


Figure 12. Bar chart interpretation of Figure 11.

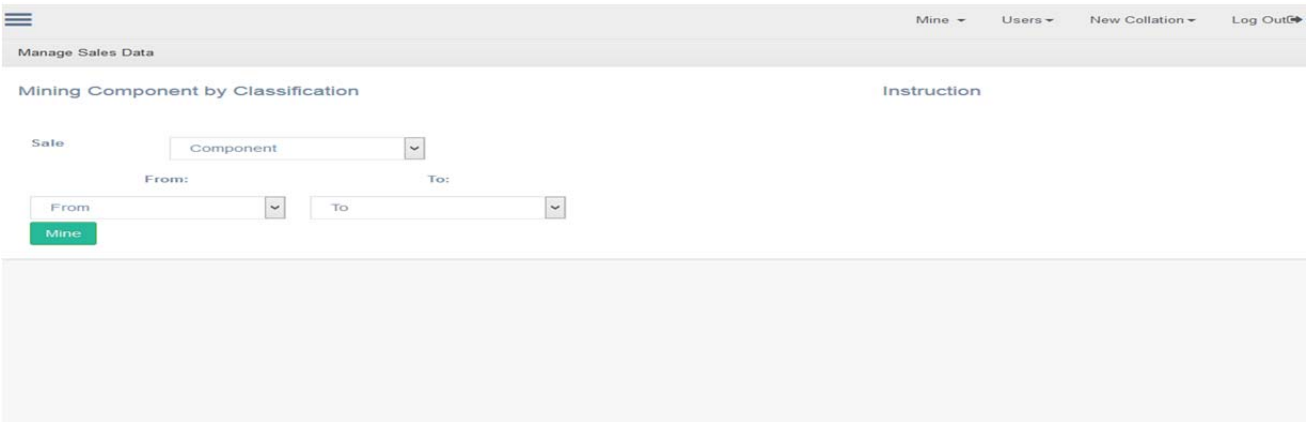


Figure 13. Mining component by classification interface.

Figure 13 is an interface where one can analyze and manage sales data for all the products considered in the work. 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), displays result which is easy to interpret. Report displayed will be analyzed and prediction would be made.

The Figure 14 shows the snapshot for mining records of airtime sales for all regions (showing region of sales, value gotten and target given) from 2008 to 2010. The Figures 15 and 16 show the interpretation using bar chart and analysis report (the airtime sales forecast for 2011) respectively.

All product data target: Figure 17 shows all the products target, sales made and difference determines whether is actualized or not with respect to years of business undertaken.

Figure 19 shows the analysis report for Figures 17 and 18 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 2010, according to the sales trends Lagos North sales will increase.

Conclusion

In this work, data mining model has been developed and implemented to enhance the operations of sales and

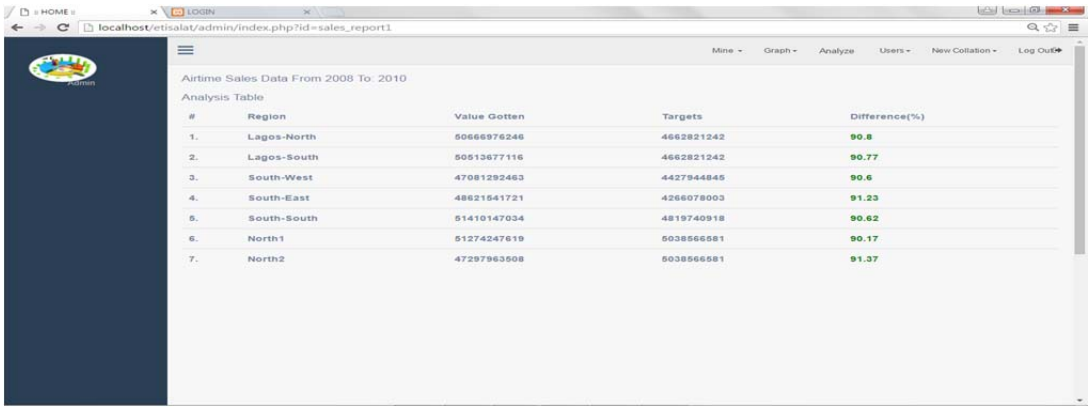


Figure 14. Result of Airtime sales department from 2008 to 2010.

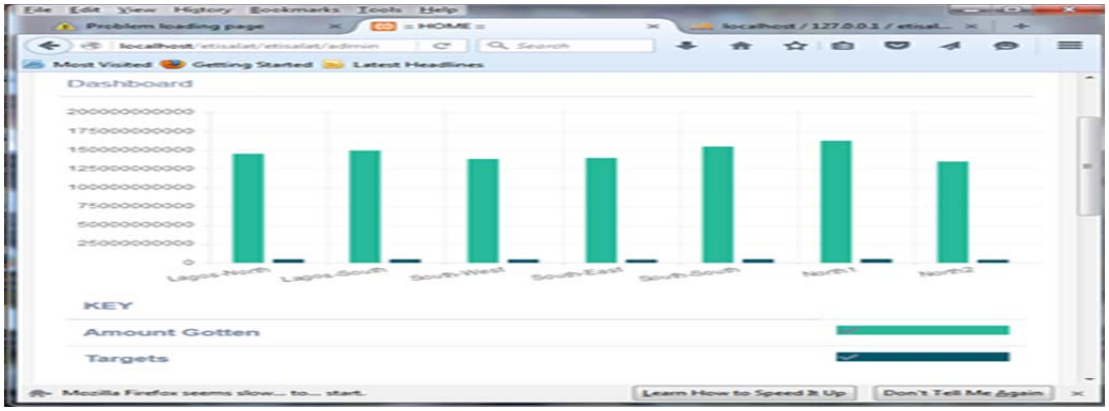


Figure 15. Bar Chart for Figure 14.

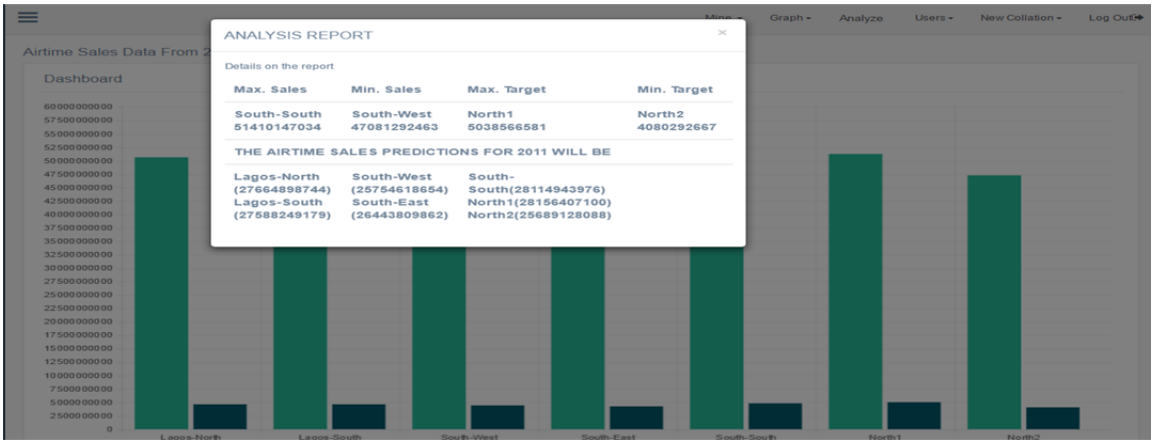


Figure 16. 2008 to 2010 Analysis report and 2011 Airtime sales forecast.

marketing departments of TSPs. This model is very robust, flexible, modular, easily understandable and has a high accuracy given a large volume of record thereby making the tedious, depth data mining analysis a simple and quick task. The model allows authorized users to gain quick precise view and predictive insights into the

#	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	2221849716	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	22538322554	21053366946
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 17. General Product/Target Report.

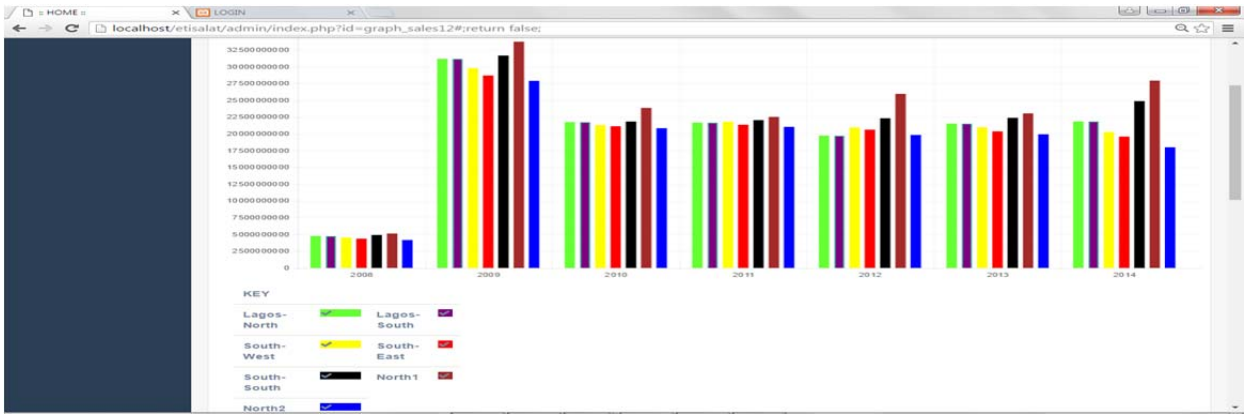


Figure 18. Bar chart interpreting Figure 17.

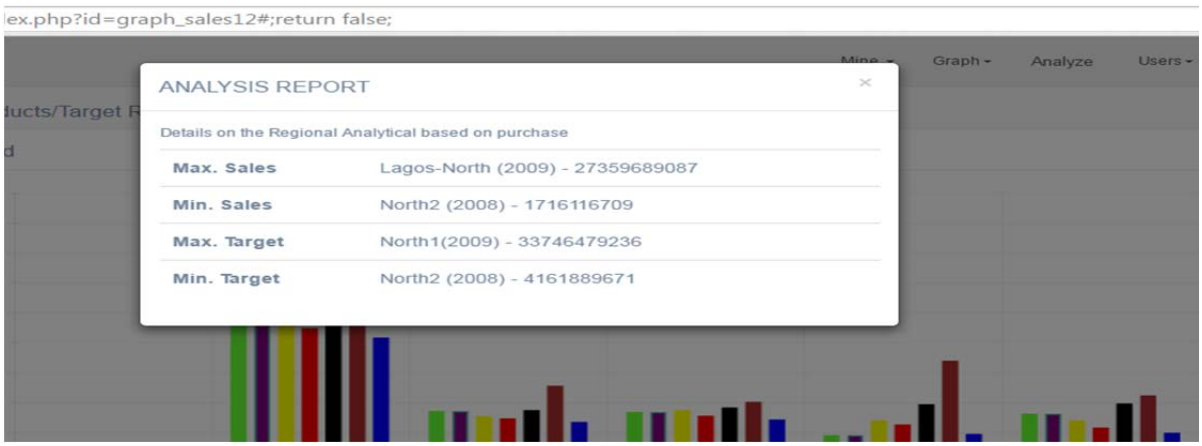


Figure 19. Analysis Report for Figures 17 and 18.

organizational large and complex datasets and reveal relationships and trends hidden in the geospatial data. Application of this model offers a great range of graphs, techniques and charts for easy description of relationships in data and knowledge acquisition which addressed the present needs in data analysis of sales

and marketing department of TSP. Therefore, based on the above listed benefit of this system, it is recommended that the new system be adopted by the TSP in Nigeria because the benefit to be realized will be outstanding.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Alves R, Ferreira P, Belo O, Lopes J, Ribeiro J, Cortesao L, 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

Cortes C, Pregibon D (2001). Signature-based methods for data streams. Data Min. Knowl. Discov. 5(3):167-182.

Cox K, Eick S, Wills G (1997). Visual data mining: Recognizing telephone calling fraud. Data Min. Knowl. Discov. 1(2):225-231.

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):140-144.

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

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.

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).

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

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

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

APPENDIX

System Pseudocode Algorithms

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
```

E-top up Sales 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
```

Airtime sales 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
```

Mine out Simcard, Airtime and Etop up Algorithm: The algorithm for mining Simcard, Airtime, E-top up have the same structure 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
```

Program flowchart

Program flowchart is the diagrammatic representation of how the entire system works.

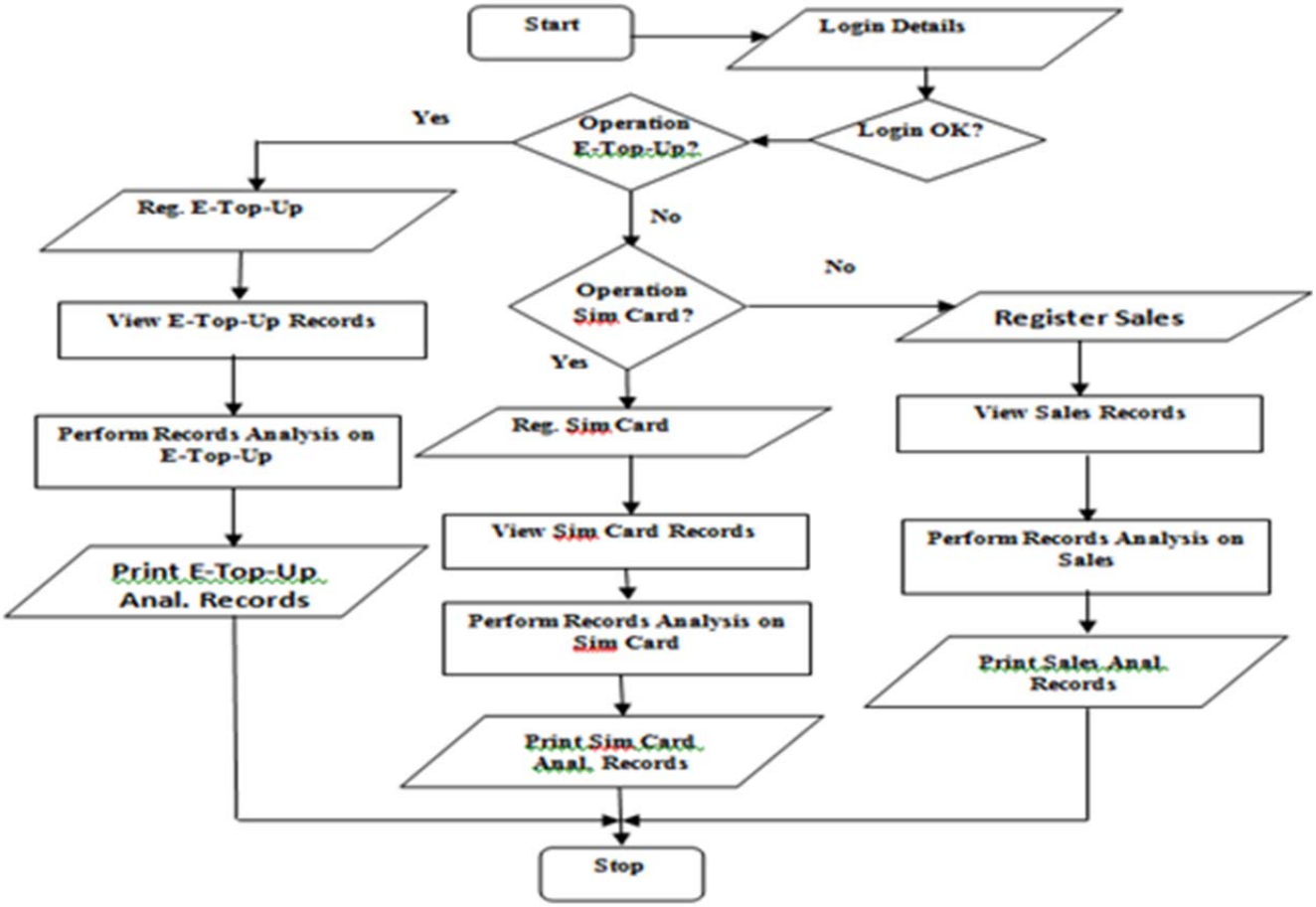


Figure 1. System flowchart.