Optimization of Data Mining for Business Intelligence Using Cobweb Clustering Algorithm- A Case Study of Nigeria Network Service Providers' Data Plan

Kelvin Ndubuisi Nnamani [*], Dr.Oliver Okechukwu [**], Chiamaka Chidimma Agu [***]

[*]*(Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*
tisaman62@yahoo.com)

[**] (*Mathematics and Computer Science Education, Enugu State University of Science and Technology, Enugu State, Nigeria*
oliverokechukwu2000@gmail.com)

[***] (*Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*
aguchiamaka46@gmail.com)

Abstract:
Business intelligence technique was employed to ascertain the efficiency of the network service providers in Nigeria: MTN, AIRTEL, ETISALAT and GLO as regards their data bundle plans. A ferventstudy was given to their data plan by interpreting the data, suggesting the loop-holes in the companies' demand and supply to facilitate more production, better performance and overall decision making strategies geared towards restoring the anomalies in the industries. WEKA software 3.8.2 JAVA incorporated was used for the data mining analysis by employing COBWEB-Algorithms. The daily, weekend, weekly and monthly data plans were data mined to help improve the quality of products towards employing price control policy to achieve a viable and productive result.From the result of the Cobweb Clustering analysis, there was 50% attributes, 22.22% instances with 60% split of full data set; 56.25% attributes and 25% instances of percentage test split emerging AIRTEL and MTN with Similar matching Link in data plan for daily and Monthly bundle which correlated with ETISALAT in Monthly data offer while GLO provides high attributes with high category of utility.

Keywords**:** *Business intelligence, Data mining, WEKASoftware, COBWEB Clustering Algorithm, Network service providers, Association Rule.*

I.INTRODUCTION

Past years, the cost for data has got reduced, while the data volume tends to increase-This is a good indication pending on consistently using the same subscription because you're not aware that MTN, AIRTEL, GLO, ETISALAT-whatever you're subscribed to, has greater data plans that you might enjoy and could be missing out on. You could be subscribing for more data than you actually need, or you could be spending much for a particular plan when there are better deals around. Theoptimal way to make a decision is by analyzing what exactly you spend your data on-Emails,videos, music, hotspot, social media apps, surfingthe webetc. To be precise, there are 1000 kilobytes (KB) in a Megabyte (MB) and 1000MB in a gigabyte (GB) **[1].**Data growth in today's world is exponential, many applications generate huge amount of data streams**[19]** at very high speed such as smart grids, sensor networks, video surveillance, financial systems, medical science data, web click streams, network data, etc.

In all nooks and crannies of the streets of Nigeria, it is a general conception that MTN Data plans(internet packages) are one of the expensive with GLO been regarded cheaper in data offer but still ETISALAT Bundles still make impact in internet access. All these are attributed to choice, financial capacity and needs **[2].**

The telecommunication industry in Nigeria, when established were supposed to provide the following services to mention a few, provide and operate public payphone, provide and operate private network links, employing cable, radio communication or satellite, within Nigeria**[3].**
In this paper the author presented a model of data analysis of four different Network service providers to a bridge the differences in customers' rejection to acceptance ratio of Network data plan towards making a judicious decision on data usage and optimizing cost-effectiveness to those networks for data mining business intelligence.

A. Business Intelligence

Business intelligence is a driven target in companies' decision making by having an in-depth overview of data mining policy since it is information-intensive and knowledge-based. The need for business intelligence is to predict the company's future improvement regarding their past performance in information delivery for better output and profit maximization. It paves way for high demand and supply.

B. Data Mining

Data mining is a process of observing data from various angles and deducing into useful information. This Different data mining techniques and software tools [4] used to achieve business intelligence are automated finding hidden relations, and predicting future events from vast amounts of data [14] in business attributes. Different data mining techniques and tools are applied and preprocess through trained data. The classification, clustering algorithm will apply to generate different classes and group of customer.

In business, data mining is the analysis of historical business activities, stored as a static data in data warehouse databases. The objective is to divulge hidden patterns and trends by employing data mining software [17] for pattern recognition algorithms to sift through large amount of data towards discovering earlier unknown strategic business information. This facilitates use of data mining for business tasks to perform market analysis to identify new product bundles.

Data mining is applied in studying huge volumes of data and gathers some useful information from them. Telecommunication industries and Technology have grown widely in communication technologies like internet, network and computer to improve in their business using data mining in ascertaining customer preferences, the pricing and the positioning of the products, the satisfaction of the customers along with the corporate profits [18].

C. WEKA Software Algorithms

The COBWEB Algorithm was developed by machine learning researchers in the 1980 for clustering objects in an object attribute dataset. The COBWEB Algorithm yields a clustering dendrogram called classification tree that characterizes each cluster with a probabilistic description. The COBWEB algorithm operates based on the category utility function that measures clustering quality [5].

WEKA'S support for clustering tasks is not as costly as its support for Classification and regression, but it has more techniques for Clustering than for association rule mining, which has up to this point not taken into consideration. Clusteringis a main task of explorative data mining, and a common technique for statistical data analysis used in many fields including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics [15]. Nevertheless, it does contain an implementation of the most well-known algorithm in this area, as well as a few other ones. These methods can be accessed via the "Associate" panel in the explorer [6].

COBWEB- Algorithm compares the host, adding a new leaf, merging the two best hosts, splitting the best host when considering where to place a new instance. Parameters for classes implementing cobweb &classist are acuity and cutoff [7, 13]. A benefit of COBWEB is that it can adjust the number of clusters in a partition [20], without the user specifying this input parameter. A disadvantage of COBWEB is that it information allows user from different viewpoints in attributes, categories, summaries and angles to identify relationship.

assumes categorical attributes are independent and it may assume correlated attributes are independent.

II. Literature Review

Business Intelligence (BI) is a concept of applying a set of technology to convert data into meaningful information. BI methods involve information retrieval, data mining (DM), statistical analysis as well as data visualization. This is done by applying different formats on large data sets for business knowledge. The interwoven of DM and BI is applied in data processing and analysis to help ease the workload for the users for better findings. The Data mining (DM) searches through data using various algorithms to discover patterns and correlations in a database [16] for effective business decision. In his paper, Data mining and business intelligence are established to identify high profit, high-value and low-risk customers.

The internet is a set of inter-connected networks [8] and includes several millions computers attached to these networks. Internet utilities can be regarded as consisting of four broad classes. One class consists of communication utilities such as E-mail, internet relay chat (IRC) or web chat, mailing lists and newsgroups. E-mail provides a simple message transfer facility. IRC or web chat enables different users to interact with each other on a real-time basis.

There was emphasis that Broadband Internet consumers are spending more time using internet applications for activities such as streaming video and online gaming, which require a high-quality network [9].Simultaneously, internet penetration is reaching a saturation point in mature markets, especially in fixed access networks, so service providers cannot depend chiefly on new subscribers as a source of revenue generation.

In the inference that business intelligence provides a platform in which information and knowledge is based to improve business operation, [10]Comparative advantage is employed in explaining customers' satisfaction beneficiary to information network providers and those responsible for decision making in related organizations

Business intelligence tools are very useful for large companies. Asexample, the Telecom Market changes. Companies are looking for better ways to take advantage in this difficult competition. The analysis of what is inside the data from systems like traffic, sales, online, accounting become more important. The data from operational systems contain information about the client and how to keep this client, how to offer solutions for giving a better price or a better service. Also, gives ideas to decision makers on how to improve communication to client, how to improve network qualities and so on [11].

It was estimated that GSM remains the most dominant technology type for accessing the internet for

subscribers **[12].**InQ4 2017, active GSM internet subscription stood at 98,391.456, representing a 5.8% growth from the previous quarter's 92,975,682 active subscribers. GSM internet subscriptions were on an increasing trend throughout the year 2017, recording a 7.1% growth in subscriptions between December 2016 and December 2017.In the quarter under review (Q4 2017) MTN retained the highest share of all

The COBWEB Algorithm is a hierarchical Clustering algorithm. In the paper, the COBWEB algorithm constructs a classification tree incrementally by **[5]** inserting the objects into the classification tree one by one. When inserting an object into the classification tree, the COBWEB algorithm

III. Data Collection and Methods.

The data collection of this study was based on the information published by Zoto Nigeria on May 22, 2017 based on the public outlook about different data bundles plans on the four telecommunication Network service providers in Nigeria. The Data mining analysis were used to build business intelligence on daily, weekend, weekly and monthly data bundle plan using preprocessing, COBWEB-clustering Algorithms and Filtered association Rule: there was filtering in the dataset. The different data sets were employed using WEKA 3.8.2 Java-based software. The Data Collected was

\

GSM active internet subscribers, capturing 36.6% of subscriptions (36,069,597), followed by GLO at 27.4% of subscription (26,997,817), Airtel at 24.3% of subscription (23,985,203), and 9Mobile (EMTS) at 11.5% of subscription (11,338,839).

traverses the tree top-down starting from the root node and find the best position to insert a new object by calculating the category utility(CU) function using four operations such as create, insert, merge, split and finally incrementally organizes records into a tree.

loaded in CSV Format before saving in Desktop for Data Mining Process.

Internet users were also interviewed on the most preferred Network on Internet Networks for web browsing on GSM/WCDMA Networks, Email creation or access, YouTube, Video streaming. The duration of Data bundles for each data bundles plan for MTN, AIRTEL, GLO, and ETISALAT were compared based on Data bundle load, price and the data usage/web surfing intervals. The data in Table 1 showcased the four Network providers in Nigeria with their daily, weekend, weekly and Monthly internet Data bundle plans, allocating different prices to each Data bundle.

TABLE 1:DAILY, WEEKEND, WEEKLY and MONTHLY (3GB & ABOVE) INTERNET DATA BUNDLE PLAN.

SERVICE PROVIDER	DAILY PRICE	DAILY BUNDLE DATA	WEEKEND PRICE	WEEKEND BUNDLE DATA	WEEKLY PRICE	WEEKLY BUNDLE DATA	MONTHLY PRICE	MONTHLY DATA BUNDLE FROM 3GB & ABOVE
MTN	₦500	150MB	₦300	3GB	₦500	750MB	₦2000	3.5GB
AIRTEL	₦500	150MB	₦100	1GB	₦300	80MB	₦2000	3.5GB
GLO	₦100	10MB	₦500	3GB	₦500	1.6GB	₦1200	3.2GB
ETISALAT	₦100	10MB	₦500	1GB	₦200	200MB	₦2000	3.5GB

IV. Results and Discussion

A. RESULTS

The result of the data presented below was data mined using WEKA 3.8.2 Java-based software, employing preprocessing, Cobweb-Algorithms, Filtered Associator Rule to make judicious decision in partitioning Internet Data bundles for MTN,AIRTEL,GLO and ETISALAT. The analysis of the results was shown in Fig 1,2,3,4, 5 and 6 shown below.

Fig1, shows a processing of data plan having 4 instances, 9 attributes, Sum of24 distinct, and the sum of weights as 4.The WEKA Explorer depicted the output in the form of a bar chart .Fig 2 shows full visualization of Fig1, showing all the attributes of the processing rule
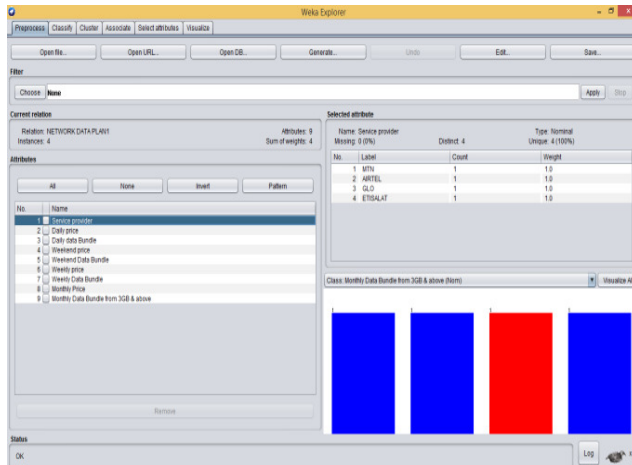
Fig 1 Preprocessing Rule of Data Plan
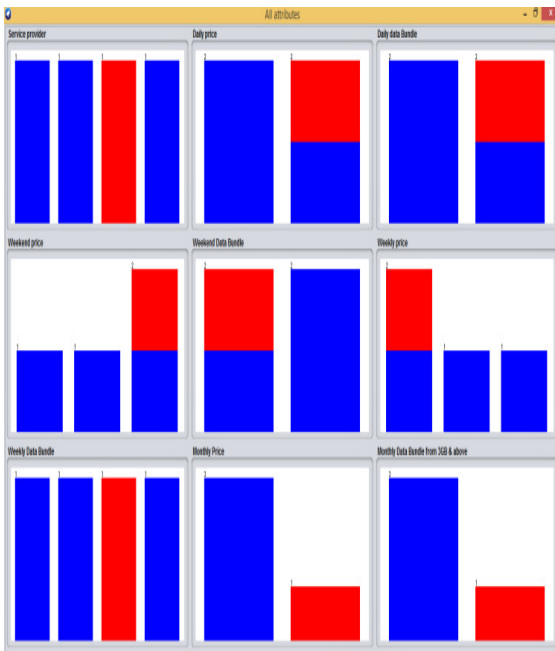
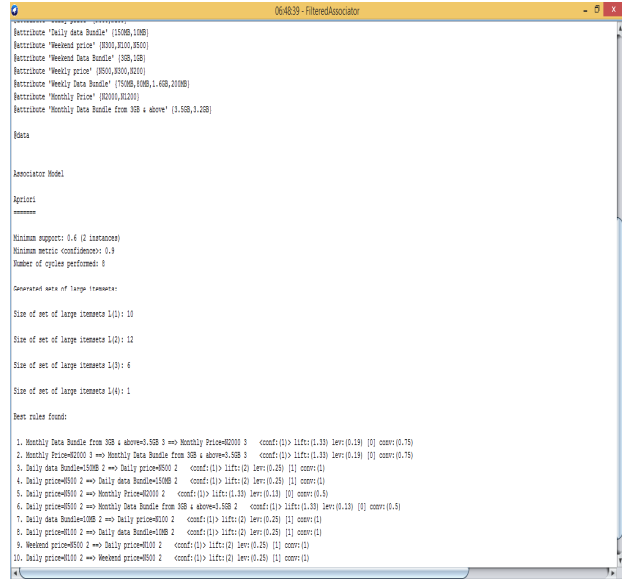Fig 2 Full Preprocessing of Data Visualization

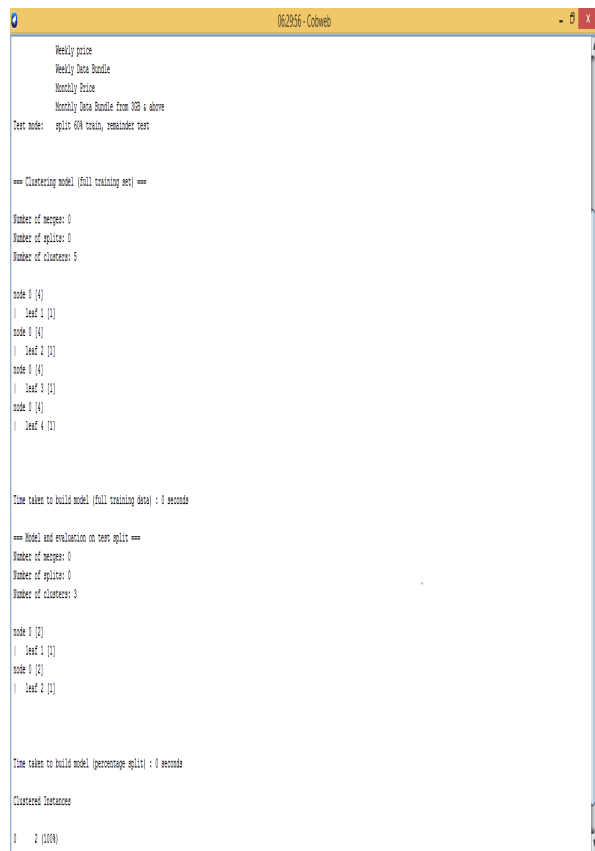Fig3 Filtered Association Model Viewed in a Separate Window

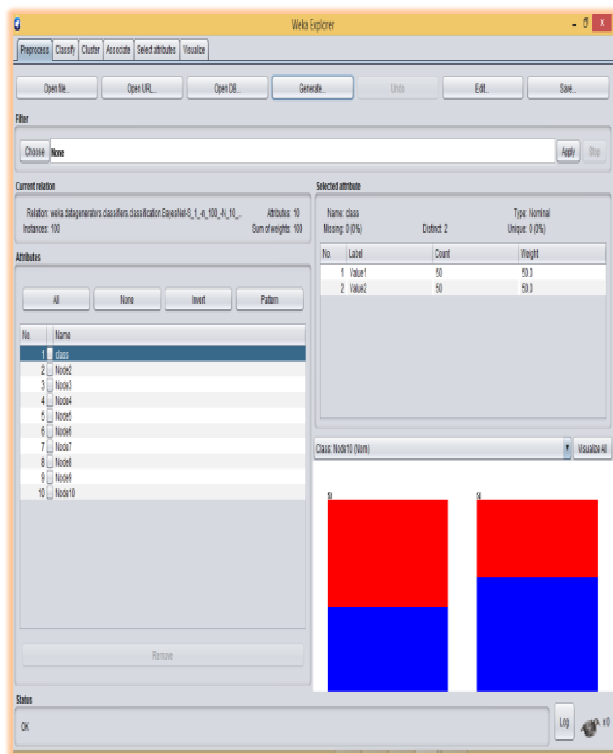Fig 4 COBWEB Clustering Algorithms for Data Analysis in a separate Window

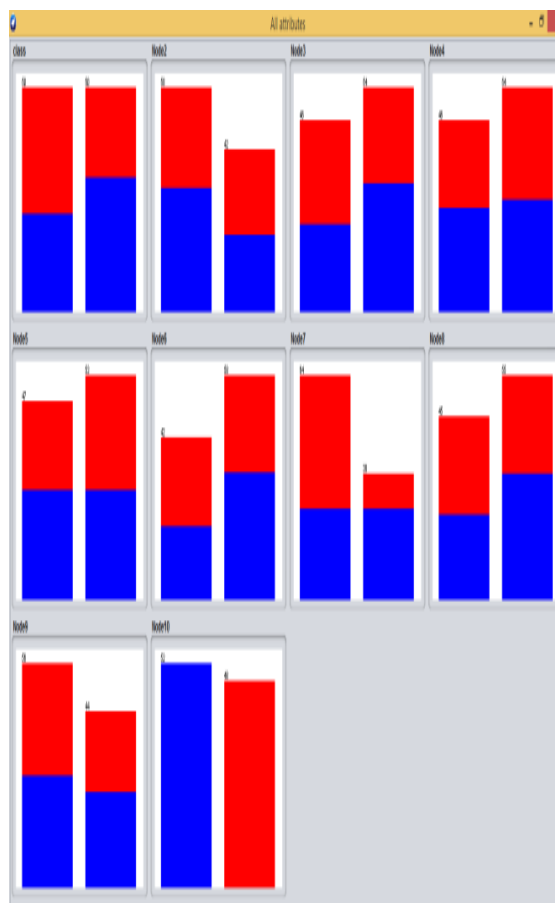Fig5 Bayes Network Generator for Preprocessing Classification into Node

Fig 6 Bayes Network Classification Full Visualization of attributes

B. DISCUSSION

The results of the data mining show: 50%attributes and 22.22%instances using COBWEB Clustering Algorithms of 60% split of full data set; 56.25% attributes and 25% instances of percentage test split. The result shows that there was 0 numbers of merges, 27.78% number of clusters in the full training data set and 18.75% number of Clusters in the test split emerging 100% instances in both cluster 0 and 2. The result indicate that AIRTEL and MTN have similar matching link in data plan for daily and Monthly bundle offer with ETISALAT correlating with them in Monthly data offer; While GLO provides high attributes in clustering and Tree visualization with high category utility as indicated by the component Bar chart in the preprocessing rule at weekly and monthly data offer. The summary of the result is shown in table 2, Fig 7 and 8 below.

TABLE 2
SUMMARY of the RESULTS of the DATASET USING WEKA RULE

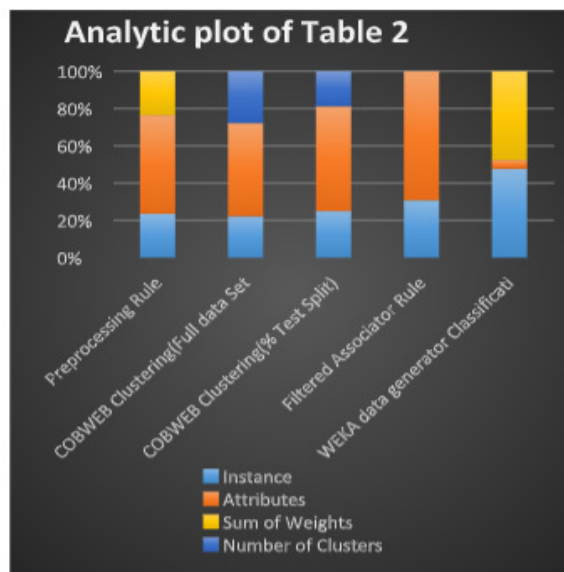| WEKA BASE RULE	Instance (%)	Attributes (%)	Sum of distinct (%)	Sum of weights (%)	Number of Clusters (%)
Preprocessing rule	9.75%	21.95%	58.54%	9.75%	___
COBWEB Clustering(full data set)	22.22%	50%	___	___	27.78%
COBWEB Clustering(% test split)	25%	56.25%	___	___	18.75%
Filtered Associator rule	26.67%	60%	13.33%	___	___
WEKA data generator classification	43.48%	4.35%	8.70%	43.48%	___
Mean Accuracy (%)	25.42%	38.51%	26.86%	26.62%	23.27%

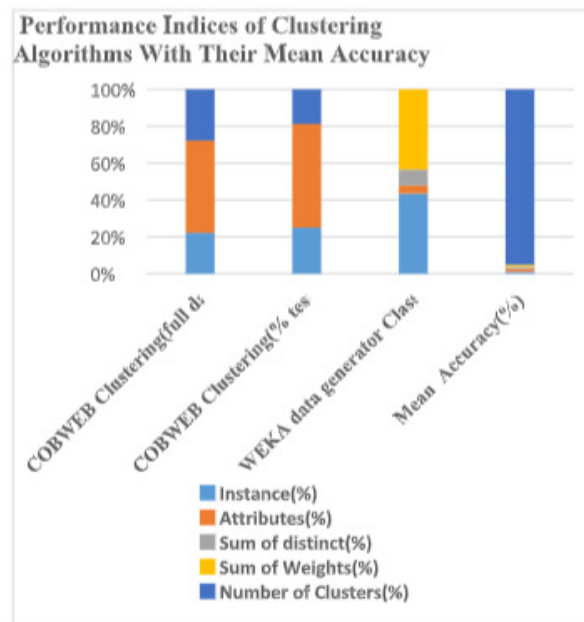Fig 7 Analysis Plot of The Results of WEKA Base Rule in Table 2.

Fig 8. Summary of Data mining Clustering Algorithm and their Mean Accuracy

V.CONCLUSION

This study was carried out for data mining of Data bundles of Four Major network providers in Nigeria using COBWEB clustering Algorithm and Association rule to make business decision through business intelligence strategy. The prediction of the models using COBWEB clustering has it that AIRTEL and MTN lead on daily and weekend Internet Data Bundle plans while GLO leads on weekly and monthly offering more data bundles at cheaper rate, even though data traffic and Environment contribute to quick exhaustion of data. In summary, GLO has better performance and Tolerance to subscribers' data Bundle plan at minimum cost and greater sales of Data bundles

REFERENCES

[1] *ZotoNigeria (2017).https://blog.zoto.com.ng/cheap mtn-airtel-gloworld.com.*

[2] *MCmedal001 (2017), "MTN Data Plans, Subscription Codes for phone andcomputer2018":https://www.geekish.ng/mtn-data-plans-subscription-codes*

[3] *Ibukun Afolabi and Fisayo Adegoke(2014), "Analysis of Customer satisfaction for competitive advantage using Clustering and Association Rules", International Journal of computer Science and Engineering (IJCSE), Vol.3, Issue 2.pp 141-150.*

[7] *S. Hanumanthsastry and Prof M.S PrasadaBabu(2013), "Analysis and Prediction of Sales Data in Sap-Erp System using Clustering Algorithms", International Journal of Computational Science and Information Technology (IJCSITY), Vol. 1, No.4.pp. 95-109.*

[8] *Badrinarayan Shankari Pawar and Ramesh Sharda (1997), "Obtaining Business Intelligence on the Internet", Long Range planning, Vol. 30, No.1, pp 110-121.*

[9] *Dave Caputo(2011), "The Fundamentals of Network Business Intelligence": info @ sandvine.com.*

[13] *Nachiketa Sahoo (2009), "Three Essays on Enterprise Information System Mining for Business Intelligence", A Doctoral Dissertation Submitted to Heinz College, Carnegie Mellon University. 127 pages.*

[14] *Drew Bentley (2017), "Business Intelligence and Analytics", Library Press. 317 pages.*

[15] *Narendra Sharma et al (2012), "Comparison the various Clustering Algorithms of Weka Tools", International Journal of Emerging Technology and Advanced Engineering, Vol.2, Issue 5, pp.73-80.*

[16] *Arti J. Ugale, P.S. Mohod (2015), "Business Intelligence Using Data Mining Techniques on Very Large Datasets", International Journal of Science and Research(IJSR),Vol.4,Issue 6, pp. 2932-2937.*

[17] *Ferero Dermino& Klawi Fortingo (2015), "What is Data Mining Methods with Different Group of Clustering and Classification?",*

[4] *Milan Milovic and Boris Milovic(2012), "Prediction and Precision making in health care using data mining", International Journal of Public Health science (IJPHS), Vol. 1, No.2,pp. 69-78.*

[5] *V. Kanageshwari and Dr. A Pethalakshmi (2007),"A Novel Approach of clustering Using COBWEB", International Journal of Information Technology (IJIT), Vol. 3 Issue 3, pp.37-42.*

[6] *Ian Witten, Holmes, et al(2014), "WEKA Data mining software": update. Research gate, Vol. 11, issue 1, pp. 10-18.*

[10] *Ibukun Fatudimu et al (2014), "An Empirical Analysis of Customer Satisfaction with Mobile network service for comparative Business Advantage", International Journal of Research in Business Management, Vol.2, Issue 3, pp. 25-34.*

[11] *Monica Lia(2015), "Customer Data Analysis Model using Business Intelligence Tools in Telecommunication Companies", Database Systems Journal, Vol .VI, No.2, pp. 39-47.*

[12] *NationalBureau of Statistics (2017),"Telecoms Data: Active Voice and Internet per State, porting and Tariff information", 97 pages American Journal of Mobile Systems, Applications and Services, Vol.1, No.2, pp. 140-151.*

[18] *Brojo Kishore Mishra et al (2016), "Business Intelligence Using Data Mining Techniques and Business Analytics", 5th International Conference on System Modeling & Advancement in Research Trends, pp. 84-89.*

[19] *Umesh Kokate et al (2018), "Data Stream Clustering Techniques, Applications, and Models: Comparative Analysis and Discussion", Big Data and Cognitive Computing, Vol.2, No.32, 30 pages.*

[20] *Vipin Kumar (2014),"Data Clustering Algorithms and Applications", a Chapman & Hall Book, Newyork. Pp.288-289.*