World Applied Sciences Journal 33 (12): 1846-1851, 2015
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2015.33.12.15629

Data Mining: a Strategy for Economic and Educational Development in West Africa

[1]Odachi Gabriel Nwabuonu and [2]Nwosu John Nwachukwu

[1]Department of Computer Science, Nwafor Orizu College of Education, Nsugbe, Anambra State, Nigeria
[2]Department of Computer Science, Federal Polytechnic Oko, Anambra State Nigeria

Abstract: Data mining is a promising and relatively new technology. It is a powerful technology with great potential for economic and educational growth. Web research carried out showed that not much is known about data mining in developing regions of which West Africa is one. This paper explores the concept, processes, technique /Methodology of data mining. It also looks at the benefits associated with data mining application to data. The result of this research will help people in West Africa to apply data mining techniques to their enormous data in order to obtain useful and hidden knowledge from them. This knowledge can be used and applied in their various areas of human endeavour to obtain maximum benefit.

Key words: Data mining · Methodology · Potential · Processed and concept

INTRODUCTION

The development of information technology has generated large amount of databases and huge date in various areas [1]. The amount of raw data stored in the data bases are exploding. Databases are now measured in gigabytes and terabytes. A terabyte is equivalent to about 2 million books. Raw data of this magnitude by themselves do not provide much information. Data mining is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data [2]. Sawyer and Williams (2007) [3] also define data mining as the computer-assisted process of sifting through and analyzing vast amount of data in order to extract hidden patterns and meaning and to discover new knowledge. This new knowledge is applied to the customers and markets to guide the marketing, investment and management strategies. We can use data mining to discover things we didn't know or what is going to happen next (prediction).

For instance, one Midwest grocery chain used data mining capacity or oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tend to buy beer. Further analysis showed that these shoppers typically did their widely grocery shopping on Saturdays. On Thursdays, they only bought a few items. The grocery management could use this newly discovered information in many ways to increase the companies' revenue. For example, they could move the beer display closer to the diaper display.

They could also increase the quantity of those goods bought on Saturdays. Data mining is thus, the process of extracting value from database (Singh and Chauhan, 2009) [4]. In today's fiercely completive business environment companies in west Africa environment need to rapidly turn to data mining to obtain useful information - information that can be used to increase revenue and cut down production cost.

Tertiary institutions can also apply data mining for effective and profitable management.

Higher education institutions are nucleus of research and future development acting in a competitive environment, with the prerequisite mission to generate, accumulate and share knowledge. Universities require an important amount of significant knowledge mined from its past and current data sets using special methods and processes. The ways in which information and knowledge are represented and delivered to the university managers are in a continuous transformation due to the involvement of the information and communication technologies in all the academic processes. Higher education institutions have long been interested in predicting the paths of students and alumni, thus identifying which students will join particular course programs and which students will require assistance in order to graduate. Another important preoccupation is the academic failure among students which has long fuelled a large number of debates [5].

Corresponding Author: Odachi Gabriel Nwabuonu, Department of Computer Science,
Nwafor Orizu College of Education, Nsugbe, Anambra State, Nigeria.

Data Mining Processes: Data Mining Processes are the steps or procedures taken to carry out data mining. These steps are as follows:

Data Source/Data Collection: Data may come from a number of sources such as point of sale transactions in files, data bases of all sorts, newswires, internet etc.

Data Fusion and Cleansing: Data from various sources are fused together and then put through a process called data cleansing or scrubbing. This is necessary because the acquired data may be of poor quality, full of errors and in consistence. So, for data mining to produce accurate result the data has to be scrubbed. That is removing the errors and checking for consistency of format.

Data/Meta-Data: The scrubbing or cleansing process yields both cleaned-up a nd variation of it known as meta-data. Meta-data is very important and necessary for the understanding of data stored in data ware-house. Meta data shows the origin of data, the changes it has undergone. Meta-data also describes the contents of the data warehouse.

The summary information of meta-data makes it more useful than the cleaned but un-integrated, un-summarized data.

Data Transport to the Data Warehouse: At this stage both data and meta-data are sent to the data warehouse, which is a special database of cleaned up and meta-data.

The data warehouse is stored on disk using storage technology such as Redundant Arrays of Independent Disks (RAID). For large data such as 500 gigabytes, massively parallel processing Computers are needed.

Analysis of Warehouse Data: Warehouse data are mined or analyzed. Patterns are searched for and interpretation and results given out for use or deployment.

Data Mining Methodology /Technique: There are many methods/techniques available for data mining practitioners. They include:

Association: Association is one of the best mining methods. In association, a pattern is discovered based on a relationship between items in the same transaction. In market basket analysis, association technique is used to identify asset of products that customer frequently purchased together. In business, retailers use association method to research customer's buying habits. Through this research, using available data, they might find out that customers always buy snacks when they buy soft drinks. Therefore, they can put soft drinks and snacks next to each other to save time for customers and increase sales. They might also discover that certain goods are bought more at a particular season than the other and therefore provide those goods in large quantity at their Favourite Seasons.

Classification: Classification method makes use of certain techniques such as neural network, decision tree etc.

Neural Network: A neural network is an artificial intelligence system which is capable of learning because it is patterned after human brain. It is composed of a large number of highly interconnected processing elements (Neurons) working in unison to solve specific problems. Artificial Neural network, like people, learn by example.

An artificial neural network is configured for a specific application, such as pattern recognition or data classification through a learning process. It is made up of 3 layers such as input node, hidden layer node and output node.

In data mining, neural networks can be used to model complex relationships between inputs and outputs or to find patterns in data. Neural networks are trained to store, recognize and associatively retrieve patterns or database entries to solve combinatorial optimization Problems. In data warehouses, neural networks are just one of the tools used in data mining. Neural networks are useful in providing information on associations, classifications, clusters and forecasting.

The back propagation algorithm performs learning on a feed-forward neural network.

Feed Forward Neural Network: One of the simplest feed forward neural networks (FFNN) is made up of three layers- an input layer, hidden layer and output layer. Fig. 3 below. In each layer there are one or more Processing elements. Processing element is meant to simulate the neurons in the brain and this is why they are referred to as neurons or nodes.

A processing element receives inputs from either the outside world or the previous layer. There are connections between the processing elements in each layer that have a weight (Parameter) associated with them. This weight is adjusted during training. Informant only travels in the forward direction through the network. There is no feedback loops.

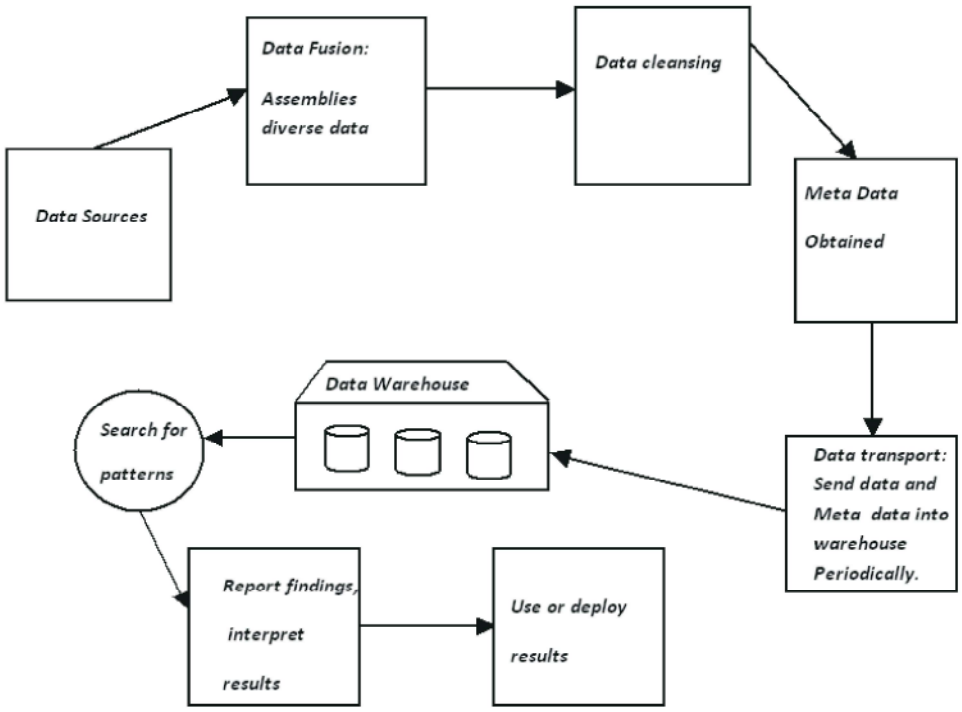

Fig. 1: Data Mining Processes

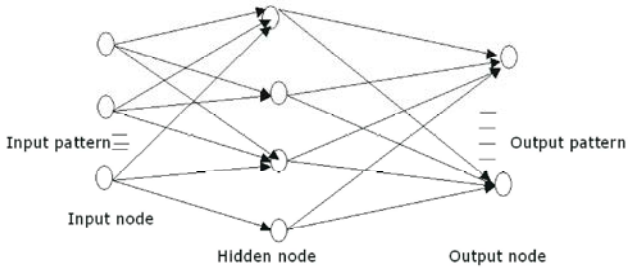

Fig. 2: Layers of Neural Network

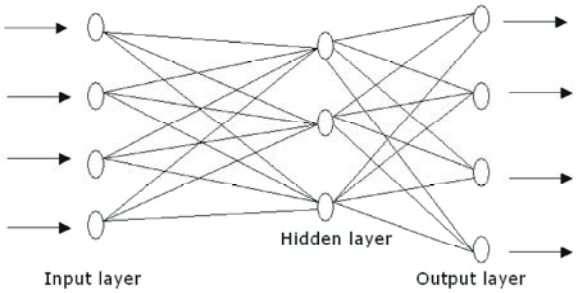

Fig. 3: Feed forward neural network

The Back Propagation Algorithm: Back propagation (propagation error) is a common method of teaching artificial neural network how to perform a given task. The back propagation algorithm is used in layered feed forward artificial neural networks. As the artificial neurons are organized in layers, signals are sent forward and then the errors are propagated backwards. The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute and then the error is calculated. The error is the difference between actual and expected results. The idea of the back propagation algorithm is to reduce this error, until the artificial neural network (ANN) learns the training data.

The simplified process for training a Feed Forward Neural Network (FFNN) is as follow:

- Input data is presented to the network and propagate the network until it reaches the output layer. This forward process produces a predicted output.
- The predicted output in subtracted from the actual output and an error valuee for the network is calculated.
- The neural network then uses supervised learning, which in most cases is back propagation, to train the network. Back propagation is a learning algorithm for weight. It starts with the weights between the output layer processing elements and the last hidden layer processing elements and works backwards through the network.
- Once back propagation is done with, the forward process starts again and this cycle is continued until the error between predicted and actual outputs is minimized.

Summary of the Technique: Present a training sample to the neural network

Compare the networks output to the desired output from that sample. Calculate the error in each output neuron.

For each neuron, calculate what the output should have been and a scaling factor, how much lower or higher the output must be adjusted to match the desired output. This is the local error.

Adjust the weights of each neuron to lower the local error.

Assign "blame" for the local error to neurons at the previous level, giving greater responsibility to neurons connected by stronger weights.

Repeat the steps above on the neurons at the previous level, using each ones "blame" as its error.

Decision Trees: Decision tree is one of the most used data mining methods because its model is easy to understand by the users.

It breaks down a dataset into smaller and smaller subsets, while at the same time an associated decision tree is incrementally developed. The final result is a decision tree with decision nodes and leaf nodes. For instance, in the diagram below, a decision root is the "outlook", which has tree branches sunny, overcast and Rainy, then the leaf node include play or not to play. See the diagram below.

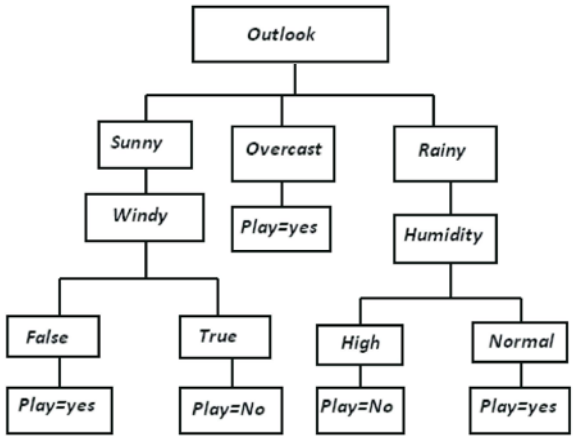

Fig. 4: Decision tree

A decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes.

R1: If (outlook = sunny) and (windy=false)
Then play=yes
R2: If (Outlook =sunny) and windy =True)
Then play = No
R3: If (outlook = Overcast) Then play=yes
R4: If (outlook = Rainy) and (Humidity = High) Then Play=No
R5: If (Outlook = Rainy) and (Humidity = normal)
Then play =yes

Decision trees are useful because they allow decision process to be unveiled from the data. Each branch in the tree represents a decision made on a particular attribute. The algorithm automatically determines which attributes are most important [6].

Decision trees are commonly used in Operations Research, Specifically in decision analysis to help identify a strategy most likely to reach a goal [7].

Clustering: It is a data mining technique that makes meaningful cluster of objects which have similar characteristics using automatic technique. This cluster method defines classes and puts objects in each class. Take for instance library management. In the library there is a wide range of books in various topics available.

The challenge now is how to keep those books in a way that readers can take several books in a particular topic without difficulty. By clustering technique, we can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name.

If readers want books in that particular topic, they would only have to go to that shelf instead of looking for it in the entire library. Depending on data and desired cluster characteristics, there different types of cluster paradigms such as representative based, hierarchical, density based and spectral clustering [8].

Prediction: The prediction, as its name implied, is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. In this case a model or a predictor will be constructed that predicts a continuous-valued-function or ordered value. Regression analysis is a statistical methodology that is most often used for numeric prediction. Prediction is used to predict some unknown or missing numeric values [9].

Data Mining Application: Data mining is increasingly popular because of the substantial contribution it can make. It can be used to control cost as well as contribute to revenue increase [10].

Data mining can help spot sales trends and other trend of events, develop better marketing campaigns and accurately predict customer loyalty and behavior. Specific uses of data mining are as follows:

- Market segmentation- This identifies the common characteristics of customers who buy the same products from your company.
- Customer churn- This predicts which customers are likely to leave your company and go to another competitor.
- Fraud detection- It identifies which transactions are most likely to be fraudulent.
- Interactive marketing- This predicts what each person accessing a website is most likely to be interested in seeing.
- Market basket Analysis- This helps to understand what product and services are commonly purchased together, eg. Soft drinks and snacks.
- Trend analysis- This reveals the difference between a typical customer this month and last.

Data mining can assist financial institutions (banks) in areas such as credit reporting and loan information.

Data mining can assist credit card issuers in detecting potentially fraudulent credit card transaction.

Data Mining can aid law enforcement officers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit and other patterns of behaviours.

Data mining speeds up researchers' data analyzing process, thus, allowing them more time to work on other projects. Data mining can be used to discover Useful knowledge from huge University database, which will help the University take important decisions. For instance data mining can assist the University to find out which students will join particular course programs and which students will require assistance in order to graduate and things like that.

CONCLUSION

As the development of information technology has generated large amount of data in various areas such as companies, tertiary institutions etc, application of data mining in our huge databases is imperative and timely too, in West Africa. This will help in extraction of useful information and pattern from huge data. This knowledge can be applied in various areas of human endeavour to obtain maximum benefits.

Recommendations: Data mining processes and techniques are recommended to the following organizations.

- Companies
- Tertiary institutions
- Commercial concerns
- Banks etc.

Data mining, when applied to the huge data will enable organizations make better decisions and obtain maximum benefit from investment.

REFERENCES

1. Ramangeri, B.M., 2010. Data Mining Techniques and Applications, India Journal of computer science and Engineering, 4: 301-304.
2. Doug, A., 2011. Data Mining (Online:www.lait.utexas.edu/~a noman/Bus/for/ course/mat/Alex).Retrieved 21/12/2015.
3. Sawyer, S.C. and B.K. Williams, 2007. Using Information Technology: A Practical Intoduction to Computers and Communication, New York, Mc Graw-Hill Publishers, pp: 424-426.
4. Singh, Y. and Chauhan, 2009. Neural Networks in Data Mining, Journal of Theoretical applied information Technology (online: www.jatit.org), pp: 37-41. Retrieved 15/11/2015.

5. Bresfelean, V.P., 2015. Data Mining Applications in Higher Education and Academic Intelligence Management (Online:http://ideas.repec.org/p/Pra/ Mprapa/2135.html)
6. Moshkovic, H.M., A.I. Mechitove and D.L. Olson, 2002. Rule Induction in Data Mining: Effect of ordinal Scale (Online: www.elsevier.com/locate/elswa). Retrieved 3/5/2015.
7. Wikipedia, 2013. Decision Tree (Online:http://en.wiki.org/ wiki/Decision Tree). Retrieved 21/12/2015.
8. Mohammed, J.Z. and M. Wagner, 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge, Cambridge University Press.
9. Jiawei, H. and M. Kamber, 2006. Data Mining Concepts and Techniques, Morgan Kaufmann Publishers.
10. Two Crows Corporation, 2005. Introduction to Data Mining and Knowledge Discovery 3rd Edition, Potomac (Online:www.twocrows.com/intro dm-pdf). Retrieved 21/12/2015.