Computing, Information Systems & Development Informatics Journal
Vol. 10 No. 2, June, 2019 - www.cisdijournal.org

A Data Mining Framework for Improving Agricultural Production in Nigeria

[1]Akpojaro, Jackson & [1]Waidor, Tamaramiebi K.
[1]Department of Mathematical Sciences
Faculty of Basic and Applied Sciences
University of Africa
Toru-Orua, Bayelsa State, Nigeria
E-mails: jakpojaro@yahoo.com, zalimaxxx@gmail.com
Phone: +2348138803299

ABSTRACT

The concept of e-Agriculture is the integration and utilization of information communication technology (ICT) in agricultural related operations. Agriculture is very significant in the development of Nigeria. Hence, all efforts of the government and other stakeholders are required to drive the agricultural potential of the country. Before the discovery of oil in the 50s, agriculture was the main economic endeavours of both governments and individuals. The incorporation of ICT into agriculture involves the integration of diverse technologies, which are capable of positively impacting the efficiency of agricultural activities to promote sustainable agricultural practice and food security. The effective application of ICT, no doubt, enhances decision making and improves production capacity for the overall development of the country. In this paper, the application of data mining and its significance in enhancing and equipping stakeholders with the right information for decision making on several issues related to agriculture and farming is investigated. It discusses the role of data mining in the context of agriculture and, also review several data mining techniques and their applications in extracting significant data in an effort to obtain knowledge and trends, to eliminate manual tasks and easier data extraction directly from electronic sources, transfer to secure electronic systems of documentation to enable production cost reduction, higher yield and higher market value. It integrates the work of various authors, which provide a useful platform for stakeholders to get information of current scenarios of data mining techniques and applications in context of agriculture and its usefulness in national development. In this effort, the paper summarizes some data mining techniques, which include Decision Tree, Artificial Neural Networks, Bayesian Network, Fuzzy Logic, Support Vector Machine, K- Nearest Neighbor and K- Means, and their applications in agriculture. Based on the peculiarity of our environment, the work proposes a suitable stakeholder framework with all the key players working in unison to achieve desired goals. Significantly, the work empowers farmers and other stakeholders with right and accessible ways of acquiring and disseminating information (knowledge) to make informed decisions anytime, anywhere with regard to improving agricultural productivity in Nigeria.

Keywords: Information communication technology, Agriculture, Environment, Stakeholders, Data mining techniques, Knowledge, Market value.

CISDI Journal Reference Format
Akpojaro, J. & Waidor, T.K. (2019): A Data Mining Framework for Improving Agricultural Production in Nigeria. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 10 No 2, Pp 61-68. Available online at www.cisdijournal.org
DOI Affix - https://doi.org/ 10.22624/AIMS/CISDI/V10N2P5

1. INTRODUCTION

Just as in several fields of endeavours, ICT is making a great inroad in the agricultural sector (Chavula, 2014). The use of ICT has helped in informed and factual decision making, which enhances productivity and marketability of agricultural process and products. Data mining which is a branch of computer science, is the discovery of knowledge using patterns from a large repository of data using some statistical or mathematical algorithms (Waidor and Akpojaro, 2018).

It is also known as knowledge discovery from data (KDD), albeit some authors argued that the term 'data mining' is just an important stage in the process of knowledge discovery (Jiawei *et al.,* 2012).This is similar to mining gold from a rock (Ali, 2009). The beauty of data mining is its multidisciplinary nature as it finds vital application in several disciplines such as education, medicine, transportation, agriculture, etc.

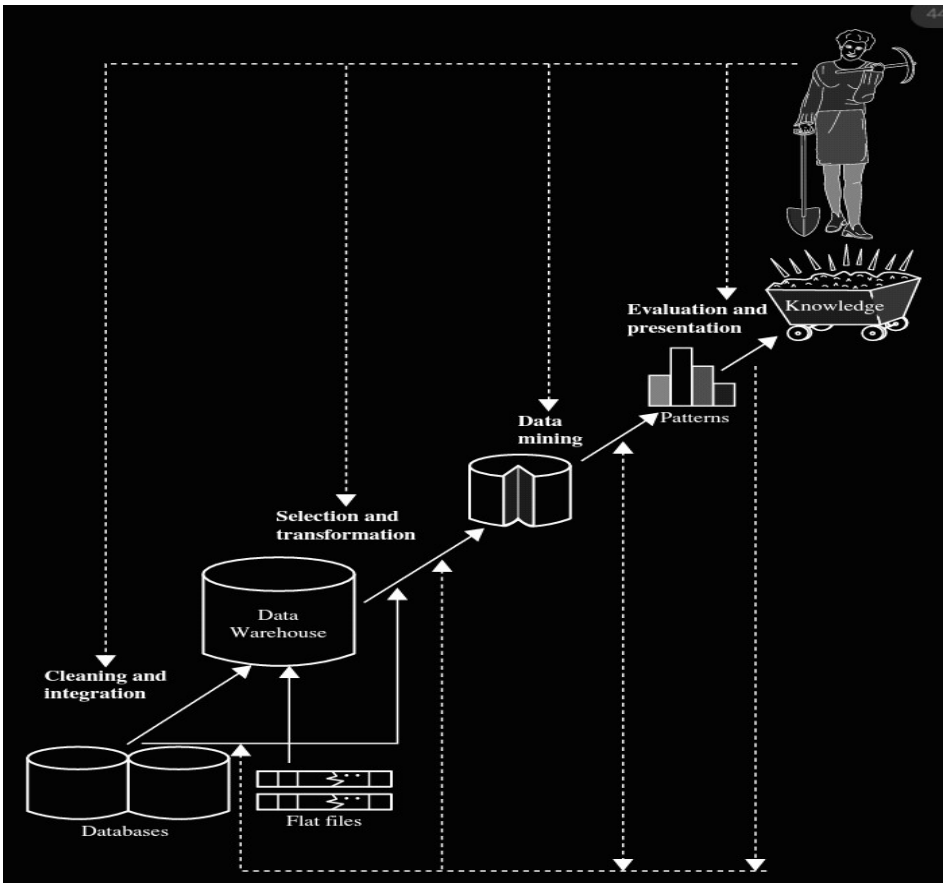

Figure 1. Data Mining Iterative Process (Jiawei et. al., 2012).

As shown in Figure 1 above, Data mining is an iterative process that includes the following steps;

- Data collection/cleaning: It always starts with the collection of data then proceed to sorting in order to remove noise and inconsistent or irrelevant data.
- Data integration: This is where multiple data sources may be combined, mostly heterogonous sources.

- Data selection. This is the stage where data relevant to the analysis decided on are retrieved from the database.
- Data consolidation/transformation: Where data are transformed into forms appropriate for mining procedures.
- Data mining: An essential process where mathematical or statistical techniques are applied to extract useful data patterns.
- Pattern evaluation: This is the stage where genuinely interesting patterns representing knowledge based on given measures.
- Knowledge presentation: This is the final stage where visualization and knowledge representation techniques are used to present mined knowledge to users for onward decision making.

Steps one and two are considered as data preprocessing stage and the outcome is usually stored in a data warehouse.

2. REVIEW OF DATA MINING TECHNIQUES

Data mining techniques could be descriptive or predictive as briefly discussed below.

Descriptive data mining: This data mining task is used to characterize the general properties of the data of interest in the database. Kwame, Hayfron-Acquah and Michael (2016), descriptive data mining examples include clustering, summarization, association rule, sequence discovery, etc.

Predictive data mining: This is used to predict unknown variables such as weather forecasting, revenue expected, and crop yield, etc. from patterns determined from known results. This is the most commonly used data mining approach. It can be used to predict variables such as future crop, appropriate kind of fertilizer and pesticide to be used, weather forecasting and expected revenue (Geetha, 2015). In this research, Geetha (2015) defined classification as a process that enables a model to learn to predict a class label from a set of training data which can be used to predict discrete class labels on new samples. Concisely put, it is the use of known values to forecast unknown and futuristic values. Kwame *et al.* (2016) employed algorithms such as classification, regression and time series analysis with focus on predictive data mining, narrowing down mainly on classification algorithms. Following from above, we proceed to take a brief overview of some classification data mining algorithms for better understanding of the subject matter.

Decision Tree: Waidor and Akpojaro (2018) defined a decision tree as a predictive model in which an instance is classified by following the path of satisfied conditions from the root of the tree until reaching a leaf, which will correspond to a class label. Neelamegam and Ramaraj (2013) described a Decision Tree to consist of nodes that form a rooted tree, that is, it is a directed tree with a node called root. Examples of the application area of the decision tree is the prediction of high or low crop productivity using factors such as humidity, rainfall, evaporation, temperature, etc.

Fuzzy Set: This is also known as possibility theory. It lets us work at a high abstraction level and offers a means for dealing with imprecise data measurement. Precisely, fuzzy set theory allows us to deal with vague or inexact facts. It can be applicable in the area of weed and disease control.

Artificial Neural Network: An artificial neural network (ANN), also known as "neural network" (NN), is a mathematical model or computational model based on biological neural networks . Neelamegam *et al.* (2013), it consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. ANN can be used to predict climatic conditions, which is a very important agricultural factors using such parameters as temperature, wind direction and speed, rainfall and humidity, etc.

K-Nearest Neighbor: These are classifiers that are based on learning by similarity (Cunningham and Delany, 2007; Salvador–Meneses, Ruiz–Chavez and Garcia–Rodriguez, 2019). The training samples are described by n-dimensional numeric attributes. Each sample represents a point in an n-dimensional space. In this way, all of the training samples are stored in an n-dimensional pattern space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the *k* training samples that are closest to the unknown sample (Neelamegam *et al.,* 2013). An example of the implementation of K-Nearest Neighbor in agriculture is the simulation of daily precipitations and other weather conditions.

Support Vector Machine (SVM): SVM is an algorithm that uses a nonlinear mapping to transform the original training data into a higher dimension (Bonaccorso, 2017; Gandhi, 2018). Within this new dimension, it searches for the linear optimal separating hyperplane (i.e., a "decision boundary" separating the tuples of one class from another). The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data. It can be applied to conduct climate impact studies.

K-Means: The k-means algorithm was developed as a partitioning technique (Li and Wu, 2012). It is most useful for forming a small number of clusters from a large number of observations. It handles variables that are continuous with no outliers. The objective of the k-means algorithm is to divide *N* observations with *P* dimensions (variables) into *K* clusters so that the within-cluster sum of squares is minimized. Since the number of possible arrangements is enormous, it is not practical to expect the best solution. Rather, this algorithm finds a "local" optimum. This is a solution in which no movement of an observation from one cluster to another will reduce the within-cluster sum of squares. The algorithm may be repeated several times with different starting configurations. This technique is used to classify soil characteristics in combination with global positioning system (GPS).

Bayesian Network: This is a statistical classifier that can predict class membership probabilities such as the probability that a given tuple belongs to a particular class (Ben-Gal, 2007; Hruschka and Nicoletti, 2013). It is based on a Bayesian Network, which represents a joint probability distribution over a set of categorical attributes.

3. DATA MINING FRAMEWORK FOR SUSTAINABLE DEVELOPMENT OF AGRICULTURAL PRODUCTION

Agriculture in Nigeria is one the few areas of sustainability that every class of her citizens freely involves either in commercial or subsistent basis. Agriculture is enshrined in the rich culture of the Nigeria people (Ogundare, 2015). It has been the main sources of livelihood of the people since pre-colonial era. It was also the main driver of the Nigeria economy until the advent of crude oil in the 50s which led to a shift in paradigm. However, after more than half a century of swimming in the topsy-turvy world of crude oil, the Nigerian government now faces the stack reality of reverting back to agriculture as the mainstay of the Nigeria economy.

For this to be possible, all hands need be on deck. As it is today, Agriculture remains the cardinal driver of the economy as regards to employment in spite of heavy reliant on oil, benefiting both the rural and urban settlers. Historically, there have been a decline in the contribution of agriculture to the gross domestic product (GDP) of the Nigerian economy. Orji (2011), there was 64% contribution in the 60's, 48% in the 70's and 19% – 20% in the 80's. However, the increase in population has led to an increase in food demand, making the need to revert back to agriculture more glaring. Nigeria now finds herself in the precarious position of importing agricultural products (such as rice, chicken, beans, etc.) that she can easily produce to guarantee food security in the country. There is no gain saying that the agricultural sector has the potential to be the industrial and economic springboard from which a country's development can take off (Orji, 2011).

Any country that depends on another country to feed her citizens is quite at a disadvantage. This perhaps is the reason why different government administrations have tried different policies to tackle this perennial problem. In the past, Nigeria has had the Operation Feed the Nation (OFN), Green Revolution (GR), Structural Adjustment Programme (SAP), etc. But implementing all these policies and programmes have not been too successful. This is in stark contrast to what was obtainable in the 50's when Nigeria was ranked very high in the production and exportation of major crops in the world (Ogundare, 2015), even becoming one of the largest exporters of palm oil, cocoa and groundnut. Today the groundnut pyramids for which Nigeria was so famous can only be seen at the back of the ₦200 note and our youths can hardly identify the cocoa tree.

Subsequently, the drive to improve agriculture in the country leads to frequent collection of data and eventually, a large volume of agricultural data that may not be useful because 'sense' cannot be got from them as they are littered everywhere. Therefore, Data mining can be used to make 'sense' out of this large repository of data in order to enhance decision making to boost agricultural productivity. Data mining can also be used to forecast future trends of agricultural processes to guide farmers in their daily and seasonal operations (Geetha, 2015). In this research work, we consider some key players that form the bedrock for sustainable agricultural development such as Government, Association of Farmers, Research and Development Centres, International Bodies, etc., and how they can collaborate and synergize for greater productivity. Furthermore, we looked at how an e-Agriculture centre can benefit these stakeholders greatly as shown in Figure 2.

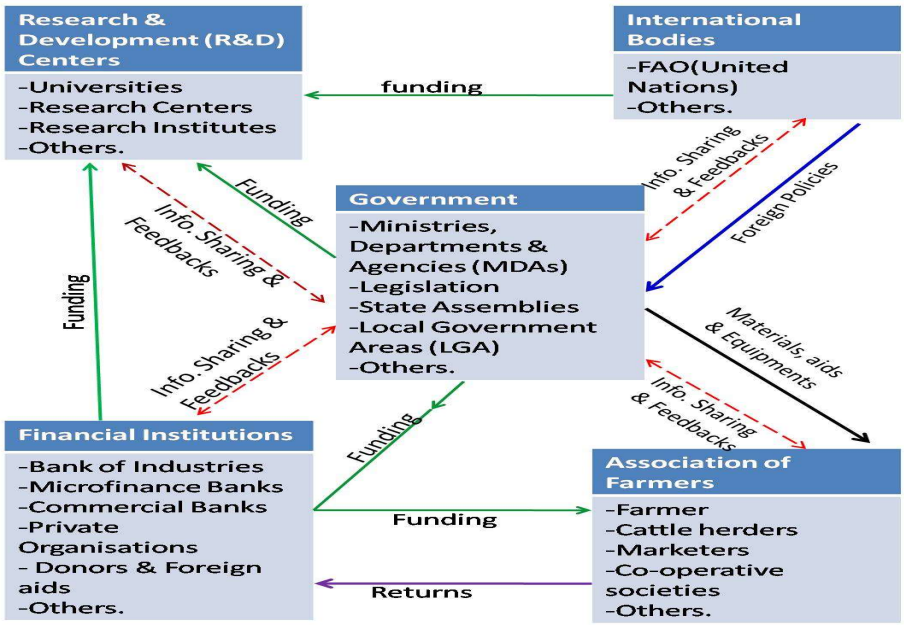

Figure 2. Stakeholders Framework

Government:

This comprises the Ministries, Departments and Agencies (MDAs) affecting Agricultural development in the country, the Legislature, State Assemblies, Local Government and others. Government must play a very cardinal role for any country desirous of agricultural growth and expansion. In short, they are at the core of the whole mix. The Legislature and State Assemblies need to formulate policies that can significantly aid agricultural growth and development at both the Federal and State levels. This would give the MDAs a focus on how to go about their businesses in line with government policies. Here every conflicting interest are sorted out to avoid bottleneck.

Government also plays the role of interacting with other stakeholders such as Research and Development Centres, International Bodies, Association of Farmers and Financial Institutions, etc. As shown in Figure 2, they provide information and receive feedbacks, take care of Farmers' education, training and enlightenment, provide financial aids as well as materials and equipment to the relevant bodies as the case may be. Furthermore, Government interfaces between other key players. For example, the Research and Development centres may through research discover an impending epidemic and channel such to the Government, who in turn disseminate this vital information to farmers' association and other stakeholders.

International Bodies:
International Bodies like Food and Agricultural Organization of the United Nations (FAO) working directly with MDAs (such as Agriculture and Rural Development; Budget and National Planning; Water Resources which houses Department of Irrigation and Environment which is responsible for forestry and climate change issues) provide strategic support to national development programmes and strategies aimed at reducing poverty, improving food security and management of natural resources. They also provide technical assistance and policy advice for the development of crops, livestock, fisheries and forestry sub sectors, with due emphasis given to institutional capacity building and development of local skills and expertise to ensuring sustainability and future local support.

Research and Development (R&D) Centres:
These too are major key players in the quest for food freedom. The universities and research centres play the role of feeding the relevant government bodies of their latest findings or discoveries or inventions as it relates to Agriculture. In return, they require constant and adequate government funding, monitoring and feedbacks.

Association of Farmers:
The voice of the farmers can be better heard when they form themselves into associations. They are the bedrock of the stakeholders' framework. They need adequate funding, up to date information, farmlands, good seeds, fertilizers, materials, equipment and a competitive market to sell their farm and allied products. Farmers also need to be in touch with each other using available social networks, local radio station and other means to disseminate and gather relevant information.

Financial Institutions:
Financial institutions such as Banks of Industry (BoI), Microfinance Banks, Commercial Banks, Foreign aids and Donors, etc. are also imperative in the stakeholders framework. No matter how good a plan may be, without funding, it will only remain at best a prototype. Agriculture needs a lots of funding to get the desired results. Government can also use these financial bodies as a channel to fund farmers instead of direct funding. This would ensure proper financial administration and accountability (an average Nigeria farmer may consider Government money as free money, but no farmer in his right sense would consider bank money as free money). Consequently, farmers are obliged to make returns to the financial institutions when yields start rolling out. Another key player in the stakeholder framework in dire need of adequate funding is Research and Development (R&D) centres. This funding could be from the Government, Financial Institutions and/or International Bodies. In return, the R&D centres are obliged to forward findings and discoveries to Government (MDAs) for decision making to share such relevant information to the farmers.

4. E-AGRICULTURE FRAMEWORK

Having looked at the stakeholder framework in the previous section, we are now acquainted with their importance in the scheme of things, and as such they form the bedrock for an e-Agriculture driven framework. Our focus here shall be on how an e-Agriculture centre can be of help to the stakeholder and the role of data miners in the development of the eAgriculture centres.

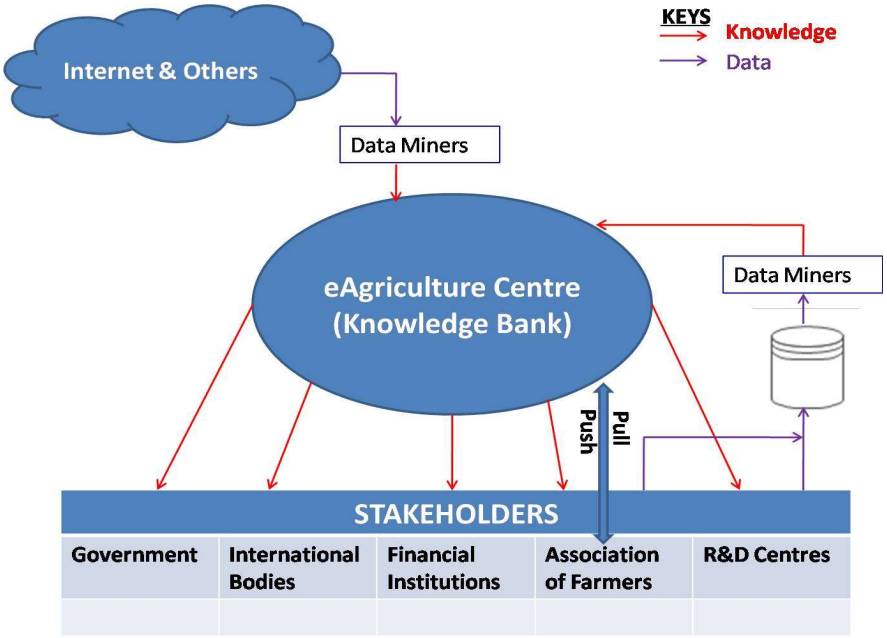

Figure 3. E-Agriculture Framework

Information (Knowledge) is the key driver of an e-Agriculture centre. And such information should be made readily available to the farmers and other stakeholders for an informed decision making. The core of the framework as shown in Figure 3 is the Knowledge Bank, where information can be easily accessed. As R&D centres make findings and discoveries, a lot of data are generated. Furthermore, even farmers generally generate quite a huge amount of data. So, from this large repository of data generated, data can be filtered to get knowledge which in turn is fed to the knowledge bank. Another great source of date which cannot be ignored in e-Agriculture is the Internet. However, the huge amount of data available here could be overwhelming and intimidating. Here again, data miners can determine useful pattern from this large repository of information and gain knowledge from them for onward transmission to the knowledge bank. It is pertinent to say that the integrity of the knowledge bank must be maintained at all times by avoiding the entrance of irrelevant and redundant data. The miners also need to be in close contact with the farmers to be more agriculture bias than being too technical. This would ensure mining more agriculture based knowledge that technical based knowledge. Farmers can get information from the knowledge bank using push or/and pull message system. This is briefly explained as follows;

Push Message: Important messages and updates unsolicited for are sent to farmers through SMS, emails, social media platforms, etc. These notifications keep the farmers abreast of trends and climatic situations.

Pull Message: This is when the farmers query the knowledge bank for specific information, and are fed with the required information.

Other means through which the farmers can get information from the knowledge bank are through websites, blog sites, bulletins, radio and television broadcast. Let not forget that Nigeria is a broadly diverse country so language barriers must be put into consideration for more effective communication with the farmers. All other stakeholders such as Government, Financial institutions and others can also gain immensely from the knowledge bank through knowledge acquired.

5. CONCLUSION

In this research paper, we reviewed some data mining algorithms that are useful in agricultural development and productivity in Nigeria. We also proposed a stakeholder framework with all the key players working in unison to achieve common goals. Then we presented an e-Agriculture framework where data miners play significant role of providing useful information for farmers and other stakeholders. Significantly, the work empowers all stakeholders with right and accessible ways of acquiring and disseminating information (knowledge) to make informed decisions anytime, anywhere with regard to improving agricultural productivity in Nigeria. Future work would aim at developing and deploying a seamlessly data mining system for farmers and other stakeholders.

REFERENCES

1. Ali, I.G.M.A. (2009). Gold Mineralization. A PhD Research Thesis, Department of Geology, Faculty of Science, Alexandria University, Egypt.
2. Ben-Gal, I. (2007). Bayesian Networks. Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons.
3. Bonaccorso, G. (2017). Machine Learning Algorithms. Packet Publishing, Birmingham, UK.
4. Chavula, H.K. (2014). The Role of ICTs in Agricultural Production in Africa. Journal of Development and Agricultural Economics, 6(7): 279-289.
5. Cunningham, P., and Delany, S. J. (2007). K-Nearest Neighbor Classifiers. Technical Report UCD-CSI-2007-4.
6. Hruschka, E. R., and Nicoletti, M. C. (2013). Roles Played by Bayesian Networks in Machine Learning: An Empirical Investigation. Springer, Berlin, Heidelberg.
7. Gandhi, R. (2018). Support Vector Machine - Introduction to Machine Learning Algorithms SVM model from scratch. Data Science.
8. Geetha, M. C. S. (2015). A Survey on Data Mining Techniques in Agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2): 887 – 891.
9. Kwame, B. A., Hayfron-Acquah, J. B., and Asante, M. (2016). An Overview of Data Mining Models (Descriptive and Predictive*). International Journal of Software & Hardware Research in Engineering*, 4(5): 53 – 60.
10. Jiawei, H., Micheline, K. M., and Pei, J. (2012). Data Mining Concepts and Techniques. Morgan Kaufmann Publishers.
11. Li, Y., and Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia 25 (2012) 1104 – 1109.
12. Neelamegam, S. and Ramaraj, E. (2013). Classification algorithm in Data mining: An Overview. *International Journal of P2P Network Trends and Technology (IJPTT),* 3(5): 2-4.
13. Ogundare, S. F. (2015). Teaching Cultural Diplomacy in Nigeria: Principles and Practices. A paper presented at the Annual Conference on Cultural Diplomacy 2015, Berlin, Germany, December 10th -13th, 2015, organized by Institute for Cultural Diplomacy.
14. Orji, O. (2011). Analysis of the Contribution of Agricultural Sector on the Nigerian Economic Development. *World Review of Business Research*, 1(1): 191-194.
15. Saurabh A. G., Priyanka M. M. (2015). E-Agriculture: Introduction and Figuration of its Application. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(1): 46-47.
16. Salvador–Meneses, J., Ruiz–Chavez, Z. and Garcia–Rodriguez, J. (2019). Compressed kNN: K-Nearest Neighbors with Data Compression. Entropy, 21, 234; doi:103390/e21030234
17. Waidor. K. T., and Akpojaro, J. (2018). The Use of Classification Algorithm for Students Academic Performance Forecasting. *In Proceedings of the 14th iSTEAMS International Multidisciplinary Conference,* AlHikmah University, Ilorin, Nigeria, 14: 153-158.