SSRG International Journal of Recent Engineering Science
ISSN: 2349 – 7157 / https://doi.org/10.14445/23497157/IJRES-V9I2P102

Volume 9 Issue 2, 8-17, Mar-Apr 2022


*Original Article*

# Big Data Mining Model to Predict Electronic Payment System using Machine Learning

Wesley Odeh Odumu[1], Ezekiel Endurance Chukwuemeke Igbonoba[2]

[1]*Department of Computer Engineering, Faculty of Engineering Technology, Plateau State Polytechnic, Barkin Ladi, Plateau State, Nigeria.*
[2]*Department of Computer Engineering, Faculty of Engineering, University of Benin, Benin City, Nigeria.*

Received: 01 March 2022 Revised: 09 April 2022 Accepted: 20 April 2022 Published: 25 April 2022

***Abstract*** *- This research presents a data mining model developed to predict the relationship between Nigeria's electronic payment (e-payment) systems. This proposes a data mining approach to establish the relationship between electronic payment and its impact on the economy. The Waikato Environment for Knowledge Analysis (WEKA) machine learning tool was used to develop the model using a simple regression technique. This predicts the usage in terms of volume and value of the following adopted electronic payment channels. The aim is to determine the performance measurement of the electronic payment system in the Nigerian banking sector. The data mining model developed can predict e-payment transactions over a number of years. The dataset used to assess and validate the authenticity of the model developed was obtained from the Nigeria Inter-Bank Settlement System (NIBSS) and the Central Bank of Nigeria (CBN). The result obtained indicates a positive relationship and contribution of e-payment networks to cost-effective progress with the modern move to a cashless economy in Nigeria. This equally impacted positively on the banking performance. The study revealed that the developed model would prove to be a preemptive and predictive tool for Nigerian banks to better policy formulation, financial advisory services, and performance measurement.*

***Keywords*** *- Big data, Data mining, E-payment, Electronic payment channel, Machine learning.*

## 1. Introduction

### *1.1. Background to the Study*

The increased convenience in banking operations inconsonant with the payment system due to the introduction of technology [1]. Big Data is turning this huge amount of data into meaningful benefits for organizations and customers. The payment system has gone beyond the traditional cash system to e-payment platforms like ATMs, internet, POS, Mobile money solutions, etc. The acceptance of transacting economic substances and safer and quicker access to funds have given the e-payment system a more acceptable platform than the cash-based system [2]. The e-payment system is fast being embraced and has gained relevance over the monetary transaction in brick-and-mortar banks in Nigeria. Therefore, the cash-based payment method is at its lowest ebb giving rise to cashless system dominance in the modern Nigerian economy [3], [4].

Moreover, customers do encounter a delay in accessing the services delivered through these electronic channels, proving that the system has some problems associated with it [5]. The internet is getting hold of popularity as a delivery conduit in the banking sector, and, at the same time, customer requests are shifting [6]. The financial stability, monetary policy, and overall economic activity in Nigeria solely depend on the well-functioning e-payment system, which formed her modern market economy [7].

Historically, the Central Bank of Nigeria (CBN) introduced a payment system facilitating e-payment in 2002. During this period, Nigeria's Automated Clearing System (NACS) was introduced as a veritable platform for developing electronic payments and reducing the time wasted in clearing cheques. The Inter-switch launched ATM services in 2003, implemented Real Time Gross Settlement in 2006, and migrated to a new uniform accounting system (NUBAN) in 2010. Subsequently, in early 2011, the Nigerian Interbank Settlement System announced instant payment services, and the first set of cash deposit ATMs was launched [7].

Research on the e-payment method in Nigeria showed that the system is progressively gaining consumers' recognition with steady growth in high proportion over the years, despite some hitches noticed in the system [8],[9],[10],[11]. Interestingly, approximately eighty billion nairas (N80b) of electronic transactions are recorded daily [7]. The citizenry will continue to enjoy the prosperity of ATMs, POS, ETF, Smart cards, e-cheques, and other forms of e-payment systems with the CBN cashless economic policy [8].

### *1.2. Big Data*

Big data is described as a large volume of structured, semi-structured, and unstructured data [12]. These are data collections whose size is away from the capacity of normally used software tools and storage systems to capture, save, manage, and process the data in an allowable lapsed period [13].

The high volume of bank data qualifies it into big data in terabytes with proponents of velocity (Batch, Real-time, and stream), variety (structured, semi-structured, unstructured data, and all of the above) has now exceeded the abilities of what information technology (IT) system can ingest, store, analyze and process it in time.

In 2012 Gartner updated the definition of the 3V model (variety, volume, and velocity) to (high variety, high volume, and high velocity) owed to boring data in a vast area like the Bank industry. Gathering a large amount of data from different sources makes big data very powerful for the banking industry to make effective decisions faster and better than traditional business tools.

The exponential growth and availability of structured, semi-structured, or unstructured data are called big data. The revealed patterns, trends, and any other possible associations of the big data set are obtained through empirical analysis. The era of big data came with both big opportunities and challenges; every field of human endeavor is experiencing an overflow of information at unpredictable volumes and speeds [14].

The Online Analytical Processing (OLAP) used for multidimensional analysis was first used, followed by Business Intelligence Analyses data before Analytics was used for statistical and mathematical derivations in the evolution of big data technology. Recently, Big Data for tremendous size and unstructured data analytics is presented in Figure 1.

Figure 1 presents the evolution of big data technology in statistical and mathematical analysis.

Big data storage, manipulation, and transformation to knowledge are the major issues associated with Big Data management. It is often thought that the huge volume of Big Data ensures that valuable knowledge is buried underneath what needs to be found. Still, analysts cannot just perceive the valuable content of the data [15]. A comparison of Big Data analytics with traditional analytics is presented in Table 1.

The traditional data analytic techniques, which work on structured data of low volume found to be inefficient in handling the variety and complexity offered by the big data, which is mostly unstructured or semi-structured and that comes in huge volume.

**Table 1. Traditional data vs. Big data analytics**

| Description | Traditional Analytics | Big Data Analytics |
|---|---|---|
| Data Sources | Trusted homogenous sources providing structured and static data | Heterogeneous sources providing unstructured/ semi-structured and streaming data |
| Data Storage | Isolated proprietary servers | Public/Private/ Hybrid Cloud |
| Database Technology | Relational data stores | NoSQL data stores |
| Data Processing | Centralized Architecture | Distributed Architecture |
| Analytics | On previously collected data | Need for real-time analytics |

### *1.3. Big Data Application*

Big data technologies have a wide and long list of their applications. It is used for search engines, logarithm processing, recommender systems, data warehousing, video, image analysis, banking & financial, telecom, retail, manufacturing, web & social media, medicine, healthcare, science & research, and social life. In politics, Mr. Barack Obama employed big data analytics to win the US presidential election in 2012.

The strategic approach to deploying Big Data technologies for preventing terrorism and crime reduction as a tool for national security and crime detection and prevention is a technology tested and implemented [16]. Equally, big data application is helpful in sport and education.

### *1.4. Big Data Analytics in the Banking Sector*

Big data technology comes with big opportunities and challenges, with every profession experiencing an overflow of information at unpredictable volumes and speeds [14]. The 21st century big data technological revolution has contributed immensely to financial service organizations, bearing in mind the cherished data management provision offered, which has solved the problem of secret fund movements. These have assisted in preventing fraudulent transactions across the globe.

The Banks harness Big Data into insights that help inform actionable, optimized, and timely decisions quickly and easily, keep risks at anticipated and acceptable levels and discover chances of leading in the competitive business. These are some meaningful benefits they attract for themselves and their customers. Fiscal organizations are adopting the introduction of big data in front-office risk management to support office skill processes [17]. The Big Data analytics flow is presented in Figure 2.

## 2. Literature Review

This section captures a comprehensive summary of previous research on the subject matter. It reviewed surveys, scholarly articles, books, and other sources relevant to the area of research.

### *2.1. E-payment System*

The electronic payment system can be seen as a broad term encompassing different dimensions of multichannel electronic delivery. E-payment could be viewed from its functions as m-payment, e-banking, e-money, online banking, internet banking, e-finance, e-broking, etc. [18], [19]. The e-payment is seen as the electronic preservation of economic substance on an intelligent device generally employed to make payments of undertakings apart from the person who issues it without involving bank accounts in the transaction, though acting as a prepaid bearer instrument elsewhere, e-payment is viewed as the use of credit cards, automated teller machines, debit cards, stored value cards, mobile wallets, and other payment systems [19].

Alternatively, it is defined as a payment facility that employs the usage of information and communications technologies and Integrated Circuit (IC) cards, cryptography, and telecommunications means to achieve its goal. However, in this study, e-payment refers to the delivery of a multichannel that provides for the electronic exchange of monetary substances without physical contact with the transacting parties. It includes all electronic transactions as well as e-cheque payments. E-payment means transacting business and settling financial commitments electronically without necessarily touching cash in a cashless society [20], [21].

### *2.2. Adoption of Electronic Payment Channels in Nigeria*

The e-payment tools, which comprise internet banking, POS, ATM, mobile money, etc., are instruments funded by banks in Nigeria with the backing of financial technology (Fin-Tech) organizations like Master Card Incorporation, Visa International, Inter-switch Company, etc. ATM and POS terminal end-users operate it with the chip activated cards. As of 2017, the e-payment channels in Nigeria have exceeded eighty-six trillion, one billion nairas (N86.1 trillion) as reported by the National Bureau of Statistics (NBS) and also reported that traditional doubt and avoidance have to lessen the expected ambiguity, which seems to be appreciated [22]. This is an approximately 32% rise above the total transaction value in 2016 of over sixty-five trillion (N65.1 trillion). In line with the objective of the payment system vision 2020, the (CBN) adopted payment system policies that will help to migrate from a cash-based economy to an e-payment-driven economy.

### *2.3. Big Data Analytics*

The data collected by organizations are employed to aid in decision-making. Big data analytics is the procedure of applying a set of rules to interpret sets of data and extract valuable and unfamiliar patterns, associations, and information [23]. It may also be applied to obtain advantageous, unknown, lawful, and secreted patterns and information from huge data collections in detecting the significant relationship between kept variables. In the context of these definitions, data should be complex and increasing in diversity, inclusive of its size. Simply considering the size of the data gives us enough oversight to understand that conventional methods would not be suitable for analyzing big data sets, and to compensate for the same, new methods and technologies are needed. The points mentioned above should be considered while going for the analysis of Big Data. Data analytics is considerably difficult compared to data collection and storage. Developing scalable and parallel machine learning algorithms for online analytics has been a serious challenge [24].

### *2.4. Waikato Environment for Knowledge Analysis (WEKA)*

This is developed and written in Java which is machine learning software. It is accessible with the General Public License (GNU) and is free. Its work surface comprises an assemblage of imagining platforms and algorithms for data study and analytical modeling in conjunction with the GUI for stress-free access [25]. Real-world data mining problems are solved by applying Java's WEKA software and running it in every system configuration. The algorithms can also be applied openly to a dataset or called from your personal Java code [25]. The processes and technological tools for collecting, storing, and analyzing big data, along with the principles that can govern its processes, are:-

- Clustering is the automation of finding correlated and meaningful data patterns within a set of data.
- Text Analytics: They rely on probability theory, rarity, and the occurrence of certain words, which are used to predict the meanings and overall ideas.
- Neural Networks: In this algorithm, nodes are activated by a signal to activate other nodes. Thus, a transfer function outputs a signal based on the received signal.
- Link Analysis: It is a subset of mathematics, and it is called graph theory. It represents a relationship between objects.
- Survival Analysis: It is called time-to-event analysis. It is a technique used to evaluate when you should start worrying about an event.
- Decision Trees: They are the most powerful data mining techniques capable of handling a diverse array of problems that can handle any data type. Decision trees split the data into small data cells. It aims at decreasing the overall entropy of data.
- Random Trees: This is the difference between possible errors and noise of an individual decision tree.
- NoSQL databases: This is also referred to as Not Only SQL, a data management method and database design beneficial for huge datasets of distributed data.

### *2.5. Related Works*

The researcher in [26] evaluated the electronic payment channels in banks in southern Nigeria, intending to determine how it affects stakeholders (financial institutions, financial regulators, product developers, customers, etc.). The researcher used a targeted sampling strategy and chi-square technique methodology. The study's findings showed that the present e-payment channels are seen to be safe, convenient, reliable, and accessible. Still, the security features, reliability of POS terminals, and accessibility of ATM systems and web services need to be improved. The researcher failed to proffer a solution to improving the system's security.

The study [27] investigated the impact of electronic payment systems on the economic growth of Nigeria. The study empirically investigated the impact of electronic payment systems on economic growth in Nigeria between 2009 and 2018. The study employed descriptive statistics, correlation analysis, and Auto-Regressive Distributed Lag (ARDL) Model to draw inferences. The findings showed that electronic payment systems positively impacted economic growth in Nigeria within the period under review. But could not state to which extent and the variables are responsible for the positive impact.

The study in [28] appraised data mining methods in e-business situations. The research aimed to review the data mining techniques in an e-business environment using secondary data sources gathered from customers' inward cycles, sellers, markets, and business environments. Apart from the gains identified, the researchers equally identified some challenges associated with data mining in the field of e-commerce, such as spider identification, data change, the flexibility of data mining algorithms, making the data mining model understandable to business users, upholding moderate changing dimensions and making data change and model structure accessible to business users. Finally, the researchers could not proffer a solution to the challenges identified.

The researchers in [29] investigated Data Mining in electronic commerce concerning its benefits and challenges. The study aimed to evaluate the use of data mining in e-commerce by aiming at structured and unstructured data collected through numerous means and cloud computing facilities to validate the significance of data mining. The researchers used a secondary data collection methodology by obtaining data from the customer's internal processes, vendors, markets, and business environment. Finally, the study revealed the benefits and challenges of data mining in e-commerce. The study could not ascertain the threat linked to the system.

The researchers in [30] researched the payment system data to forecast economic activity. The research aimed to find out the monthly data available on Italy's retail settlement system for long-term and short-term forecasting. The scholars embraced a mixed-frequency factor model approach established on a large-scale data set to forecast Italian GDP and its key constituents. The researchers' findings revealed that the payment data track economic activity with different aggregates of payment system flows in Italy with other indicators adopted in macroeconomic forecasting. Finally, the researcher could not give a convincing, detailed economic growth associated with the efficiency of the e-payment system.

The researchers in [31] examined the current dynamics in increasing e-payment in Iran by data mining methods. The research determined to ascertain the elements that inspire customers to agree to take e-banking throughout the nation by means of the data and information obtained from the Central Bank with the help of data mining techniques. The K-Means clustering algorithm methodology was applied to analyze the secondary data obtained from the Central bank. The indices of social-economic, Information and Communication Technology improvement, and business explosion were the greatest in improving the tradition of e-payment approaches as obtained from the decision tree rules result. The researcher failed to identify the threats associated with the system.

## 3. Materials and Method

This section deals with the research design, data collection, and data analysis methods. Simple linear regression is used to predict a variable from an independent variable; it estimates exactly how much *the Y* variable will change when the *X* variable changes by a certain amount. The data was collected from the Central Bank of Nigeria (CBN) and the Nigeria Inter-Bank Settlement System (NIBSS). It is a small dataset that contains the year, volume, and value in the naira of eight electronic payment channels from 2012 to 2017. WEKA data mining tool was used to evaluate the data.

### *3.1. Research Design*

This is a thorough framework of exactly how a study will be performed. This includes how data was collected, what instruments were employed, how the instruments were used, and the means for analyzing data collected. For this study, an experimental research design is used to establish a relationship between the cause and effect of Electronic Payment Channels (EPC) on the economy. This research design was adopted because it is a highly practical research design method that contributes to solving the problem. Secondly, a set of data from an existing database was used.

### *3.2. Data Collection*

This research work employed a secondary data collection method. The data was obtained from the Central Bank of Nigeria (CBN) and Nigeria Inter-Bank Settlement System (NIBSS), verified and validated by the National Bureau of Statistics, Nigeria (NBS).

### *3.3. Regression*

This is a commonly used data study processing statistical tool, and it describes the strength of the relationship between the dependent and independent variable(s) in a statistical investigation. It allows you to infer a relationship between two or more variables.

### *3.4. Mathematical Equation*

Simple linear regression models the relationship between the magnitude of one variable and that of a second. For example, as *X* increases, *Y* also increases. Or, as *X* increases, *Y* decreases [32]. The easiest regression model is the simple linear regression presented in Equation 1:

$$Y = \beta_0 + \beta_1 * x_1 + \varepsilon \tag{1}$$

Where $\beta_0$ is a *constant* and is the intercept of the regression line with the y-axis. $\beta_1$ is the slope of the regression line. It shows how much *y* changes for each unit change of *x*. The last term is the epsilon (ε), representing the estimation error. The error is the actual difference between the observed data and the regression-predicted data. *Y* is the variable we are trying to predict and is called the dependent variable. X is an independent variable. From the model equation in (1), our predictor is the number of years, and the response variables are the volumes and the value in naira.

### *3.5. Creating the Regression Model with Weka*

The data set from the repositories of CBN and NIBSS is loaded in WEKA using Attribute-Relation File Format (ARFF) method by supplying each row of the dataset in a comma-delimited format and building the Linear Regression Model on it. The list of instances sharing a set of attributes describes the ARFF file in the ASCII text file.

The *arff* file format requires a declaration of @RELATION (associates a name with the dataset), @ATTRIBUTE (specifies the name and attribute of an attribute) and @DATA (denotes the start of the data segment).

The model considered the years after 2012 to 2017 and beyond for this study. The linear regression on Years after 2012 is calculated thus: current year minus reference year.

E.g. 2018 – 2012 = 6.

The model is given as thus: -458705.71 * Years after 2012 + 14088764.29

So, then applying the formula for the volume-outcome, we have: -458705.71 * (6) + 14088764.29 = 11,336,530.03
The model calculation for the value in terms of naira is -492528000000 * Years after 2012 + 7876666666666.67

So, then applying the formula above for the value outcome, we obtained: -492528000000 * (6) + 7876666666666.67 = 4,921,498,666,666.67.

The model developed can determine the volume for a particular year, supposedly knowing the volume for the previous year by feeding the data into the model, and the volume for that particular year will be determined. Similarly, when known for the previous year, the value in naira can be fed into the model, for the value of that year can be determined too. The predicted volume and value that the linear regression model obtains consider the previous volume, value, and year in consideration.

### *3.6. Process of Creating a Linear Regression*

Firstly, you get sample data; then, you can design a model that explains the data; finally, you use the model you've developed to predict the whole population. Traditionally, the electronic spreadsheets application (Microsoft Excel) allows users to enter data in rows and columns, calculate the mean, perform statistical analyses, create tables, and produce other financial schedules. Spreadsheet software also has features that allow the creation of analytical graphics. When viewed on a monitor or printed out, analytical graphics or business graphics help make data and the generated report easy to comprehend and analyze for decision-making by management. The data in an MS Excess worksheet is presented in Figure 3.

### *3.7. Waikato Environment for Knowledge Analysis (WEKA)*

This is a collection of machine learning algorithms mining software [33]. It includes schemes for classification, numeric prediction, meta-schemes, and clustering. Linear regression is one of the implemented numeric prediction schemes. Weka uses Attribute-Relation File Format (ARFF). The ARFF file format requires the declaration of @RELATION (associates a name with the dataset), @ATTRIBUTE (specifies the name and attribute of an attribute), and @DATA (denotes the start of the data segment). The preprocessing phase in Weka is represented by the necessary actions that load the data. Once the data is loaded, a linear regression on the dataset may be performed. To perform this analysis, the Linear Regression must be chosen. It may be found under Classify tab right at the functions leaf. Selecting the dependent variable is the last stage in creating our model (the column we are looking to predict). The WEKA preprocess view is presented in Figure 4.

The data preprocessing, clustering, classification, regression, visualization, and feature selection data mining tasks are supported by WEKA [34]. Weka's performances are centered on the postulation that the data is obtainable as only a flat file or relation, where a secure figure of attributes defines every single data argument. The Java Database connectivity can process the outcome resumed by a database query owed to the delivery of WEKA in the SQL databases. Converting the pool of associated database tables into a single table is done with a separate software appropriate for handling Weka but incapable of multi-relational data mining. The key region is not enclosed in the Weka distribution algorithms in sequence modeling.

### *3.8. Format of Dataset in WEKA*

The normal technique of signifying datasets that comprise independent, unordered cases and do not contain relationships between cases is called an ARFF file. ARFF files have two distinct units, i.e., the Header and Data information. The name of the relation, a list of the attributes (the columns in the data), and their types ARFF file Header. The following graphical interfaces can be seen when WEKA is called on: the Explorer, the Knowledge Flow, the Experimenter, and the command-line interface. The WEKA Explorer view is presented in Figure 5.

## 4. Result and Discussion

This section describes the data presentation, result, discussion, and analysis of data obtained from the repositories of the CBN and NIBSS concerning the volume and value of the e-payment transaction in the Nigerian banking system. These data were analyzed using a simple linear regression method to predict the volumes and values of ATM and POS e-payment channels from 2012 to 2017. In the analysis, the actual data from 2012 to 2017 were presented. The model was tested and validated using only 2017 data. Table 2 presents the actual volume and value of EPC in ATM transactions from 2012 to 2017 as evaluated. The result obtained from the process is presented in Table 3.

**Table 2. Volume and value of EPC in ATM (Source: NIBSS and CBN)**

| Description | Year | Volume | Value (Naira) |
|---|---|---|---|
| ATM | 2012 | 375,513,<br>154 | 1,984,990,<br>636,830 |
| | 2013 | 295,416,<br>724 | 2,830,533,<br>105,570 |
| | 2014 | 400,269,<br>140 | 3,681,980,<br>955,458 |
| | 2015 | 433,695,<br>748 | 3,971,651,<br>486,420 |
| | 2016 | 590,238,<br>934 | 4,988,133,<br>401,544 |
| | 2017 | 800,550,<br>000 | 6,440,000,<br>000,000 |

**Table 3. Comparison of predicted and actual volume and value data of ATM for 2017**

| Year | Predicted Volume | Actual Volume | Predicted Value (Naira) | Actual Value (Naira) |
|---|---|---|---|---|
| 2017 | 699,976,<br>625.85 | 800,550,<br>000 | 6,056,988,<br>567,280.3<br>5 | 6,440,000,<br>000,000 |

The result obtained in Table 3 showed a difference of 100, 573, 374 .15 in volume in favor of the actual, and the predicted value presented a difference of 383,011,432,719.65 in favor of the actual value.

Table 4 presents the actual data of POS transactions in volume and value from 2012 to 2017 obtained from the CBN and NIBSS. The predicted result for the year 2017 compared to the actual is presented in Table 5.

**Table 4. Volume & Value of POS EPC (Source: NIBSS and CBN)**

| Description | Year | Volume | Value (Naira) |
|---|---|---|---|
| POS | 2012 | 2,587,595 | 48,461,883,431 |
| | 2013 | 9,418,427 | 161,212,840,665 |
| | 2014 | 20,817,423 | 312,071,736,903 |
| | 2015 | 33,720,933 | 448,512,548,727 |
| | 2016 | 63,715,203 | 758,996,505,702 |
| | 2017 | 146,272,162 | 1,409,798,000,00<br>0 |

**Table 5. Comparison of the actual and predicted volume and value data of POS for 2017**

| Year | Predicted Volume | Actual Volume | Predicted Value (Naira) | Actual Value (Naira) |
|---|---|---|---|---|
| 2017 | 109,961,2<br>43.21 | 146,272<br>,162 | 1,147,210,68<br>7,622.21 | 1,409,798,0<br>00,000 |

THE predicted result in Table 5 presents a difference of 36,310,918.79 in volume, while the predicted value presented a difference of 262,587,312,377.78 in value in favor of the actual

### *4.1. Result Validation*

The linear regression model analysis summary for the ATM transaction for the year 2017 presented a correlation coefficient of 0.89 for the volume and 0.98 for the value. Alternatively, the POS presented a correlation coefficient of 0.98 for volume and 0.93 for value. However, validating both results using the statistics from the correlation coefficient showed that the model developed is a good model by presenting values very close to 1. These presented a close match between the actual and predicted data for the two tested parameters. The summary of the validation statistics is shown in codes 1 and 2 in Appendix A.

## 5. Conclusion

Big Data mining has offered the Nigerian banking system an opportunity for global banking integration through e-payment transactions. This has enhanced the financial services offered to the customers and has increased the banks' performance, thereby affecting the economy positively. The ease of doing business has improved with technology. The model developed will serve as a tool for predicting economic fortune for Nigeria concerning e-payment transactions.

## References


[1] K. Donovan, Mobile Money for Financial Inclusion, World Bank E-Library, 2012. [Online]. Available: https://elibrary.worldbank.org

[2] C.K. Ayo, and J.O. Adewoye, "The State of e-Banking Implementation in Nigeria: A Post-Consolidation," *Journal of Emerging Trends in Economics and Management Sciences*, vol. 1, no. 1, pp. 37-45, 2010. [Google Scholar] [Publisher Link]

[3] Trimisiu Tunji Siyanbola, "The Effect of Cashless banking on Nigerian Economy," *e-Canadian Journal of Accounting and Finance*, vol. 1, no. 2, pp. 9-19, 2013. [Google Scholar] [Publisher Link]

[4] Omotunde Muyiwa, Sunday Tunmibi, and A.T. John-Dewole, "Impact of Cashless Economy in Nigeria," *Greener Journal of Internet, Information and Communication Systems*, vol. 1, no. 2, pp. 40-43, 2013. [Google Scholar] [Publisher Link]

[5] A. O. Olakah, "Benefit, Challenges and Prospects of a Cashless Economy," *Journal of the Chartered Institute of Bankers of Nigeria*, Lagos, 2012. [Google Scholar]

[6] D. P. Acharjya, and Kauser Ahmed P, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 2, pp. 511-518, 2016. [Google Scholar] [Publisher Link]

[7] Central Bank of Nigeria Communication, The Cashless Nigeria Project, 2011. [Online]. Available: https://www.cbn.gov.ng

[8] F. N. Echekoba and E. G. Kasie, "Electronic Retail Payment Systems: User Acceptability and Payment Problems in Nigeria," *Arabian Journal of Business and Management Review*, vol. 1, no. 6, pp. 111-123, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[9] Olugbade Adeoti, and Kehinde Osotimehin, "Adoption of Point of Sale Terminals in Nigeria: Assessment of Consumers' Level of Satisfaction," *Research Journal of Finance and Accounting*, vol. 3, no. 1, pp. 1-5, 2012. [Google Scholar] [Publisher Link]

[10] E. S. Odior, and F. B. Banuso, "Cashless Banking in Nigeria: Challenges, Benefits and Policy Implications," *European Scientific Journal*, vol. 8, no. 12, 2012. [Google Scholar] [Publisher Link]

[11] Odi Nwankwo, and Onyekachi Richard Eze, "Electronic Payment in Cashless Economy of Nigeria: Problems and Prospect," *Journal of Management Research*, vol. 5, no. 1, 2013. [Google Scholar] [Publisher Link]

[12] M. Kerznerand, and S. Maniyam, Hadoop Illuminated, 2015. [Online]. Available: https://goodreads.com/hadoop-illuminated/hadoop-book

[13] Wayne R. Kubick, "Big Data, Information and Meaning," *Applied Clinical Trial Digital Edition*, vol. 21, no. 2, pp. 26-28, 2012. [Google Scholar] [Publisher Link]

[14] V. Mayer-Schonberger, and K. Cukier, Big Data: A Revolution that Will Transform How We Live, Work, and Think, Houghton Mifflin Harcourt: Boston Massachusetts, USA, 2013.

[15] Tim Kraska, "Finding the Needle in the Big Data Systems Haystack," *IEEE Internet Computing*, vol. 17, no. 1, pp. 84-86, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[16] Babak Akhgar et al., *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies,* Butterworth-Heinemann: Oxford, 2015. [Google Scholar] [Publisher Link]

[17] Aisha Siddiqa et al., "A Survey of Big Data Management Taxonomy and State of the Art," *Journal of Network and Computer Applications*, vol. 71, pp. 151-166, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[18] David Humphrey et al., "Benefits from a Changing Payment Technology in European Banking," *Journal of Banking and Finance*, vol. 30, no. 6, pp. 1631-1652, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[19] Oyewole Oginni Simon et al., "E-banking and Bank Performance: Evidence from Nigeria," *International Journal of Scientific Engineering and Technology*, vol. 2, no. 8, pp. 766-771, 2013. [Google Scholar] [Publisher Link]

[20] R. Nzaro, and N. Magidi, "Assessing the Role of Electronic Payment Systems in Financial Institutions: A Case of a Savings Bank in Zimbabwe," *Global Journal of Management and Business Research*, vol. 14, no. 2, 2014. [Google Scholar]

[21] Tatenda Kavu et al., "An Electronic Payment Model for Small and Medium Enterprises in Zimbabwe," *International Journal of Scientific and Engineering Research*, vol. 4, no. 1, pp. 1-8, 2013. [Google Scholar] [Publisher Link]

[22] O. Kolawole, and SA Mustapha, Impact of Cashless Policy on Bank's Profitability in Nigeria, 2018.

[23] Niall M. Adams, "Perspectives on Data Mining," *International Journal of Market Research*, vol. 52, no. 1, pp. 11-19, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[24] Amr Osman, Mohamed El-Refaey, and Ayman Elnaggar, "Towards Real-Time Analytics in the Cloud," *IEEE Ninth World Congress on Services*, pp. 428-435, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[25] [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

[26] Francis Bukie Osang, "E-Banking: Evaluating Electronic Payment Channels In Southern Nigeria," *NOUN Journal of Physical and Life Sciences*, vol. 1, pp. 137-159, 2020. [Google Scholar] [Publisher Link]

[27] Udeme Okon Efanga et al., "An Empirical Investigation of the Impact of Electronic Payment Systems on Economic Growth of Nigeria, 2020. [Publisher Link]

[28] Sajan Mathew, and Dr. Jitendra Sheetlani, "Review of Data Mining Techniques in E-Business Environments," *Journal of Critical Reviews*, vol. 7, no. 15, pp. 4867-4872, 2020. [Google Scholar] [Publisher Link]
[29] Mustapha Ismail et al., "Data Mining in Electronic Commerce: Benefits and Challenges," *International Journal of Communications, Network and System Sciences*, vol. 8, no. 12, pp. 501-509, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[30] Valentina Aprigliano, Guerino Ardizzi, and Libero Monteforte, "Using Payment System Data to Forecast Economic Activity," *International Journal of Central Banking*, vol. 15, no. 4, pp. 55-80, 2019. [Google Scholar] [Publisher Link]
[31] Fateme Moslehi, Abdolrahman Haeri, and Mohammad reza Gholamian, "Investigation of Effective Factors in Expanding Electronic Payment in Iran using Data Mining Techniques," *Journal of Industrial and Systems Engineering*, vol. 12, no. 2, pp. 61-94, 2019. [Google Scholar] [Publisher Link]
[32] P. Bruce, and A. Bruce, *Practical Statistics for Data Scientists*, O'Reilly Media Inc, 2017.
[33] Mark Hall et al., "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10-18, 2009. [CrossRef] [Google Scholar] [Publisher Link]
[34] Sunita B. Aher, and Louis Lobo, "Applicability of Data Mining Algorithms for Recommendation System in e-Learning," *Proceedings of the International Conference on Advances in Computing, Communications, and Informatics*, pp. 1034-1040, 2012. [CrossRef] [Google Scholar] [Publisher Link]

## Appendix A

Code 1 and 2 are the screenshots of the WEKA program output on loading and running the model using the 2017 data as extracted from Tables 2 and 4, respectively.

**Code 1.**

```
===Run information ===
Scheme: weka.classifiers.functions.LinearRegression-
        S 0 -R1.0E-
Relation:     Research data_ATM
Instances:    6
Attributes:   2
                  Years after    2012
                  Volume ATM
Test mode: evaluate on training data
=== Classifier model  (full training set) ===
Linear Regression Model
VoLume ATM =
36545070.3404 * Years  after 2012 + 265251274.145
=== Summary ===
Correlation coefficient          0.8864
Mean absolute error        70278418
Root mean squared error        77527647.4336
Relative absolute error          49.5425%
Root relative squared error     46.283%
Total Number of Instances       6
===Run information ===
Scheme: weka.classifiers.functions.LinearRegression
          -S 0 -R 1.0E-
Relation:     ATMData_Value
Instances:    6
Attributes:   2
                  Years after 2012
                  ATM value in naira
Test mode:        evaluate on training data
 === Classifier model (full training set) ===
Linear Regression Model ATM value in Naira=
825643426912.1417 * Years after 2012 +
1508771432719.6458
=== Summary ===
Correlation coefficient       0.5333
Mean absolute error     221755617941.4279
Root seat squared error       262270321549.569
Relative absolute error       19.214%
Root relative squared error  18.2011%
Total Sober of Instances      6
```

**CODE 2**

```
=== Run information ===
Scheme: weka.classifiers.functions.LinearRegression
          -S 0 -R l.OE
Relation:     Research Data_POS
Instances:    6
Attributes:        2
                  Years after 2012
                   POS =
Test mode:    evaluate on training data
=== Classifier model (full training set) ===
Linear Regression Model
POS = 25545047.7485 * Years after 2012 + -17783555.5389
=== Summary ====
Correlation coefficient       0.8914
Mean absolute error      19445294.7078
Root mean squared error       22184384.7678
Relative absolute error       49.5169%
Root relative squared error  45.3214%
Total Number of Instances     6
===  Run information ===
Scheme: weka.classifiers.functions.LinearRegression
           -S 0 -R  1.0E-
Relation:     POSData_Value
Instances:    6
Attributes:   2
                   Years after 2012
                   POS Value in Naira
Test mode:   evaluate on training data
 ===Classifier model (full training set) ===
Linear Regression Model
```

POS Value in Naira= 249613816153.4867 * Years after 2012 + -100858393145.2167

===Summary===

| | |
|---|---|
| Correlation coefficient | 0.9346 |
| Kean absolute error | 141455721981.91 |
| Root mean squared error | 162198604608.6175 |
| Relative absolute error | 37.8074% |
| Root relative squared error | 35.5612% |
| Total Number of Instances | 6 |

## FIGURES

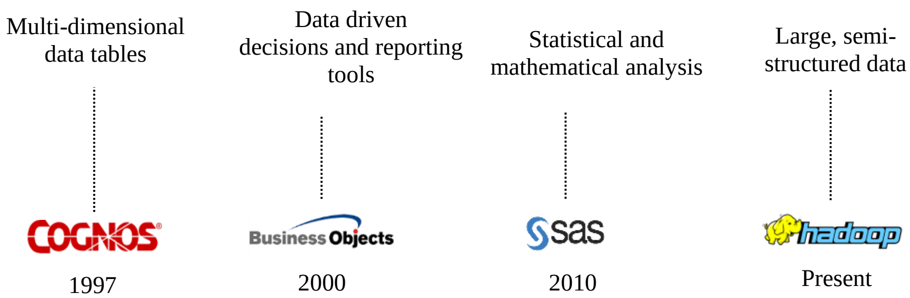


**Fig. 1 Trends of data analytics**

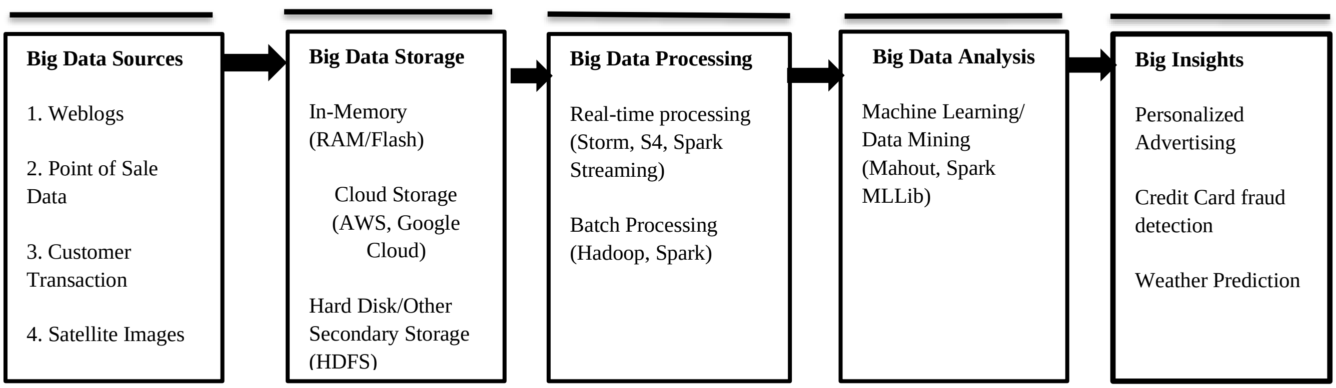


**Fig. 2 Big data analytics flow**

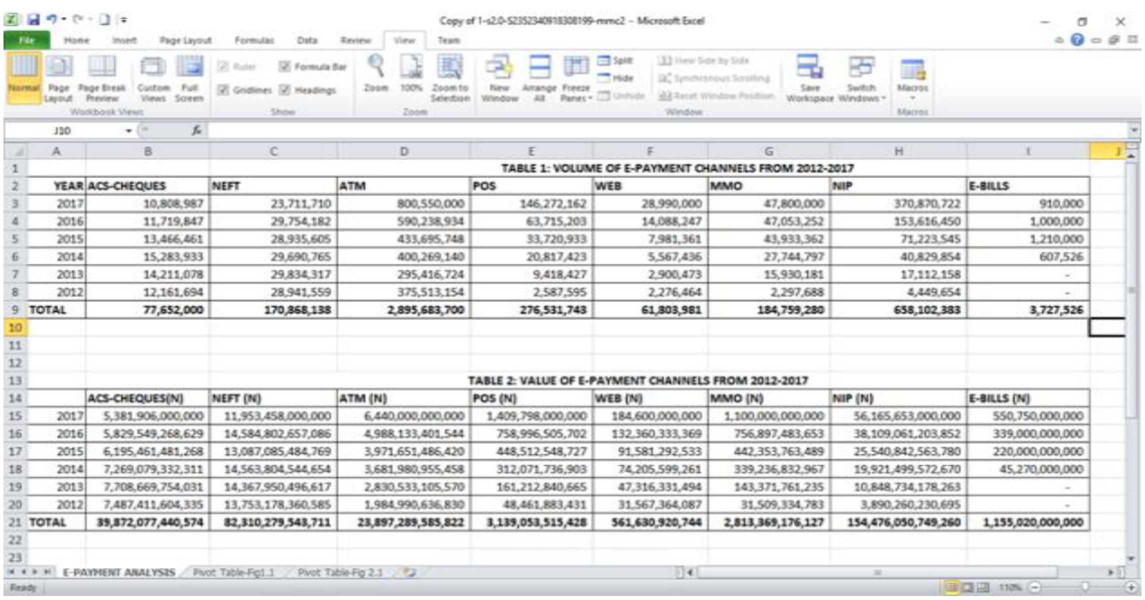

TABLE 1: VOLUME OF E-PAYMENT CHANNELS FROM 2012-2017

| YEAR | ACS-CHEQUES | NEFT | ATM | POS | WEB | MMO | NIP | E-BILLS |
|---|---|---|---|---|---|---|---|---|
| 2017 | 10,808,987 | 23,711,710 | 800,550,000 | 146,272,162 | 28,990,000 | 47,800,000 | 370,870,722 | 910,000 |
| 2016 | 11,719,847 | 29,754,182 | 590,238,934 | 63,715,203 | 14,088,247 | 47,053,252 | 153,616,450 | 1,000,000 |
| 2015 | 13,466,461 | 28,935,605 | 433,695,748 | 33,720,933 | 7,981,361 | 43,933,362 | 71,223,545 | 1,210,000 |
| 2014 | 15,283,933 | 29,690,765 | 400,269,140 | 20,817,423 | 5,567,436 | 27,744,797 | 40,829,854 | 607,526 |
| 2013 | 14,211,078 | 29,834,317 | 295,416,724 | 9,418,427 | 2,900,473 | 15,930,181 | 17,112,158 | - |
| 2012 | 12,161,694 | 28,941,559 | 375,513,154 | 2,587,595 | 2,276,464 | 2,297,688 | 4,449,654 | - |
| TOTAL | 77,652,000 | 170,868,138 | 2,895,683,700 | 276,531,743 | 61,803,981 | 184,759,280 | 658,102,383 | 3,727,526 |

TABLE 2: VALUE OF E-PAYMENT CHANNELS FROM 2012-2017

| | ACS-CHEQUES(N) | NEFT (N) | ATM (N) | POS (N) | WEB (N) | MMO (N) | NIP (N) | E-BILLS (N) |
|---|---|---|---|---|---|---|---|---|
| 2017 | 5,381,906,000,000 | 11,953,458,000,000 | 6,440,000,000,000 | 1,409,798,000,000 | 184,600,000,000 | 1,100,000,000,000 | 56,165,653,000,000 | 550,750,000,000 |
| 2016 | 5,829,549,268,629 | 14,584,802,657,086 | 4,988,133,401,544 | 758,996,505,702 | 132,360,333,369 | 756,897,483,653 | 38,109,061,203,852 | 339,000,000,000 |
| 2015 | 6,195,461,481,268 | 13,087,085,484,769 | 3,971,651,486,420 | 448,512,548,727 | 91,581,292,533 | 442,353,763,489 | 25,540,842,563,780 | 220,000,000,000 |
| 2014 | 7,269,079,332,311 | 14,563,804,544,654 | 3,681,980,955,458 | 312,071,736,903 | 74,205,599,261 | 339,236,832,967 | 19,921,499,572,670 | 45,270,000,000 |
| 2013 | 7,708,669,754,031 | 14,367,950,496,617 | 2,830,533,105,570 | 161,212,840,665 | 47,316,331,494 | 143,371,761,235 | 10,848,734,178,263 | - |
| 2012 | 7,487,411,604,335 | 13,753,178,360,585 | 1,984,990,636,830 | 48,461,883,431 | 31,567,364,087 | 31,509,334,783 | 3,890,260,230,695 | - |
| TOTAL | 39,872,077,440,574 | 82,310,279,543,711 | 23,897,289,585,822 | 3,139,053,515,428 | 561,630,920,744 | 2,813,369,176,127 | 154,476,050,749,260 | 1,155,020,000,000 |

**Fig. 3 Data in an MS Excel worksheet**

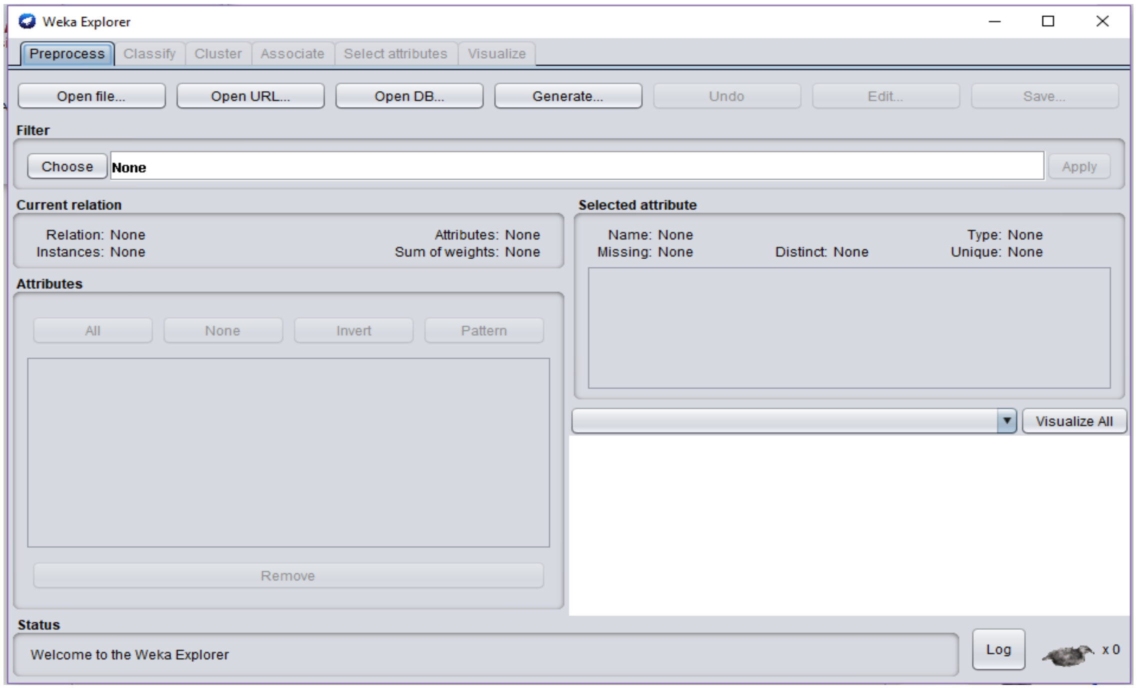


**Fig. 4 WEKA preprocess view**


**Fig. 5 WEKA explorer view**