Data Mining and its Application in Banking Industries in Nigeria

[1]Amanze, Bethran, [2]Eze, Irene, [3]Ariahu, Chinwe & [4]Dennis, Mary
[1&2]*Department of Computer Science, Imo State University, Owerri*
[2]*Department of Computer Science, Imo State Polytechnic, Umuagwu*
[3]*Department of Computer Science, Akanu Ibiam Federal Polytechnic, Unwana Afikpo*

A b s t r a c t

The banking industry has undergone various changes in the way the conduct the business and focus on modern technologies to compete the market. The banking industry has started realizing the importance of creating the knowledge base and its utilization for the benefits of the bank in the area of strategic planning to survive in the competitive market. Data mining is new and powerful technology of extracting useful information from large databases. It helps in improving business decisions, enhance the value of each customer and help in customer requirements. The aim of this paper is to give a view on uses of data mining in detecting bank frauds which can be harmful to bank in future and also to find those transactions which can be useful for bank. This paper provides an overview of the applications of data mining to enhance the performance of some of the core business processes in banking industry and different frauds in banking. The banking industries consist of public industry, private industry and foreign industries, apart from smaller regional and co-operative banks. IT-based banking products, services and solutions are available in market. Phone banking; ATM facility; credit, debit and smart cards; Internet banking and mobile banking; SWIFT Network and INFINET Network are most common in banking, connectivity of bank branches to facilitate anywhere banking.

Keywords: *Data mining, Banking industry, Fraud detection, Phone banking.*

Corresponding Author: Amanze, Bethran

Background to the Study

Data mining can contribute to solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the volume of data is too large or is generated too quickly to screen by experts. The managers of the banks may go a step further to find the sequences, episodes and periodicity of the transaction behavior of their customers which may help them in actually better segmenting, targeting, acquiring, retaining and maintaining a profitable customer base. Business Intelligence and data mining techniques can also help them in identifying various classes of customers and come up with a class based product and/or pricing approach that may garner better revenue management as well. Data Mining is the process of extracting knowledge hidden from large volumes of raw data. The knowledge must be new, not obvious, and one must be able to use it. Data mining has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is "the science of extracting useful information from large databases". Data mining is one of the tasks in the process of knowledge discovery from the database.

In business, scientific domain and in all sectors the need of the database applications has been increased, so the need of data and the storage device are also increased. This explosion in the amount of electronically stored data was accelerated by the success of the relational model for storing data and the development and maturing of data retrieval and manipulation technologies. This technology is used for storing the data as the demand increased; little stress was paid to developing software for analyzing the data until recently. When companies realized that hidden within these masses of data was a resource that was being ignored. The large amount of stored data contains knowledge about a number of aspects of their business waiting to be harnessed and used for more effective business decision support. Database Management Systems used to manage these data sets at present only allow the user to access information explicitly present in the databases. In database the stored data is only a small part of the 'iceberg of information' available from it. Contained implicitly within this data is knowledge about a number of aspects of their business waiting to be harnessed and used for more effective business decision support. This extraction of knowledge from large data sets is called Data Mining or Knowledge Discovery in Databases. It is defined as the non-trivial extraction of implicit, previously unknown and potentially useful information from data. The benefits of Data Mining have resulted in a lot of resources being directed towards its development.

IT has helped the banking industry to deal with the challenges the new economy poses. Nowadays, Banks have realized that customer relationships are a very important factor for their success. Customer relationship management (CRM) is a strategy that can help them to build long-lasting relationships with their customers and increase their revenues and profits. CRM in the banking sector is of greater importance. The CRM focus is shifting from customer acquisition to customer retention and ensuring the appropriate amounts of time, money and managerial resources are directed at both of these key tasks. The challenge the bank face is how to retain the most profitable customers and how to do that at the lowest cost. At the same time, they need to find and implement this solution quickly and the

solution to be flexible. Traditional methods of data analysis have long been used to detect fraud. They require complex and time-consuming investigations that deal with different domains of knowledge like financial, economics, business practices and law. Fraud instances can be similar in content and appearance but usually are not identical. In developing countries like Nigeria, Bankers face more problems with the fraudsters. Using data mining technique, it is simple to build a successful predictive model and visualize the report into meaningful information to the user.

Objectives of the Study
(1)   It help optimize business decisions
(2)   Increase the value of each customer and communication
(3)   Improve customer satisfaction
(4)   It can help banks to gain a clear advantage over its competitors.
(5)   Data mining can help banks in better understanding of the vast volume of data collected by the CRM systems.
(6)   It is a valuable tool by which banks can identify potentially useful information from the large amounts of data.

Data Mining Algorithm and Techniques
There are several data mining techniques and algorithms have been developed and used in data mining like association, classification, clustering, prediction and sequential patterns, Regression, Neural Networks etc. We will briefly examine those data mining techniques.

Classification
Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we make the software that can learn how to classify the data items into groups. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fake and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper judgment. The algorithm then encodes these parameters into a model called a classifier.

Clustering
Clustering can be said as identification of similar classes of objects. This is the technique of combining the transactions with similar behavior into one group, or the customers with same set of queries or transactions into one group. Classification approach can also be

used as effective mean of distinguishing groups. So clustering can be used as preprocessing approach for attribute subset selection and classification. For Example: The customer of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intending to take risks, likewise the same set of service class people in rural areas have a the preferences for some particular brands which may differ from their counterparts in urban areas. This information will help the organization in cross-selling their products, Instead of mass pitching a certain "hot" product, the bank's customer service representatives can be equipped with customer profiles enriched by data mining that help them to identify which products and services are most relevant to callers. This technique will help the management in finding the solution of 80/20 principle of marketing, which says: Twenty percent of your customers will provide you with 80 per cent of your profits, then problem is to identify those 20 % and the techniques of clustering will help in achieving the same.

Predication
Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

Association
Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. Association and correlation is usually used to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.

Applications of Data Mining in Banking Sector
Data Mining can help by contributing in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases. The banking industry across the world has undergone tremendous changes in the way the business is conducted. With the recent implementation, greater acceptance

and usage of „electronic" banking, the capturing of transactional data has become easier and, simultaneously, the volume of such data has grown considerably. It is beyond human capability to analyses this huge amount of raw data and to effectively transform the data into useful knowledge for the organization. Data Mining can help by contributing in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases. By using data mining to analyze patterns and trends, bank executives can predict, with increased accuracy, how customers will react to adjustments in interest rates, which customers will be likely to accept new product offers, which customers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable.

Marketing
One of the most widely used areas of data mining for the banking industry is marketing. The bank's marketing department can use data mining to analyze customer databases. Data mining carry various analyses on collected data to determine the consumer behavior with reference to product, price and distribution channel. The reaction of the customers for the existing and new products can also be known based on which banks will try to promote the product, improve quality of products and service and gain competitive advantage. Bank analysts can also analyze the past trends, determine the present demand and forecast the customer behavior of various products and services in order to grab more business opportunities and anticipate behavior patterns. Data mining technique also helps to identify profitable customers from non-profitable ones. The data mining techniques can be used to determine that how customers will react to adjustments in interest rates, the risk profile of a customer segment for defaulting on loans.

Fraud Detection
Another popular area where data mining can be used in the banking industry is in fraud detection. Being able to detect fraudulent actions is an increasing concern for many businesses; and with the help of data mining more fraudulent actions are being detected and reported. Two different approaches have been developed by financial institutions to detect fraud patterns. In the first approach, a bank taps the data warehouse of a third party and use data mining programs to identify fraud patterns. The bank can then cross-reference those patterns with its own database for signs of internal trouble. In the second approach, fraud pattern identification is based strictly on the bank's own internal information. Most of the banks are using a hybrid approach.

Customer Relationship Management
Data Mining can be useful in all three phases of customer relationship cycle: Customer Acquisition, Increase Value of the customer and Customer Retention (Prakash *et al.*, 2012; Ping and Liang, 2010). Financial organizations especially banking sector recruits Relationship Managers or team of executives to pay proper attention to their customers. Due to the tight competition exists in the market (Sergio *et al.*, 2011; Wang *et al.*, 2009; Chen *et al.*, 2009), customers will always with banks which provide better facility and more secured transaction option. Data Mining techniques (Prakash *et al.*, 2012; Wikum *et al.*, 2009) can be used to determine the list of customers as per the set of definitions (Sergio *et*

al., 2011; Wang *et al*., 2009; Corne *et al*., 2012) and interest and the institution can offer better facilities to them (Abdullah and Titus, 2010) customers are varying from their approach in banking, like certain customers interested only electronic banking while others want banking through the counter. Classifying such customers can easily done using data mining techniques and provide better facilities. Data mining can be used to find out customers holding one product (Wu and Chou, 2011) having interest in similar to other one, there by promoting the product which benefits the organization. Not only can data mining help the banking industry to gain new customers, it can also helps to maintain the existing customers with better service (Tremblay *et al*., 2010; Kontonasios *et al*., 2012; Liu *et al*., 2012). Within the context of Customer Relationship Management (CRM), data mining can be seen as a business driven process aimed at the discovery and consistent use of profitable knowledge from organizational data (Wu and Chou, 2011). It can be used to fasten the decision making and guide to forecast the effects of decisions (Prakash *et al*., 2012). Data Mining can be used to increase the response rate of marketing campaign. This can be done by segmenting the customers into groups with their needs and characteristics, it can predict how likely an existing customer is to take the business to a competitor (Mimno, 2012). Each of the CRM elements can be supported by different data mining models (Vaillancourt, 2010; Abdullah and Titus, 2010; Herawan and Deris, 2011; Akbar *et al*., 2010; Shinde, 2012; Delamaire *et al*., 2009) which generally include classification, association, sequence discovery, clustering, regression, forecasting and visualization.

Risk Management
Data Mining is used to identify the risk factors in each department of banking business (Moradi *et al*., 2013). Credit Approval authorities in the financial organization used data mining techniques to determine the risk factors in lending decisions (Chen *et al*., 2009; Chen and Huang, 2011) by analyzing the data based on nationality, repayment capacity and so on. Retail marketing department uses data mining methodologies to find the reliability (Yap *et al*., 2011) and the behavior of credit card applicant (Delamaire *et al*., 2009) while selling the credit cards. They uses data mining techniques on existing customers to sell credit cards or increase customers credits or top up on credit card loans (Bhattacharya *et al*., 2011). In commercial lending, data mining plays a vital role.

In commercial lending, risk assessment is usually an attempt to quantify the risk of default or loss to the lender when making a particular lending decision or approving a credit facility (Chen and Huang, 2011). Here credit risk can be quantified by the changes in the value of credit products or of a whole credit customer portfolios, which is based on changes in the high risk tendency, default probability, instrument's rating and recovery rate (Yap *et al*., 2011; Ravisankar *et al*., 2011; Liu *et al*., 2012) of the instrument in case of default. The major part of implementation and care of credit risk management system (Raj and Portia, 2011) will be a typical data mining problem: the modeling of the credit instrument's value through the default probabilities, recovery rates and rating migrations (Fung *et al*., 2010). Data Mining can be used to derive credit behavior (Delamaire *et al*., 2009) of individual borrowers with parameters card loans, mortgage value, repayment and using characteristics such as history of credit, employment period and length of residency. A score is thus produced that allow a lender to evaluate the customer and decide whether

the person is a good candidate for a loan, or if there is a tendency to become high risk of default (Raj and Portia, 2011). Customers who have been with bank for a longer periods of time, remained better with bank and have good credit history and have higher salaries/wages, are more likely to receive a loan than a new customer who has no credit history with the bank, or who earns low salaries/wages (Ravisankar *et al.*, 2011). Bank can reduce the risk factors to maintain a better position by knowing the chances of a customer to become default (Dorr and Anne, 2009; Tsai, 2012).

Money Laundering Detection

Money Laundering is the process of hiding the illegal origin of "black" money so as to legitimize it (Khac and Kechadi, 2010). Banks are commonly used as channels to launder money. Therefore governments and financial regulators require banks to implement processes, systems and procedures to detect and prevent money laundering transactions. Failure to detect and prevent such illegal transactions can invite hefty fines both monetarily and operationally which can prove very costly for the bank and even can make its survival difficult. Conventional rule-based transaction analysis based on reports and tools will not be sufficient to detect more complicated transaction patterns like smurfing and networked transactions (Khac *et al.*, 2011). Here data mining techniques can be applied to dig out transaction patterns that can lead to money laundering. Typically such systems take client risk assessment data, transaction risk measurement data and patterns and behavior patterns into consideration for detecting money laundering patterns. Transactions are then grouped into clusters based on their similarities found in these chosen attributes (Khac *et al.*, 2011). In a large database of banking transactions, it is possible that a huge number of patterns emerge and will be classified as money laundering transactions thereby increasing false positives. Statistical false reduction methods based on decision tree classification are employed to limit the number of false patterns detected (Anuar *et al.*, 2008).

Investment Banking

Investment is an action of investing money into an asset or item for profit/income. Banks often offer investment services to their customers. There are a vast number of financial instruments in the market. Data mining like K-means clustering can be applied to choose the best investments based on customer's profile (Ingle and Meshram, 2012). Capability to predict asset prices (for example stock prices) from historic prices can increase returns from investment tremendously. Data mining techniques for prediction like neural networks and linear regression can be employed for prediction of prices for stocks (Naeini *et al.*, 2010). Data mining can also be applied in time series analysis for financial applications (Tak-chung, 2011).

Conclusion

Banks use data mining in various application areas like marketing, fraud detection, risk management, money laundering detection and investment banking. The patterns detected help the bank to forecast future events that can help in its decision-making processes. More and more banks are investing in data mining technologies to be more competitive. Data Mining techniques can be of immense help to the banks and financial

institutions in this arena for better targeting and acquiring new customers, fraud detection in real time, providing segment based products for better targeting the customers, analysis of the customers' purchase patterns over time for better retention and relationship, detection of emerging trends to take proactive approach in a highly competitive market adding a lot more value to existing products and services and launching of new product and service bundles.

References

Vivek, B. (2011). Application of data mining in banking sector. *International Journal of Computer Science and Technology, 2* (2), 199-202.

Sreekumar, C. & Balan, D. (2013). Data mining in banking and its applications-A review. *Journal of Computer Science, 9* (10), 1252-1259.

Vikas, B. & Balan, D. (2013). A review on Data mining in banking sector. *American Journal of Applied Sciences, 10* (10), 1160-1165.

Chitra, H. & Subashini, R. (2013). Data mining techniques and its applications in banking sector. *International Journal of Emerging Technology and Advanced Engineering, 3* (8), 219-226.

Rajdeepa, N. & Nandhitha, K. (2015). Fraud detection in banking sector using data mining. *International Journal of Science and Research, 4* (7), 1822-1825.

Kazi, D. & Qazi, R. (2012). Use of data mining in banking. *International Journal of Engineering Research and Applications, 2* (2), 738-742.

Abdullah, Z. & Titus, H. (2010). Mining significant least association rules using fast SLP-growth algorithm. *Advanced Computer Science Information Technology, 60* (59), 324-336.

Akbar, S., Rao, K. N. & Chandulal, J. A. (2010). Intrusion detection system methodologies based on data analysis. *International Journal Computer Application, 5,* 10-20.

Chen, I., L. Chi-Jie, L., Tian-Shyug &. Chung-Ta, L (2009). Behavioral scoring model for bank customers using data envelopment analysis. *Stud. Comput. Intell*., 99-104.

Corne, D., Clarisse, D. & Laetitia, J. (2012). Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal Operational Research, 2* (21), 469-479

Delamaire, L., Hussein, A. & John, P. (2009). Credit card fraud and detection techniques: A Review. *Banks Syst., 4,* 57-68.

Dorr, D. M. & Anne, M. D. (2009). Establishing relationships among patterns in stock market data. *Data Knowl. Eng., 68,* 318-337.

Fung, B.C., Wang, K., Chen, R. & Yu, P.S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surveys.*

Bhattacharya, S., Jha, S., Tharakunnel, K. & Westland, J.C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Syst. 50,* 602- 613

Hu, Y.C. & P.C. Liao, (2011). Finding critical criteria of evaluating electronic service quality of internet banking using fuzzy multiple-criteria decision making. *Applied Soft Comput., 11,* 3764-3770.

Kontonasios, K. N., Spyropoulou, E. & Bie, T. D. (2012). Knowledge discovery interestingness measures based on unexpectedness. *Wiley Interdisciplinary Rev, Data Min. Knowl. Discovery, 2,* 386-399.

Mimno, D. (2012). Computational historiography: Data mining in a century of classics Journals. *Journal of Computer Cultural Heritage*

Moradi, M., Salehi, M., Ghorgani, M. E. &. Yazdi, H. S. (2013). Financial distress prediction of Iranian companies by using data mining techniques. *Organizacija, 46,* 20-27.

Ping, Z. L. & Liang, S. Q. (2010). *Data mining application in banking-customer relationship management.* Proceedings of the IEEE International Conference on Computer Application and System Modelling, Oct. 22-24, IEEE Xplore Press, Taiyuan, pp: 124-126.

Prakash, B.V. A., Ashoka, D.V. & Aradhya, V.N.M., (2012). Application of data mining techniques for software reuses process. *Proc. Technol., 4,* 384-389.

Raj, S.B.E. & Portia, A. A. (2011). *Analysis on credit card fraud detection methods.* Proceedings of the International Conference on Computer Communication and Electrical Technology, Mar. 18- 19, IEEE Xplore Press, Tamilnadu, pp 152-156.