INTERNATIONAL JOURNAL OF ENGINEERING AND MANAGEMENT SCIENCES

www.scienceandnature.org

DATA MINING TECHNIQUES FOR KNOWLEDGE DISCOVERY FROM FINANCIAL INSTITUTION DATABASE

[1]Abdulsalam S. O., [1]Adewole, K. S., [2]Bashir, S.A., [3]Jimoh, R.G. & [4]Olagunju, M.
[1]Department of Computer Science, Kwara State University, Malete, Nigeria.
[2]Department of Computer Science, Federal University of Technology, Minna, Nigeria.
[3]Department of Computer Science, University of Ilorin, Ilorin, Nigeria.
[4]Department of Computer Science, Kwara State Polytechnic, Ilorin, Nigeria

ABSTRACT
Due to its importance for the investment decision-making and risk management, describing and predicting stock represents a key topic in stock analysis. Stock or shares are valued and analyzed by stock investors using fundamental analysis and technical analysis. Stock investors describe and predict stock manually based on individual experiences. Employing manual procedure in analyzing stock, most especially technical analysis is always very cumbersome and inefficient; because of much time that is consumed in examining the past records of a respective firm, an investor is willing to invest in. This paper presents a better way of describing and predicting stock, especially technical analysis, by employing data mining techniques. A database was developed employing 360 records of daily activity summary (equities) and 78 records of weekly activity summary (equities) spanning through 18 months that is, from January 2007 to June 2008. These data were obtained from the daily official list of the prices of all shares traded on the stock exchange published by the Nigerian Stock Exchange being the financial institution that runs Nigerian stock market; using banking sector of Nigerian economy with three banks namely:- First Bank of Nigeria Plc, Zenith Bank Plc, and Skye Bank Plc. A data mining software tool was developed and employed in identifying patterns and relationships from the database to generate new knowledge about the data set in the database through the use of data mining techniques that employ regression analysis.

KEYWORDS: Data Mining, Stock Exchange, Financial Institution, Decision-Making, Risk Management, Regression Analysis

INTRODUCTION

In all large enterprises irrespective of whether it is a manufacturing organization, a commercial organization, service providers, university or any other enterprises, a large amount of data is collected over a period time and kept in files. In most of the cases, files of the current year or perhaps one or two previous years are generally referred to very frequently and files belonging to distant past are just bundled up and stored. They are rarely opened and their contents are considered only of historical importance. In recent times it is gradually dawning on the organizations that a wealth of information is contained in these files. A deeper study of such historical material would give us certain relations between items, individuals and events which have not been suspected earlier. Organization stored these vast amounts of information in a proper manner thereby forming a database, so as to make the retrieval of information simple, quicker and reliable (Ravindranath, 2003). In the database, the items of information, which have some link with each other, are generally stored in such a manner that the relations can be easily established and necessary information obtained quickly. From the developed database, data mining techniques can be used for knowledge discovery. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information – information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in a large database (Ravindranath, 2003).

We examine the use of data mining technique in stock exchange, as an analytical tool to describe and discover knowledge about stock market returns. The uncover patterns of behaviour derived from the use of data mining technique serve as a tool to make complex decision by the investors in the stock market so as to reduce risk in investment.

The act of investing in stocks unarguably poses series of risks to the investor as long as the investment lasted. Total risk associated with stock investment is defined as the variation in returns (of range of probable incomes) which result from undertaking an investment (Hayes, 1997). The technical analysis which is purely based on trend – market trend interpreted through graphs and charts is employed in valuing and analyzing shares. Trend analysis describes the serial movement of stock prices over time which can be analyze through graphs and charts. The serial movements of stock prices over a period of time extracted from the daily official list of Nigerian Stock Exchange are used in developing a large database and data mining tool which employed the use of regression analysis was used in discovering meaningful new patterns. However, in the context of this paper, the banking sector with three banks namely – First Bank Nigeria PLC, Zenith

Bank PLC and Skye Bank PLC are taken into consideration; using the uncover patterns and relationship to describe and discover knowledge about values of other variables in the database.

REVIEW OF STOCK MARKET AND DATA MINING CONCEPTS

The Stock Market and Stock Exchange

A stock market (or capital) is a place where stocks, bonds, and other securities are traded. A stock exchange is the body that runs a stock market. Some stock markets are not run by any major body, but are coordinated by their dealers (Nwawu, 2004). These are termed over the counter (OTC) e.g. the NASDAQ in America. The stock market in Nigeria is run by the Nigerian Stock Exchange (NSE).

The stock exchange helps companies generate capital. As a primary market, it provides an avenue for them to sell new shares and bonds to investors. The companies can then use the proceeds from these sales to expand their businesses, develop new products, buy new equipment etc. The stock market also provides a means for investors to trade in the shares of companies they own among themselves. In other words, it serves as a secondary market. For example, one who bought the shares of a company at a particular price may sell it to another investor. The investors are the one who profit from this type of trade – companies do not.

The stock exchange also has the function of upholding rules and regulations so that shady people do not cheat investors of their hard earned money. It gives investors security.

How to Read a Share – Price Column

Using Table 1.1 which is a reprint of part of the Banking sector section of the daily official list, as one could find it in Financial or Business Times. Go down to the entry First Bank. The first column you will find is the **public quotation price**

Table 1.1: The Share price columns

Ordinary Shares	Public Quotation Price(₦)	Current Market Price(₦)		Ex Div	Ex Sc	Dividend			EPS	PE Ratio
						Date Pd	Inter	Final		
SKYE BANK PLC	0.50	19.00	+	×	×	7/2/2007		1.12	1.10	17.27
ZENITH BANK PLC	0.50	22.67	+			12/2/2007		1.07	1.14	19.89
FIRST BANK OF NIG. PLC	0.50	28.75	+			4/8/2006		1.50	1.19	24.2

Business Done			This Year		Last Ex Date	Last Ex-Sc Date
Price(₦)	Date	Quantity				
18.96	18/10/07	69,301	22.47	17.88	18/01/07	18/01/07
23.17	18/10/07	1,158,341	26.45	20.39	20/01/07	
29.39	18/10/07	15,763,363	31.00	20.00	15/07/06	15/07/06

Against First Bank is written 50k, which is the par value of the share. Next you see the current market price of ₦28.75; the current market price column indicates the last agreed market price for the shares. After that, you see a + against the share. A + indicates that there were more buyers than sellers of that particular share in the stock market, thus the share is likely to rise in the near future. Conversely, a – indicates that there were more sellers than buyers. Next you will find no cross in the **ex-dividend** and **ex-scrip** column. A cross in the ex-dividend column means that if you buy shares in First bank you will not be entitled to the recently declared dividend. Likewise a cross in the ex-scrip column means you will not receive the recently declared scrip issue.

Next you see the **business done** column. This gives you the details of the last transaction carried out on that particular share – the price at which the share was traded; the date and the quantity of the transaction are listed. We see that 5,763,363 First bank shares were traded at a price of ₦29.39 on 18/10/07. This column helps us know whether a share is being actively traded or not. A share which has a very old business done date means that nobody wants to buy it. These shares, which are difficult to sell, are almost 'dead'.

The next column shows **'This year's high, low'**. This shows the highest and lowest prices of the share in the last 52weeks. We can see that in the 1year period, the share price of First Bank reached a max of ₦31 and low of ₦20. Some make a decent profit by buying at the low price and selling at the high price in a year cycle.

The **last ex-div** date shows the date after which any shares you buy will not take part of the recently declared dividend. The **last ex-scrip** date shows the date after which any shares you buy will not take part of the recently declared scrip issue i.e. the scrip issue belongs to the seller after that date. In the dividends column, we see the **dividends date paid**. The dividends date paid is the date when you will actually be sent dividends. First Banks last dividend was paid on 4th August, 2006.

The next two columns show **interim** and **final dividend** payments. We see that no interim dividend was paid by First Bank but a final dividend of ₦1.50 per share was paid by the company.

Next is the **EPS** or earning per share which is a very important figure. It tells us that a share selling at ₦20 earns ₦1.19 for you, the investor every year. At this rate, it will take an investor about 24.2 years to recoup his investment. This 24.2 is the **price earnings (P.E) ratio**. It compares the price of the share to how much it can earn for you. Companies with higher PE ratio have chances of higher growth performance and profit.

The formula for calculating PE ratio is as follows:

PE = Current Market Price / Earnings Per Share (EPS) 1.1

Data Mining and its Economic Use

Translating Data mining word by word means, the mining or digging in data with the purpose of finding information or respectively knowledge. Coming to the more abstract and very well known definition of Frawley, Data Mining is defined as "*The nontrivial extraction of implicit, previously unknown, and potentially useful information from data*" (Frawely et al., 1992).

Groth (1988) mentions another interesting aspect of Data mining. He describes it as "*the process of automating information discovery*". Today Data mining is a term that covers a broad range of techniques to analyze data. The techniques use specific algorithms to identify and extract patterns and establish unknown relationships in order to discover hidden and valuable information in a huge amount of data. Most companies already collect massive quantities of data. Data mining techniques can be implemented on existing software and hardware platforms to enhance the value of existing information resources (Thearling, 2005).

In the words of Moxon (1996) "*Data mining is the process of discovering meaningful new correlation, patterns and trends by sifting through large amounts of data, using pattern recognition technologies as well as statistical and mathematical techniques*".

Data mining techniques can be considered to be descriptive or predictive. Descriptive data mining intends to summarize data and to highlight their interesting properties, while predictive data mining aims to build models to forecast future behaviours (Han and Kamber, 2001).

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information – information that can be used to increase revenue and cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from different dimensions or angles, categorise it and summarize the relationships identified.

KDD – Knowledge Discovery in Databases

Knowledge Discovery in Databases, also often used with the abbreviation KDD, "*is the concept of extracting previously unknown and potentially useful information from large sets of data*" (Witnessminer, 2005). So KDD is only the concept of a multistage process that identifies pattern in data in order to find new information. Data mining is only one stage in the KDD process as shown in figure 1.1 and it is concerned with applying computational techniques to find patterns in data. This step consists of algorithms which delivers patterns in an acceptable time out of a defined database. Other stages in the KDD process are the comprehensibility and the validity of the discovered patterns. In theory and practice the expressions KDD and Data mining are often mixed. But it is important to understand that KDD is the whole concept and Data mining is only a step in this concept of extracting data. Simplified, KDD is the concept and Data mining is the tool (Witnessminer, 2005).The five main processes that are common in almost all of the methods are: Task Analysis, Pre-processing, Data mining, Post-processing and Deployment as shown in figure 1.2.

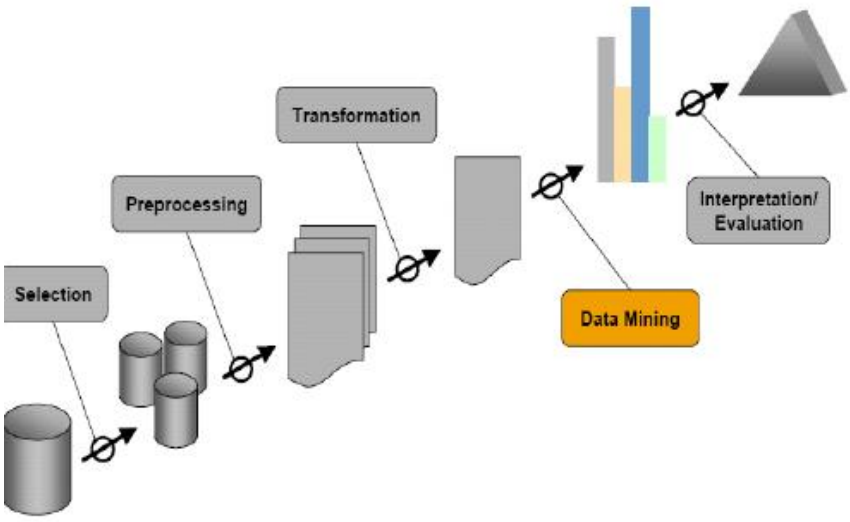

Figure 1.1: Knowledge Discovery in Databases (Source: Lesley, 2004)

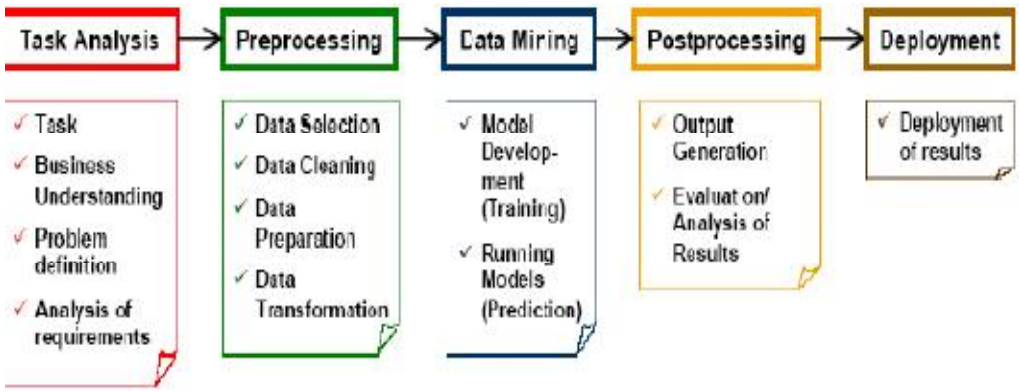

Figure 1.2: Knowledge Discovery in Databases (Source: Lesley, 2004)

Data Mining Functions

The main goals of data mining are prediction (using values of variables in the data set to predict future values of other variables) and description (identifying patterns and relationships in data, and generates new knowledge about the data set). Based on the above explanations, the basic data mining tasks are divided into predictive and descriptive. Regression, time series analysis and prediction which fall under predictive task are briefly discussed as follows:

Regression

It is the data mining functionality that maps a data item to a real value prediction variable, using a function. The data item is assumed to fit into a type of function (linear, logistic), and using error analysis the best function to model the data is discovered. Multiple regressions are used when more than one predictor variable is involved. Nonlinear regression can be transformed to a linear model. Logistic regression is used for modeling the probability of some event occurring as a linear function of a set of predictor variables. Log-linear regression is used to estimate discrete multidimensional probability distributions.

Prediction

It is a type of classification, but it is predicting future state instead of current state, based on past and current data. Some examples of applications are speech recognition and pattern recognition. The future values can be predicted using time series analysis or regression techniques. When it deals with time series data, this task is also called forecasting; when it deals with finding the numerical value of a variable, it is called regression. Some of the most common predictive techniques are: Regression analysis, Regression trees, Neural networks, K Nearest Neighbour, Box-Jenkins methods and Genetic algorithms. For instance, an international company's future income can be predicted using variables such as exchange rate, inflation rate and so on. Also, the gas consumption of a power plant can be forecast based on daily temperature and the day of the week.

Time series analysis

It is a data mining function that examines how the value of an attributes varies over time. The values are taken at evenly spaced points in time (like weeks, months, years). In general, the more values that are considered in developing the model, the better it is for its accuracy. In order to visualize the time series, a time series plot can be used. The most common example of using time series data mining is analyzing the trends of stock market prices.

METHODOLOGY

To achieve the objectives of this research, necessary data were obtained from the daily official list of the prices of all shares traded on the stock exchange published by the Nigerian Stock Exchange using three banks namely:- First Bank of Nigeria Plc, Zenith Bank Plc and Skye Bank Plc from the banking sector of Nigerian economy. Microsoft Access Database Management System was used to develop a database based on the collected data. The data mining software tool was developed using Microsoft Visual Basic Programming Language. This tool explores data mining technique that uses regression analysis to identify patterns and relationships in the data collected in order to generate new knowledge about the data stored in the database.

In this paper, we use regression analysis data mining method for stock market returns description. The discovered patterns and relationships were used to describe the market trends of the stock prices which were interpreted through graphs and charts. Crystal Report software was used to plot the graphs, draw of the charts and generate other forms of reports.

Linear regression

Linear regression is one of the most common methods for predicting the future value of variable based on the linear relationship it has with other variables. Basically, it assumes there is a straight line that approximates the data set, and bases the forecast on it.

There is only one independent variable and the formula that describes this relationship is the one that defines a straight line:

$$y = a + bx \qquad \text{---------- 1.2}$$

Where, y is the dependent variable
x is the independent variable
a and b are the line's coefficients

Moreover, the least square method was employed in finding the regression line and it is expressed as follows:

$$y = a + bx \quad \text{---------- 1.3}$$

Where, $a = \bar{y} - b\bar{x}$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} \quad \text{---------- 1.4}$$

In relating to this paper, each variable denotation is as follows:

y = current market price
x = percentage earning (P.E.) ratio

$\bar{y}$ = Mean occurrence of current market price
$\bar{x}$ = Mean occurrence of percentage earning (P.E.) ratio
n = Total number of occurrence of the variables

The above variables were derived from the relationship between the percentage earning (P.E.) ratio, current market price and the earnings per share (EPS) as earlier shown in equation 1.1.

The various operations that can be carried out with the data mining software tool are shown in the system flowchart below:

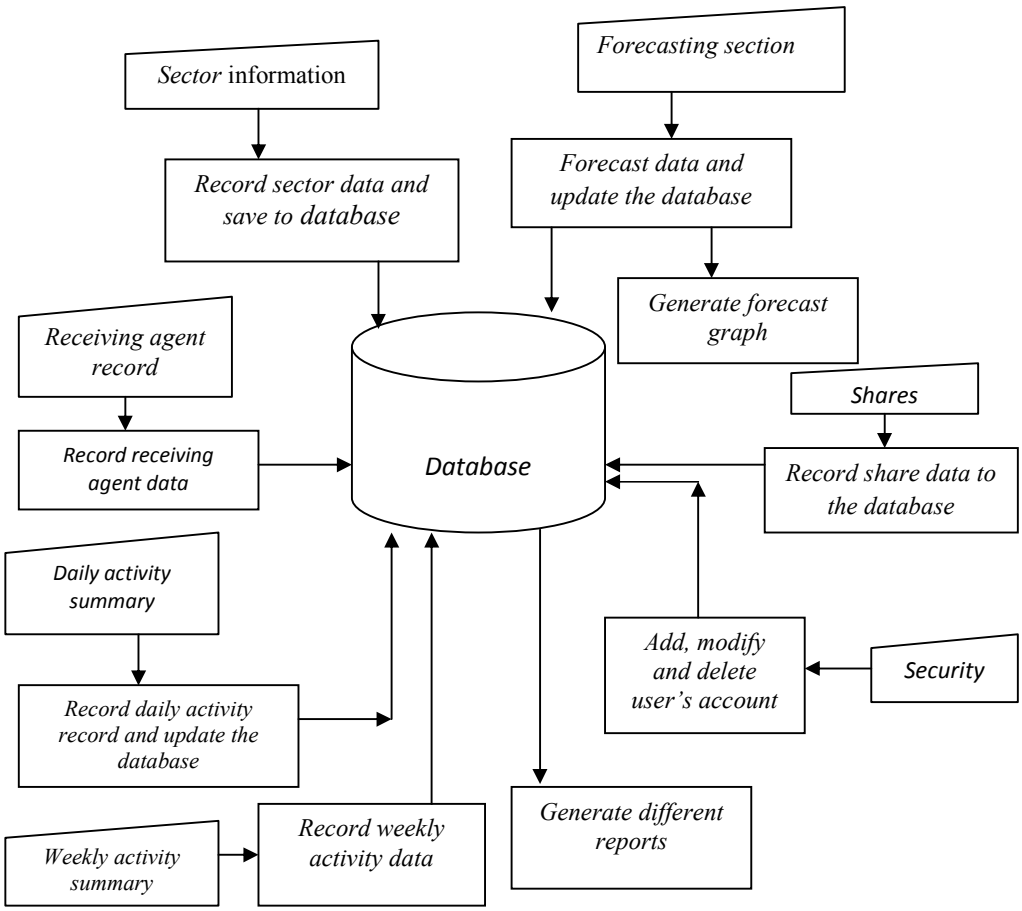

Figure 1.3 System flowcharts for the Data Mining Software Tool

RESULTS AND DISCUSSION

Figure 1.4 shows the overall Daily official list records as specified in the Nigeria Stock Exchange Daily official list for each bank on 21st March, 2008 (Financial Standard, March, 2008). It can be seen from this figure that the P.E. ratio of First Bank is higher than that of Zenith Bank and the Zenith Bank's P.E. ratio is higher than that of Skye Bank. This simply means that, the possibility of First Bank having the highest value in the high column is guaranteed since it has the highest P.E. ratio. This can be seen from the high column where First Bank has 54.36, Zenith Bank 51.00 and Skye Bank 19.78. One would have concluded that investing in First Bank stock is better than that of the two other banks at this period of time. But in Figure 1.5, where Data mining technique that employed regression analysis as described in equations 1.1, 1.3 and 1.4 was used to uncover pattern and trends of stock market returns of these three banks. It was discovered that investing in Zenith Bank stock would be the most preferred to any of

the other two banks has revealed by the trends of the stock chart.

NIGERIAN STOCK EXCHANGE (NSE)

NSE DAILY OFFICIAL LIST

AS AT: 21/03/2004

Month: MARCH Week: 3 Day: FRI Year: 2008

Ordinary Share	Public Quot Price(N)	Current Market Price(N)	Price(N)	Quantity	High	Low	Last Ex-Div Date	Last Ex-Sc Date	Date Paid	Final	E.P.S	P.E Ratio
FIRST BANK PLC	N0.50	N46.08	N46.08	79,452,780	54.96	40.53	07/06/2008	07/06/2008	07/06/2008	1.00	0.75	31.14
ZENITH BANK PLC	N0.50	N50.00	N50.00	7,639,535	51.00	46.09	07/06/2008	07/06/2008	07/06/2008	1.10	1.08	23.26
SKYE BANK PLC	N0.50	N16.84	N16.84	18,346,195	19.78	16.84	07/06/2008	07/06/2008	07/06/2008	0.42	0.76	16.67

Figure 1.4: Daily Official List Report of the three banks (Source: Financial Standard, 2008)

MARKET PRICE GRAPH

FOR THE YEAR

2007-2008

Figure 1.5: Market Price Graph showing the Market Trends of Prices of the three banks over a period of time

To further complement the result shown in figure 1.5 which is in form of line graph, a quarterly bar chart was plotted to show the trends of the stock market returns of the three banks in figures 1.6 and 1.7. Data mining also revealed from these two figures that investing in Zenith Bank stock in the most preferred to any of the other two banks.

MARKET PRICE GRAPH

FOR THE YEAR

2007

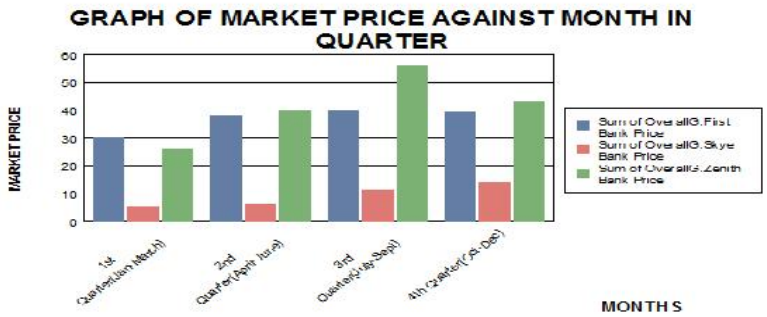

Figure 1.6: Bar chart showing the Market Trends of Prices of the three banks over a period of time

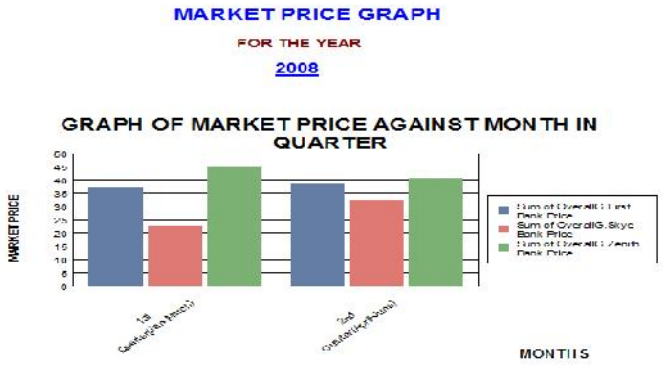

Figure 1.7: Bar chart the market price for the three banks in quarterly format for the year 2008

CONCLUSION

In this paper, we employed data mining techniques that use regression analysis to uncover pattern and relationship from financial institution database. These uncovered patterns are used to describe and discover knowledge about the values of other variables in the database, which will allow investors to better their choice of investment when buying and selling stock in financial market in Nigeria.

REFERENCES

Moxon, B. (1996). *Defining Data Mining, The Hows and Whys of Data Mining, and How It Differs from Other Analytical Techniques*: Online Addition of DBMS Data Warehouse Supplement.

Lesley, C. (2004). *A Presentation on Data Mining with SAP BW 3.5. SAPNET.*

Financial Standard. Available at: http:// www.financialstandardnews.com. [Accessed 07 May 2008].

Hand, D., Manilla, H. and Smyth, P. (2001). *Principle of data mining*: Bradford Books.

Hayes, M. (1977). *Irwin Guide to Stock Market Cycles*. Illinois, USA: Dow Jones – Irwin Homewood.

Nwawu, C. I. (2004). *The Beginner's Guide to Investing in the Nigeria Stock Market*. Available at:

http:// www.stockmarketnigeria.com/forums. [Accessed 05 May 2008].

Han, J. and Kamber, M. (2001). *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publishers.

Kantardzic, M. (2003). *Data mining: Concepts, Models, Methods and Algorithms*. Wiley: IEEE Press.

Thearling, K. (2005). *An Introduction to Data Mining: Discovery hidden value in your data warehouse*. Available at http://databases.about.com/gi/dynamic/offsite.htm?site=htt p%3A%2F%2Fwww.thearling.com%2Ftext%2Fdmwhite%2Fdmwhite.htm. [Accessed 13 August 2008].

Chapple, M. (2005). *Data Mining: An Introduction*. Available at: http://databases.about.com/library/weekly/aa100700a.htm. [Accessed 13 August 2008].

Groth, R. (1988). *Data Mining: A Hands-on Approach for Business Professionals*. Upper Saddle River, New Jersey: Prentice Hall.

Frawley, W., Piatetsky-Shapiro, G., and Matheus, C. (1992). *Knowledge Discovery in Databases: An Overview*. Fall: A1 Magazine.

Ravindranath, B. (2003). *Decision Support Systems and Data Warehouses*. New Delhi: New Age International Limited. ISBN: 81-224-1454-0.

Witnessminer (2005). *KDD*. Available at: http://www.witnessminer.com/kdd definition.htm. [Accessed 13 August 2008].