REVIEW ARTICLE

A Review of Data Mining as a Tool for Organization's Growth and Productivity

Oluigbo Ikenna[1], Nwakanma Ifeanyi[2], Roland Justina[3], Nwachukwu Nnamdi[4], Ujunwa Anuli[5], Ndanike Paul[6]

[1-6]Department of Information Management Technology, Federal University of Technology Owerri, Imo State, Nigeria

[1] ikenna.oluigbo@gmail.com, [2] fraircos@yahoo.com, [3] just2009great@yahoo.com,

[4] skipnroll@yahoo.com, [5] anulilynda@yahoo.com, [6] agabusinlondon@gmail.com

---

Abstract - The way at which organizations collect and use data, especially for decision making, is growing at a phenomenal rate. There is therefore the need to properly analyse these data into more sophisticated information so that it can be of great value. Data Mining, also referred to as Knowledge Discovery, uses different techniques to analyse data from different forms or patterns and summarize it into valuable information; so as to increase competitiveness, efficiency and market value of private and public organizations. Organization Growth is a necessity to meet the demands of an increasingly dynamic and complex environment. Knowledge discovery and evidence-informed decision making are instrumental in organizational growth. This paper studies the impact of implementing data mining as a strategic tool for organization's growth and productivity. Areas of successful Data mining application were also introduced. Finally, issues that inhibit effective implementation of data mining in an organization were discussed.

Keywords: Data Mining, Knowledge Discovery, Organizational Growth, Data mining techniques, Data mining Applications

---

I.   INTRODUCTION

Storing information in a data warehouse alone does not provide all the benefits to drive an organization forward. In most cases, it is often necessary to extract the hidden knowledge repository within the data warehouse. However, as an organization grows, so does the complexity of its data warehouse; thus it becomes increasingly difficult (if not impossible) and very tedious for such organization to identify behaviours and relationship in the data.

Such organization might be faced with these pressing problems:

1. Abundance of data in the data warehouse and need for more powerful data analysis tool
2. "Data tombs" – a situation where data archives are rarely visited
3. Decisions are made not on the abundance of data, but on a decision maker's intuition
4. No tool to extract data embedded in vast amount of data
5. Users or domain experts manually inputs knowledge; which is time consuming, costly and prone to biases and errors. [1]

Data mining is one of the best ways for an organization to analyse data using software techniques for finding hidden and unexpected pattern and relationship from a huge amount of data.

According to reference [2], Data mining is a knowledge discovery using a sophisticated blend of techniques from traditional statistics, artificial intelligence and computer graphics. Simply put, Data mining is the process of extracting valid, previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions.

Basically, the goals of data mining are threefold [2];

1. Explanatory: To explain some observed event or condition, such as why marketing of foot wears have increased in Nigeria
2. Exploratory: To analyse data for new or unexpected relationships, such as finding employees with similar job description

3. Confirmatory: To explain a hypothesis, such as whether female employees in an organization are more likely to be highly motivated than their male counterparts.

Data mining can provide huge paybacks for companies who have made significant investment in data warehousing. Data mining is a technology that has already been integrated into business operations of organizations for their growth and development.

1.1   RELATED WORK

Reference [3] worked on the role of data mining in the successful implementation of Customer Relationship Management (CRM). They described the success factors of CRM as Change Management, Operation Management, Knowledge Management, Marketing Management, Customer Support Management, Sales Management, Technology Management and Project Management.

Reference [4] worked on the role of data mining in strategic management. They analysed the problems involved with implementation data mining as a strategic management tools. The problems include Data quality problems, Training the Staff, Applying DM Tools to Business Problems Blindly, The Environment May Change over Time, and Difficult to Convince Management.

II.   DATA MINED IN AN ORGANIZATION

Whether it is a simple quantitative data such as employee's salary and age, or complex qualitative elements such as multimedia and hypertext documents, different organization mine various kind of data for its productivity and growth. According to [5], the kind of data that can be mined from an organization includes but not limited to the following;

2.1   Business Proceedings: Transactions can either be between one or more organizations, or within an organization. The type of business transactions can be for marketing, sale, purchases; banking e.t.c. channelling an organization towards carrying out the right business proceeding at any point in time can help boost productivity, hence increasing organizational growth

2.2   Memos and reports: Organizations collects and exchanges memos and reports in textual form either between or within organizations. These memos and reports are usually exchanged through E-mails. These messages are stored for future use to build a strong digital library and mined whenever needed.

2.3   World Wide Web Repository: World Wide Web is a large repository of multimedia and hypertext document, covering a broad variety of topics covered and the infinite contributions of resources and publishers. Even though WWW can be redundant and inconsistent sometimes, information mined from the WWW can be of great importance to improve organization productivity.

2.4   Digital Media: The ease, economy of scale and scalable nature of digital information have increased its wide spread use in organization. Most organizations are beginning to convert their data repositories into digital format.

2.5   CAD and Software engineering data: There are a number of Computer Aided Software Engineering (CASE) tools which organizations can use to design their databases and data warehouse. These tools generates large amount of objects, codes and function libraries which needs to be properly managed and maintained.

III.   HOW ORGANIZATIONS MINE DATA

The choice of an appropriate method to mine data depends on the nature of the data to be analysed, as well as the size of the data set. An organization might choose to perform data mining against either the data marts (a small-scale warehouse) or the enterprise data warehouse or both. Several different techniques (also known as Data mining task) can be used to mine data, with each technique having its own choice of operation for implementation. These operations include Predictive Modelling, Database segmentation, Link analysis, and Deviation detection. The technique used and results obtainable includes;

3.1   Regression: it is widely used to map data item to a real valued prediction variable. In simple terms, regression tests or discovers relationships between historical data (dependent and independent variables) in the data warehouse. The types of regression model used may include Linear Regression, Multivariate Linear Regression, Non Linear Regression, and Multivariate Non Linear Regression.

Result obtainable: Successful implementation of Regression will help in successfully forecasting business context (i.e. determining the increase or decrease trends relative to the dependent and independent variables.

3.2   Decision Tree Induction (DTI): This technique test or discovers if certain conditions are obeyed, and then rules for decision propensity. Basically, Decision trees are tree-shaped structures that represent sets of decisions. These decisions generate guidelines for the classification of a dataset.

Result obtainable: DTI as a Data Mining technique serves as a support system that empowers managerial decision making process in an organization.

3.3   Clustering: Clustering technique can be used alongside signal processing to discover similar subgroups and segments, or similar classes of objects. According to [6], clustering techniques can help further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.

Result obtainable: Clustering can help organize business functions in hierarchy; by maximizing the similarity between business functions of the same hierarchy and minimizing the similarity between business functions of different hierarchy.

3.4   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. 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 [6]. Classification can be used with Clustering, Regression or with DTI.

Result obtainable: Successful implementation of Classification can help an organization determine the behavioural pattern of their employees or customers, and grouping those employees/customers with similar behaviour together.

3.5   Neural networks: This technique develops predictive models based on principles modelled after the human brain. It resembles the biological neural network in structure, and it has the ability to detect meaning and trends from complicated data, that are too complex to be noticed by human or computer technique.

Result obtainable: Since neural networks learn through training, many organizations have deployed neural networks in solving real business problems.

3.6   Fractals: This is a data mining technique that is used to compress large databases into simpler structures without losing information. This helps reduce memory requirements to store information, and saves time in accessing large repository of data in a database. This in turn can increase an organization's productivity.

IV.   APPLICATIONS OF DATA MINING

Data mining techniques have been successfully used for a wide range of real-world applications. Data mining applications are growing rapidly for the following reasons:

1. The amount of data in data warehouses and data mart s is growing exponentially. Organizations need the type of automated techniques provided by data mining tools to mine the knowledge in these data

2. New data mining tools with expanded capabilities are continually being introduced

3. Increasing competitive pressures and productivity are forcing organizations to make better use of the information and knowledge contained in their data.

The general wide range application of data mining in business functions of an organization includes the following;

4.1   Profiling Population: Developing profiles of high-value customers, credit risks, and credit card fraud. In-depth knowledge of customers and business partners helps an organization build a healthy business-to-business or business-to-customer relationship

4.2   Analysis of business trends: involves identifying the growth trend in the market. Understanding the trends of goods and services (either below or above average) will help an organization remain competitive.

4.3   Target Marketing: involves identifying customers (or customers segments) for promotional activity. Having a good relationship with a particular market segment will help boost the organization's productivity.

4.4   Customer retention and Churn Analysis: examining the behaviour of customers who have left for competitors to prevent remaining customers from leaving. Organizations, especially Telecommunication companies are interested in predicting the churning behaviour of their customers beforehand so that selective marketing strategies can be developed towards these groups to reduce churning rate [7].

4.5   Financial Forecasting: Data mining techniques can be used for forecasting prices, returns or volatilities of various financial securities, so that investment companies can develop more accurate portfolio allocation models and suggest creative diversification alternatives to their customers [8].

4.6   Up-Selling: Identifying new products or services to sell to a consumer based upon critical events and life-style changes.

4.7 **Product Affinity**: helps an organization identify products that are purchased concurrently, or the characteristics of consumers for certain product groups

4.8 **Fraud Detection**: Fraud detection models identify which transactions are likely to be fraudulent. Predicting the likely fraudulent behaviour may cause lots of savings for financial institutions especially banks for credit card fraud or telecommunication companies for telephone call fraud. For example online DM models running behind operational systems can quickly identify and monitor suspicious transactions [9].

More specifically, data mining can be applied in different organizations to achieve different objectives. These objectives in turn increase the productivity, and promote the growth of these organizations.

Data Mining in Health care

With the growth in electronic health records (EHRs), more and more facilities are gathering huge amounts of digitized patient data. Because EHRs contain so much data, they provide a rich source for healthcare organizations to tap for information about how to improve patient care and reduce costs. Healthcare providers can use data mining to uncover previously unknown patterns from vast data stores and then use this information to build predictive models. Data mining in health care have the following benefits:

1. Healthcare workers can use data mining applications to identify and track chronic diseases and high-risk patients, and design appropriate interventions. Hence, Patients can receive better, more affordable healthcare services.

2. Healthcare organizations can use data mining to make better patient-related decisions. It can also help these organizations influence revenues, costs, and operating efficiency while maintaining high-quality care.

Data Mining in Banking and Insurance

Applying Data mining in the banking sector can help banks detect patterns of fraudulent credit card use, identify and reward loyal customers, and predicting customers likely to change their credit card affiliation. Similarly, a credit card company can leverage its vast warehouse of customer transaction data to identify customers most likely to be interested in a new credit product. In insurance, data mining can help insurance companies predict which customers will buy new policies.

Data Mining in Retail Marketing

Data mining in retail marketing assists in identifying the buying patterns of customers, predicting response to mailing campaigns, finding associations among customer demographic characteristics, and in Market basket analysis.

Data Mining in Education

Data Mining is a relatively new technology in educational institutions. The more data these institutions collect about their students, the more the value these institutions will deliver to them. And the more value you can deliver to them…the more the revenue generate. Most of the data mining techniques mentioned above can be applied in education. For example, the concepts of fractals can be used to breakdown large students and staff data repository into smaller structure for easy data access and retrieval. Likewise, previously unknown and actionable data can be mined from an institutions database to support decision making.

V.   BENEFITS AND ISSUES OF DATA MINING FOR ORGANIZATION'S GROWTH

Deploying data mining in an organization can be of great advantage;

1. **Competitive Advantage**: Data mining can provide insights that will increase customer's loyalty, Unlock hidden profitability within the organization, and increase staff/client/customer retention [11].

2. **Insight into the future**: Comparing the information derived from past data mined in the organization with current information can give new insights to the organization that otherwise might be overlooked. For example, an organization that knows there is always a sharp decline in sales in a particular month can give extra attention and effort into maximizing the output from that month.

3. **Economy of scale**: Mining only valuable and useful data from the data warehouse can potentially reduce the cost expended on having to work a large collection of information.

4. **Increased Productivity**: Applying complex Data mining techniques and algorithm to solve problems in an organization can reduce the time spent on achieving certain objectives, thereby increasing productivity.

5. **Induced Growth and development**: An organization having a competitive advantage in the market will have great customer and supplier relationship, thereby causing the organization to grow and develop.

6. **Potential high returns on investment**: An organization must commit a huge amount of resources to ensure the successful implementation of data mining and the cost can vary enormously. However studies show that average five-year returns on investment (ROI) in data mining reached over 35%, with more than 80% organization achieving over 35% ROI.

5.1   Issues of Data Mining
While data mining is still in its infancy, it is becoming a trend. Before data mining develops into a conventional, widely accepted discipline, many still pending issues have to be addressed. The following are the problems faced with implementing data mining in an organization;

1. **Problem of poor data quality**: Applying Data mining techniques to a business problem requires availability of high quality data. This data may be primary data; those generated within the organization from various business functions, or secondary data; those data derived from other sources. Implementing data mining in an organization requires high quality data accumulated in the data warehouse of the organization. Even high quality data can be expensive; the return on investment of the data mining processes will cover the cost of the data. On the other hand low quality data is usually cheaper but most likely not cover return on investment for a Data Mining process.

2. **Employee empowerment**: Although Data mining can solve a whole lot of problems in an organization, finding a well-trained data mining expert for a Data mining project can usually pose a challenge. Adequate plans must be made to empower employees by providing trainings and compensation [10].

3. **Data integrity and security issues**: Integrity and security are important issues with any data collection that is shared and/or used in an organization. In addition, when data is collected for use, large amounts of sensitive and private information about individuals or companies is gathered and stored. This becomes controversial given the confidential nature of some of this data and the potential illegal access to the information. Moreover, data mining could disclose new implicit knowledge about individuals or

groups that could be against privacy policies, especially if there is potential dissemination of discovered information [5].

4. **Complexity of Integration**: The most important area for the management of a data mining project is the ability to effectively integrate it into the organization's business processes. This means that an organization must spend a significant amount of time determining how well the various different data mining techniques can be integrated into the overall solution that is needed. This can be a very difficult task, as there are a number of techniques that can be implemented for various business operations of the organization.

VI. CONCLUSION AND FURTHER STUDY

This study is aimed at reviewing the role that data mining plays on organization's growth or its level of productivity. Firstly, an introduction and the threefold goals of Data Mining are presented. Then we reviewed the kind of data that can be mined in an organization and the various techniques an organization can use in mining such data. The work was concluded with how and where data mining can be applied to boost productivity and growth, and the various advantages attached to implementing data mining projects in an organization.

As a future work we plan to conduct a field study to identify the critical success factors that affects Data mining implementation in an organization.

REFERENCES

[1] Olufade F.W.O. (2013) Fundamentals of Data Mining; Department of Computer science, University of Ibadan, Ibadan.

[2] Weldon J.L. (1997) State of the Art: Warehouse Conerstones; Bvte, January. pp. 87

[3] Abdullah S.A, Zahid U. and Farrukh S. (2009) Successful Implementation of CRM: The Role of Data Mining; 2009 International Conference on Computer Engineering and Applications IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore.

[4] Bertan B. and Omer L. (2006) Role of Data Mining in Strategic Management

[5] Osmar R.Z. (1999) Introduction to Data Mining; CMPUT690 Principles of Knowledge Discovery in Databases, University of Alberta

[6] Bharati M.R. (2012) Data Mining Techniques and Application; Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305

[7] Berry, M. J. A., Linoff, G. S. (2004) Data Mining Techniques for Marketing, Sales and Customer Relationship Management, 2ed, Wiley Publishing

[8] Kovakerchuk, B., Vityaev, E. (2000) Data Mining in Finance, Kluwer Academic Publishers

[9] Fawcett, T. E., Provost, F. (2002) Fraud Detection, in Klösgen, W. (Edited), Zytkow, J. M. (Edited), Handbook of Data Mining and Knowledge Discovery, Oxford University Press, pp. 726-731

[10] Oluigbo I. and Anyiam K. (2014) The Role of Compensation On Employee Performance in an Information Technology Organization; IJRIT International Journal of Research in Information Technology, Volume 2, Issue 4, April 2014, Pg: 548- 558

[11] 10 Ways Data Mining Can Help You Get a Competitive Edge, https://blog.kissmetrics.com/data-mining/, Accessed date: 10th September 2014