

The Impact of Emerging Technologies on Data Mining

Atkinson Ike Madukwe¹, Jennifer Somtochukwu Madukwe²,
Justina Ogochukwu Osonwa³ and Obiora Chukwuemerie Ernest⁴
¹Chief Librarian, ²Librarian, ^{1&2}Spiritan University, Nneochi, Abia State, Nigeria
^{3&4}Graduate Assistant, Nnamdi Azikiwe University, Awka, Nigeria
E-mail: madukweatkinson@gmail.com

Abstract - Business analytics has improved tremendously in recent past providing business leaders' insights, particularly from operational data stored in transactional system. An example is e-commerce data analysis, which has recently come to be viewed as a killer appropriate for the field of data mining.

Keywords: Impact, Emerging Technologies, Data Mining

I. INTRODUCTION

Data mining is the use of automated data analysis techniques to uncover previously undetected relationships among data items. It often involves the analysis of data stored in a data warehouse. Generally, three of the major data mining techniques are regression, classification and clustering. Data Mining also popularly known as Knowledge Discovery in Databases (KDD) refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. In general, Data mining is used for a variety of purposes in both the private and public sectors. Industries such as insurance, banking, medicine, and retailing commonly use data mining to increase sales, enhance research, and reduce costs (Anmol, 2014).

Before the big mining era began, businesses assigned relatively low value to the data they were collecting that did not have immediate value. When the data mining began, this investment in collecting and storing data for its potential future value changed, and organizations made a conscious effort to keep every potential bit of data. This shift in behavior created a virtuous circle where data was stored and then, because data was available, people were assigned to find value in it for the organization. The success in finding value led to more data being gathered and so on. Some of the data stored was a dead end, but many times the results were confirmed that the more data you have, the better off you are likely to be. The other major change in the beginning of the big data era was the rapid development, creation, and maturity of technologies to store, manipulate, and analyze this data in new and efficient ways. After the success of the data processing results in every step is determined from the choice of data and ending with an explanation of the anomalies in the results. To make decisions based on data collected by an organization, they must either rely on data analysis to extract information from the data or employ analytic applications that blend data

analysis technologies with task-specific knowledge (Sumathi & Sivanandam: 2006). Enormous amount of data is generated every minute. A recent study estimated that every minute, Google receives over 4 million queries, e-mail users send over 200 million messages, YouTube users upload 72 hours of video, Facebook users share over 2 million pieces of content, and Twitter users generate 277,000 tweets (domo). With the amount of data growing exponentially, improved analysis is required to extract information that best matches user interests.

II. OBJECTIVES OF THE STUDY

The main objective of this paper is to determine the Impact of Emerging Technologies on Data Mining. The specific objectives are to

1. Find out the challenges of emerging technologies on data mining.

III. LITERATURE REVIEW

A. Emerging Technique for Data Mining

Data mining has great potential to produce useful information for companies which can benefit the way they manage their problems. Big data analysis is becoming indispensable for automatic discovering of intelligence that is involved in the frequently occurring patterns and hidden rules. These massive data sets are too large and complex for humans to effectively extract useful information without the aid of computational tools. Emerging technologies such as the Hadoop framework and MapReduce offer new and exciting ways to process and transform big data, defined as complex, unstructured, or large amounts of data, into meaningful knowledge (Natarajan, 2014).

B. Hadoop

Hadoop consists of distributed file system, data storage and analytics platforms and a layer that handles parallel computation, rate of flow (workflow) and configuration administration (Priya & Chandrakant 2014). Hadoop is a scalable, open source, fault tolerant Virtual Grid operating system architecture for data storage and processing. It runs on commodity hardware; it uses HDFS which is fault-tolerant high bandwidth clustered storage architecture. It

runs MapReduce for distributed data processing and it works with structured and unstructured data (Puneet & Sanchita 2013).

C. Map Reduce

Map Reduce is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster. Hadoop Map Reduce is a programming model and software framework for writing applications that rapidly process vast amounts of data in parallel on large clusters of compute nodes (Puneet & Sanchita 2013). MapReduce consists of two functions, map and reduce. Mapper performs the tasks of filtering and sorting and reducer performs the tasks of summarizing the result. There may be multiple reducers to parallelize the aggregations (Richa, Sunny & Anuradha 2014).

D. Emerging Trends in Business Analytics

The key consumer is the business user, whose job possibly in merchandising, marketing, or sales, is not directly related to analytical tools to improve the results of some business process along one or more dimensions (such as profit and time). Fortunately, data mining analytic applications, and business intelligence systems are now better integrated with transactional systems than they were once, creating a closed loop between operations and analysis that allows data to be analyzed and the results reflected quickly in business actions. The mined information today is deployed to a broader business audience taking advantage of business analytics in its everyday activities. Analytics are now routinely used in sales, marketing, supply chain optimization, and fraud detection. Today's business environment is more dynamic, so businesses must be able to react quicker, must be more profitable, and offer high quality services that ever before. Here, data mining serves as a fundamental technology in enabling customer's transactions more accurately, faster and meaningfully. Data mining techniques of classification, regression, and cluster analysis are used for in current business trends. Almost all of the current business data mining applications are based on the classification and prediction techniques for supporting business decisions, thus creating strong Business Intelligence (BI) system.

E. Automated Prediction of Trends and Behaviors

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.

F. Web and Semantic Web

Web is the hottest trend now, but it is unstructured. Data mining is helping web to be organized, which is called Semantic web. The underlying technology is Resource Description Framework (RDF) which is used to describe resources. FOAF is also a supporting technology, heavily used in Face book and Orkut for tagging. But still there are issues like combining all RDF statements and dealing with erroneous RDF statements. Data mining technologies are serving a lot to make the web, a semantic web.

IV. METHODOLOGY

The study adopted a descriptive research method which is concerned with the classification of materials and the representation of events or situations in clear and understandable manner. To achieve this, descriptive research engages on the use of figures and the outlining of events in sequential order Onyishi & Omemma (2010). This paper describes Impact of Emerging Technologies on Data Mining. The approaches used are the authors' personal experiences and observation working in a university library, observations of responses with respect to quarries, as well as the reflections among the students and patrons were done using Structured questionnaire which were mailed out through E-mails, WhatsApp, Facebook, Telegram and Twitter. Pretest self-administered questionnaire and observation checklist were used to collect data on different variables. Data entry and analysis were done using Excel 2016 and SPSS 25.

TABLE I FIND OUT THE CHALLENGES OF EMERGING TECHNOLOGIES ON DATA MINING

Sl. No.	Questionnaire Items	SA (4)	A (3)	D (2)	SD (1)	N	Mean	Remark
1	Ubiquitous Data Mining (UDM)	116	51	16	4	58	3.2	Accept
2	Hypertext and hypermedia Data Mining	152	48	6	1	58	3.6	Accept
3	Multimedia Data Mining	172	24	8	3	58	3.6	Accept
4	Prone to fraud	180	30	4	1	58	3.7	Accept
5	It does not lead to Security and Privacy concerns	40	36	34-	19	58	2.2	Rejected
6	I feel in the long run it will lead to blackmail	208	12	2	2	58	3.8	Accept

Source: Field Survey, 2021

V. RESULTS

Table I shows the distribution of responses for questionnaire items measuring challenges of emerging technologies on data mining. The statistics of those who strongly agreed, agreed, undecided, disagreed and strongly disagreed were succinctly captured in the Table. However, the analysis was based on the mean of the individual questionnaire items. The decision rule is to accept any questionnaire item with a mean of 3 and above as been entrenched or happening in the organization or being true while any questionnaire item that is below 3 is rejected as not being applicable or true with the respondents cumulatively.

Starting with questionnaire items used in measuring Ubiquitous Data Mining poses a major challenge to the users this according to them introduces additional cost due to communication, computation, security, and other factors this is in agreement with (Annan 2012) who stated that Ubiquitous Data Mining cost due to communication, computation, security, and other factors are a major challenge to UDM. Similarly, Human-computer interaction is another challenging aspect of UDM. Visualizing patterns like classifiers, clusters, associations and others, in portable devices are usually difficult. The respondents agreed that Hypertext and hypermedia Data Mining are a challenge to them some complained that are at risk of losing their mind over this. Which the attributed to receiving unsolicited emails, text and notifications from media companies. The respondents where almost unanimous in voicing their privacy and security concerns, this according to them some apps in their phones are spying on them sending and receiving unsolicited messages to data companies.

VI. CONCLUSION

The amounts of data are growing exponentially worldwide due to the explosion of social networking sites, search and retrieval engines, and media sharing sites, stock trading sites, news sources and so on. Data is becoming the new area for scientific data research and for business applications. Data analysis is becoming indispensable for automatic discovering of intelligence that is involved in the

frequently occurring patterns and hidden rules. Data analysis helps companies to take better decisions, to predict and identify changes and to identify new opportunities. In this paper we briefly reviewed the various data mining trends and applications from its inception to the future. This review puts focus on the hot and promising areas of data mining. Though very few areas are named here in this paper, yet they are those which are commonly forgotten. This paper provides a new perspective of a researcher regarding the impact of emerging technologies on data mining.

REFERENCES

- [1] K. Anmol, K. T. Amit and K. T. Surendra, "Data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work," *International Journal of Emerging Technology and Advanced Engineering* Website: www.ijetae.com (ISSN 2250-2459 (Online), Vol. 4, No. 1, February 2014.
- [2] N. P. Annan, "Data mining: future trends and application," *International Journal of Modern Engineering Research (IJMER)* www.ijmer.com Vol. 2, No. 6, pp. 4657-4663, Nov-Dec. 2012, ISSN: 2249-6645,2012.
- [3] Clifton, Brian, "Advanced Web Metrics with Google Analytics," 2nd edition, *Sybex*. 2010.
- [4] [Online]. Available: <http://www.domo.com/blog/2014/04/data-never-sleeps-2-0>.
- [5] Kaushik, Avinash, "Web Analytics 2.0 - The Art of Online Accountability and Science of Customer Centricity," *Sybex, Wiley*. [Online]. Available: <http://nlp.stanford.edu/IRbook/html/html5edition/evaluation-of-ranked-retrieval-results-.html>. 2009.
- [6] S. Sumathi and S. N. Sivanandam, "Emerging Trends and Applications of Data Mining," *Studies in Computational Intelligence (SCI)*, Vol. 29, pp. 165-183, Springer-Verlag Berlin Heidelberg, 2006.
- [7] M. Natarajan, *et al.*, "Issues, Challenges, and Solutions: Big Data Mining Netcom, Csit, Graph-Hoc," *Sptm* – 2014 pp. 131-140, 2014. Cs & It-Csep 2014 Doi: 10.5121/Csit.2014.41311. 2014.
- [8] P. S. Priya and P. N. Chandrakant, "Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution", *IJCSIT*, Vol. 5, No. 2, pp. 2126-2131, 2014.
- [9] S. D. Puneet and P. Sanchita, "Big Data Analysis: Challenges and Solutions," *Int. Conf. on Cloud, Big Data and Trust, RGPV*, 2013.
- [10] G. Richa, G. Sunny and S. Anuradha, "Big Data: Overview", *IJCTT*, Vol. 9, No. 5, 2014.
- [11] T. O. Onyishi and D. A. Omemma, "Understanding Social Research," *John Jacob's Classic Publishers Ltd*, Enugu, 2010.