Techniques and Tools of Big Data Analytics at the Technical University of Kenya and Strathmore University

*Lucy Wachera Kibe, Tom Kwanya, Ashah Owano
The Technical University of Kenya
*Email: * kibelucy871@gmail.com*

Abstract

Big data is the term used to refer to any group of datasets so huge and composite that it is difficult to process the same using traditional data processing applications. Big data analytics is a set of procedures and technologies that entail new forms of integration to uncover large unknown values from large datasets that are various, complex, and of an immense scale. Analysing big data is a challenging task as it contains huge dispersed file systems which should be fault-tolerant, flexible and scalable. There is an immense need of constructions, platforms, tools, techniques and algorithms to handle big data. Some of the tools used to anlyse big data are Hadoop, Map Reduce, Apache Hive, and No SQL, among others. Techniques for big data analytics include descriptive, diagnostic and prescriptive analytics. This chapter compares the techniques and tools used for big data analytics by the Technical University of Kenya with those used by Strathmore University. The study on which this chapter is based was conducted as a mixed method research to enable deep understanding of the concept. Primary data was collected through structured questionnaires and interviews with clientele and information communication technology staff from the two institutions in Nairobi, Kenya. Secondary data was collected through document analysis. Data was analysed and presented using descriptive statistics. The findings revealed that the tools used frequently for big data analytics were SQL and Java. The two academic institutions mostly used descriptive big data analytics techniques. There was variance in the use of some techniques where SU applied predictive and TUK diagnostic techniques, SU used rules and algorithms to detect the patterns. They also employed statistical analysis, data mining and machine learning to get meaning from data. On the other hand, TUK employed diagnostic analytics to examine their big data.

Keywords: *Big Data, Big Data Analytics, Big Data Analytics Tools, Big Data Analytics Techniques*

1 Introduction

The value of big data is realised when an organisation is able to leverage on it to make decisions. Organisations need to turn big data into meaningful insights in order to benefit from it. Labrinidis and Jagadish (2012) assert that in order to extract value from big data, organisations need to conduct data management and analytics. Through data management, organisations are able to attain, store and retrieve data for analytics. After acquiring the data, organisations can analyse it to gain insight from it. Big data analytics is the procedure where the large amount of data is analysed to gain insight from it and show correlations (Sagiroglu & Sinanc, 2013). The use of big data analytics enables organisations to gain insight from the large volumes and varieties of data. Big data analytics depends on the organisation's ability to employ technologies that can assist them to analyse the big data they hold. The technologies include data management tools such as Hadoop, MapReduce, arithmetical analysis and advanced visualisation tools. The advancement of technology especially in the big data analytics has made it possible to retrieve data of every kind not withstanding its size, variety and retrieval speed required. Big data analytics has created organisational value through increased transparency; created adaptive organisation models; and supported decision making (Wamba, Akter, Edwards, Chopin & Gnanzou, 2015). This helps organisations to enhance their competitive edge through the creation of information based collections pools that assist in decision making.

2 Techniques of big data analytics

According to Laney (2012); Banerjee, Bandyopadhya and Acharya (2013); Evans and Lindners (2012); Eckerson (2007); Chapelle, Schölkopf and Zien (2006) and Basu (2013), techniques of big data analytics include descriptive, diagnostic, predictive and prescriptive big data analytics.

Descriptive analytics (what happened?)

Organisations use descriptive analysis to recognise previous and present organisational performance in order to make informed decisions. Through this, organisations are able to classify, describe, combine and categorise data to translate it to valuable information for understanding organisational performance. The analysis assists in summarising the data into meaningful charts and reports.

Diagnostic analytics (why did it happen?)

Banerjee *et al.* (2013) perceive diagnostic analytics as an exploratory analysis of existing data to discover the root cause of a problem in an organisation. The analytics is also termed as generic extrapolative analytics which examines data to answer the question "why did it happen?" It uses techniques such as drill-down, data discovery, data mining and correlations. It looks at the causes of events and behaviours based on data.

Predictive analytics (What might happen?)

Predictive analytics examines historical data and combines it with rules and algorithms. It can detect patterns or relationship in data. According to Evans and Lindners (2012), predictive analytics extrapolates data relationships. This analysis involves statistical analysis, data mining and machine learning to find meaning from large amounts of data. There are two major types of predictive analytics. These are supervised and unsupervised analytics (Eckerson, 2007).

(i) Supervised predictive analytics

Supervised prediction is based on the previously solved cases or historical data that contains the results one is trying to predict. The approaches applied include classification, regression and time series analysis. For example, classification can be used if one wants to know which clients are likely to respond to a new direct mail campaign while regression is used in forecasting. Variance analysis and time series assist in understanding the unique properties of time and calendars (Chapelle, Schölkopf & Zien, 2006).

(ii) Unsupervised predictive analytics

According to Eckerson (2007), the unsupervised predictive analytics directly infer the properties of the probability using descriptive statistics to examine the natural pattern and relationships that occur within the data. It can be used to identify products and content that goes well together hence used a lot in market analysis. According to Chapelle *et al.* (2006),) the concept of semi-supervised predictive analytics algorithms is applied with some supervision information but not necessarily for all examples.

Prescriptive analytics (How can we make it happen?)

Basu (2013) explains that prescriptive analytics enable organisations to not only look at the future but also to look at the opportunities that are potentially out there. It goes beyond describing, explaining and predicting by associating alternatives with the predictions of outcomes. It helps analysts to know what might happen in future and optimise to get the best in achieving the organisational objectives with the limited resources they have. Riabacke, Danielson and Ekenberg (2012) assert that prescriptive analytics assists to forecast what lies ahead and suggests means to take advantage of this predicted future without compromising other significances. Examples of prescriptive analytics are recommendation systems such as those used by Netflix, Google, and Amazon. Figure 1 presents examples of the types of analytics.

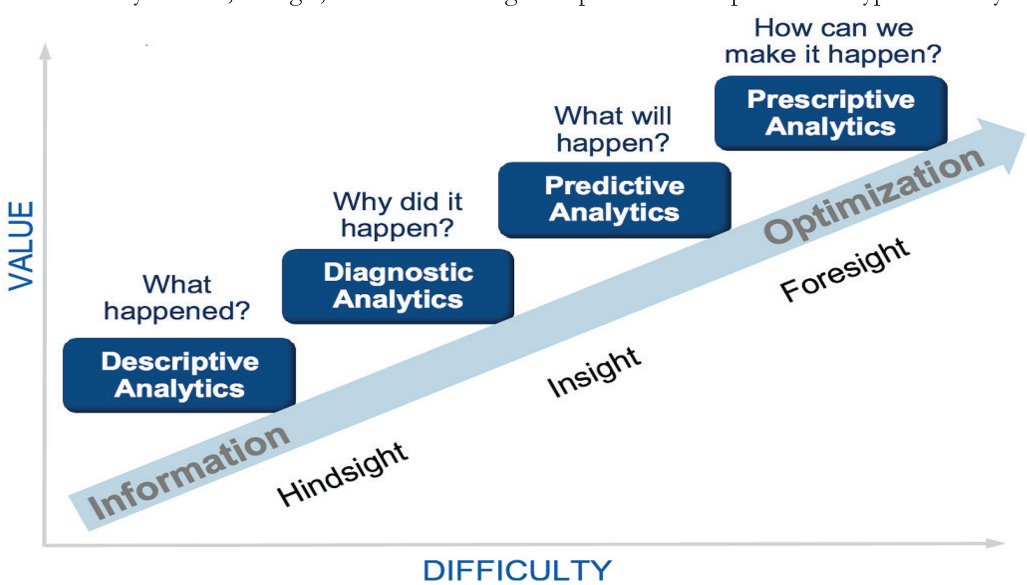

Figure 1: Big data analytics techniques

Source: Laney (2012)

3 Tools for big data analytics

To conduct big data analytics, there are tools required to help in analytics and getting insights. According to Ajit-Kumar (2016), the tools used for big data analytics depend on five concepts which are:

1. Processing - This is where the processing is hosted, for example, on distributed servers or cloud such as Amazon EC2;
2. Storage - This is where big data is stored;
3. Programming model - Distributed processing by use of MapReduce;
4. Indexing of big data - Entails the use of high schema databases such as MongoDB; and
5. Processes performed on data - Examples include analytical or semantic processing.

Researchers such as Loshin (2013), Raghupathi and Raghupathi (2014), Shvachko, Kuang, Radia and Chansler (2010), Zikopoulos and Eaton (2011), Fan and Bifet (2013) as well as Cuzzocrea, Song and Davis, (2011) identify some of the tools and platforms for big data analytics as indicated in Table 1.

Table 1: Big data analytics tools

Tools/platform	Role in big data analytics
Hadoop	Storage of big data using Hadoop clusters.
MapReduce	Dividing tasks to be done during data analytics and assembling of outputs. During the execution of tasks, it tracks the processing and shows progress.
Hive	Works on the Hadoop platform. It allows querying on SQL to retrieve big data. Can be used for data mining.
Cassandra	Used to handle big data that is on different servers. Also supports querying of No-SQL type.
MongoDB	Can be an alternative to relational databases. It is used for managing big data that keeps on changing and which is unstructured or semi-structured. It helps to store and analyse big data.
NoSQL	Used for storing unstructured data. Also can be used to store large set of data.
RapidMiner	Offers advanced analytics where the users do not require to write any code. It works as a data mining tool for data dispensation and imagination.
Talend	Big data integration to cloud controlling and also simplifies the processing of big data.
Bokeh	Used to generate data applications, dashboards and plots.
Plotly	A tool for visualisation of big data. It can be used even by organisations with low skills on big data. It also enables the sharing of the visualised data.
Cloudera	Used for creating data repositories.
R	Software used for analytics and graphics. It also assists with data mining and statistical analysis and presentation.

Source: Research Data

4 Rationale of Study

Academic institutions have not been left behind in the production of big data in terms of student and staff records; research output and innovations; as well as administrative, logistics, financial and procurement records. These records are produced very fast, in vast volumes and diverse formats. This creates enormous challenges in identifying, processing and applying data produced by the academic institutions for decision making and general operations of the organisations. Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Byers (2011) assert that the concepts of big data and analytics have developed into hotspots that fascinate academic institutions, industries and most government in the world. These concepts have infiltrated most of today's organisations making it a crucial influence of performance. Similarly, the same academic institutions are operating in increasingly complex and competitive environments which they need to forecast and respond to appropriately. Clarke, Nelson and Stoodley (2013) are of the opinion that academic institutions should realise the need to make decisions based on the synthesis of the vast data that they generate to help them understand the rapidly changing contexts of the academic sector; otherwise, the potential value of big data is not realised. They recommended that academic institutions need to invest on tools and techniques that would unravel insight form the big data they generate. This will in turn enable them to generate meaning from the big data. There is a dearth of literature on how academic institutions in Kenya analyse the big data they generate. This study seeks to bridge the gap by analysing tools and techniques used for big data analytics by academic institutions in Kenya. The specific objectives of the study were to: investigate the techniques used in big data analytics at the Technical University of Kenya and Strathmore University; investigate the tools used in big data analytics at the Technical University of Kenya and Strathmore University; and recommend tools and techniques that academic institutions in Kenya can use to analyse big data.

5 Context of Study

The mandate of the Technical University of Kenya is to provide technical education and training so as to contribute towards the progression of society through research and innovation. Therefore, the university has engaged in innovation projects such as the development of the "green tuk tuk" that uses biodiesel instead of fuel, among other innovations (TUK, 2014). The university has also not been left behind by the information explosion emanating from the vast amounts of information being produced from its processes, staff and students. To make sense out of this content, the university conducts big data analytics to support its decision making process. For example, every end of semester, the university conducts a financial analysis to determine the number of students who have paid school fees and those who have not as a means of establishing those who are eligible to sit examinations. Based on this analytics, the university management sends out memos to academic departments with lists of the students who should be allowed to sit examinations. Similarly, the university from time to time requests academic staff to submit information on various issues including publications and research projects, among others. The TUK represents universities which generate and manage big data without using conventional big data analytics tools and procedures. The university, therefore, offers the researcher a big data context which is in its infancy and ad hoc. Apart from the environment being representative of several other public universities, it also offers the researcher a good context to examine the potential of big data analytics in enhancing organisational performance.

According to the July 2017 and 2018 webometrics ranking of universities, Strathmore University emerged as the best private university in Kenya. It is perceived as the best performing private university in Kenya. On 21st July 2015[4] Strathmore University, under @iLabAfrica Research Centre, signed a memorandum of understanding (MOU) with the Dell EMC, which is a United States firm, to offer access to platforms for big data analytics, information on data storage management and the modern research and training resources on cloud infrastructure and data science. Through this, the university was able to get assistance from the firm in line with its vision of providing the latest of global technology tools and un-equalled training in Africa. This followed another engagement in which the university met with the Data Science Centre (DSC) in June 2015[5]. The DSC thereafter collaborated with the university to introduce the use of data analytics at the university to enhance research in the university and enhance its performance. These projects emanated from the understanding that big data research inspires new ways of transforming processes, organisations, industries and society at large. This researcher holds the view that the university provides an appropriate big data analytics environment which has a great potential to contribute to the objectives of this study.

6 Methodology of Study

The study from which this chapter is extracted was designed as a mixed method research design. According to Creswell and Plano-Clark (2015) mixed method design enables the understanding of a problem of study by gaining different corresponding data and enhancing validation. This research adopted a convergent parallel design. The data collected and analysed was both qualitative and quantitative. The authors merged the results from both sets of data for comparison and validation. This enhanced the interpretation of similar and dissimilar concepts. Primary data was collected through structured questionnaires and interviews from the Technical University of Kenya (TUK) and Strathmore University (SU). The study targeted 15,020 students and information communication technology staff from TUK and 7,030 from SU as study population. Therefore, the total population of the study was 22,050 respondents. Information-oriented purposive sampling was used to select information rich subjects where class representatives were selected for the study. This gave TUK a sample size of 580; while SU was 114. Thus, total sample was 694 respondents. Questionnaires were distributed to all the selected respondents using drop and pick techniques and face to face interviews with key informants. Data was analysed using SPSS and presented using descriptive statistics. Secondary data on the understanding of the concept was collected through documentary analysis.

6 Findings and discussions

A total of 576 (83%) respondents responded to the study while 118 (17%) of them did not provide usable responses. The response rate was considered admissible since Mugenda and Mugenda (2012) recommend that a response rate of at least 50% is adequate for analysis; 60% is generally good; while a response rate of above 70% is excellent. Kothari (2014) concurs with this position and adds that a response rate of above 70% is deemed to be very good. The findings are presented hereunder according to the objectives of the study.

6.1 Field of study and titles for staff respondents

The research investigated the areas of study and job titles of the staff respondents at TUK and SU. This was done in order to establish whether the people dealing with big data analytics are trained in the field. None of the respondents indicated that they were big data analysts but from the field of study they indicated that they studied analytics and information technology as shown in Table 2.

4 http://www.nation.co.ke/business/corporates/Strathmore-deal-with-US-firm-to-boost-research-and-training/-/1954162/2801342/-/bc3hm0/-/index.html

5 https://www.strathmore.edu/news/ims-establish-research-data-science-centre-industrial-mathematical-program/

Table 2: Fields of study and titles of the staff's respondents

Fields of study	Mobile telecommunication and IT
	Information systems and security
	Mobile innovation and telecommunication
	Telecommunication and innovation
	Analytics and business intelligence
	computer science
	hardware maintenances
	Information technology
	Information and communication technology
	Telecommunications technology
	Business intelligence
	Machine learning and data mining
Job titles	Software developer
	Project coordinator
	Business intelligence analyst
	Senior technologist
	System developer
	Technologist
	Technician
	Business intelligence specialist
	Analytics designer
	Programmer

Source: Research Data

It is evident from Table 2 that none of the staff members indicated they were big data analysts. However, their fields of study and job titles closely resemble the descriptions of big data analysts. Patil (2011) asserts that big data analysts are often trained in mathematics, computer science, business intelligence, statistics, database management, and social science. They can also emerge from any field that focuses on computational and data management. Although none of the respondents indicated that they were big data analysts, their field of study and job titles agree with the description of Patil (2011) and Harris, Murphy and Vaisman (2013) on big data analysts.

6.2 Technical skills for big data analytics

The study investigated the technical skills of the respondents relevant to big data analytics. This resulted in multiple responses since the respondents were free to choose more than one entry to the question that sought to establish technical skills, competencies and expertise of staff as shown in Figure 2.

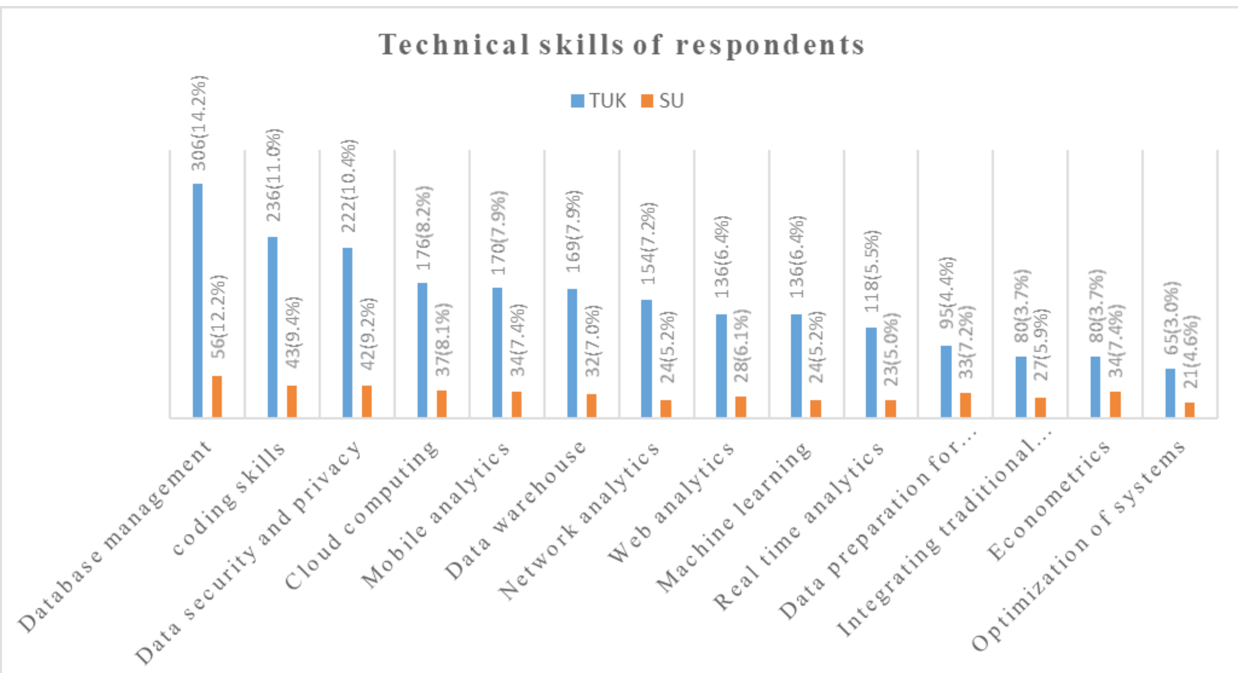

Figure 2: Technical skills of respondents

The most popular technical skill was database management with 306(14.2%) respondents from TUK and 56(12.2%) from SU. The other notable competences included coding skills indicated by 236(11.0%) respondents from TUK and 43(9.4%) from SU; as well as data security and privacy skills which was identified by 222(10.4%) respondents from TUK and 42(9.2%) from SU. The prominence of coding skills is understandable because it is one of the basic competency requirements for big data analysts. Data analysts need to master technical skills such as mathematics and statistics, data architecture design, and databases and data warehouse knowledge (Miller, 2014). Competency in data management seems to be the requirement for big data analysts. Among the study participants, competency in subjects related to data management such as data architecture design, databases, integration of different kind of data and preparing data for analytics was generally at a good level.

6.3 Tools for big data analytics

The staff were requested to rank their proficiency in the use of tools for big data analytics using a scale of 1-4 (where 1 was None; 2- Basic; 3- Proficient; and 4- Expert). Figure 3 presents the means of the responses from the two universities.

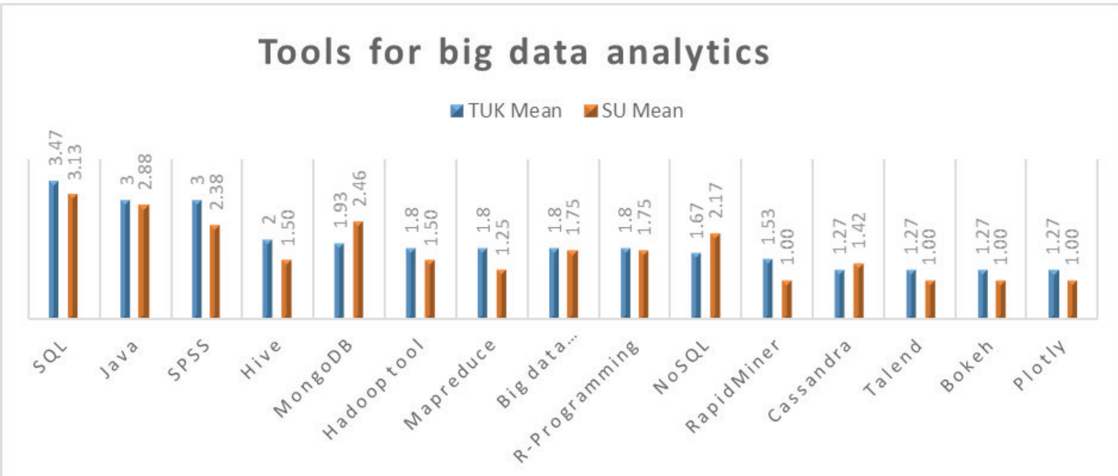

Figure 3: Staff means of tools for big data analytics

SQL tools had the highest mean with TUK having a mean of 3.47, while SU had 3.13 followed by Java as shown in Figure 3. This is an indication that most of the staff from the two univerities are proficenct in using SQL and Java tools for big data analytics. Based on the staff proficiency, the tools that had the least mean at TUK were Cassandra, Talend, Bokeh and Plotly, which had a mean of 1.27. At SU, the tools were RapidMiner, Cassandra, Talend, Bokeh and Plotly with a mean of 1. This meant that the staff had no profiecincy in using the tools.

When the respondents from the two universities were asked to identify tools they have used for the big data analytics, the question yielded multiple responses because a respondent was required to tick as many tools as they are proficient in as shown in Table 3.

Table 3: Student's responses on tools for big data analytics

Technical University of Kenya			Strathmore university		
Tools for big data analytics	**Responses**	**Percentages**	**Tools for big data analytics**	**Responses**	**Percentage's**
Java	310	20.70%	Java	56	19.50%
SQL	241	16.10%	R-Programming	50	17.40%
SPSS	180	12.00%	SQL	42	14.60%
R-Programming	162	10.80%	SPSS	36	12.50%
NoSQL	118	7.90%	Hadoop tool	31	10.80%
Hadoop tool	97	6.50%	Mapreduce	21	7.30%
Mapreduce	75	5.00%	NoSQL	13	4.50%
Hive	71	4.70%	MongoDB	12	4.20%
Rapid-Miner	63	4.20%	Hive	8	2.80%
Cassandra	61	4.10%	Rapid-Miner	5	1.70%
MongoDB	50	3.30%	Cassandra	4	1.40%
Cloudera	42	2.80%	Cloudera	4	1.40%
Bokeh	14	0.90%	Plotly	3	1.00%
Plotly	14	0.90%	Bokeh	2	0.70%
Total	**1498**	**100.0%**	**Total**	**287**	**100.0%**

Source: Research Data

Table 3 shows that Java was the most popular tool in the two institutions with 310 (20.7%) at TUK and 56 (19.5%) at SU. The least popular tools were Bokeh and Plotly 14 (0.90%) for TUK, and Bokeh 2 (0.7%) at SU. Concerning tools, the basic requirements for data analysts seem to be competence in Java, SQL, and SPSS. Competence in SQL also seems to be needed in many cases.

6.4 Techniques for big data analytics

The research sought to establish whether the concept of big data analytics was employed in the two academic institutions. The respondents were asked to explain why and how they undertake big data analytics. Two of the respondents expressed their views as follows:

> *"Getting relevant information from big data requires approaches that help gain the insight from the data. This requires looking at all the data that you have and understanding what it contains to help you to decide what to do with any other data similar to it in future." [TUK]*

> *"To make meaning from big data, there needs to be analytics done to it. This will help organisations to get value from the big data through insights. It can be done by use of statistical analytics, for example, analysing how many students are admitted and from them how many manage to graduate." [TUK]*

A number 16 (67%) of the respondents from SU reported that they used the theme of descriptive analytics to try and make meaning from data they had to support current decision making. The rest 8(33%) said they used predictive analytics which is the use statistics and algorithms to show patterns of events and relationships. Similarly, 12(80%) respondents from TUK indicated that most of them use descriptive analytics to make meaning from the big data while 3(20%) use diagnostic analytics. Below are some of the verbatim responses:

> *"Analytics today is the bedrock of any growing business. This is because analytics can help us in bringing improvements in performance, enhance security, productivity, and efficiency. There are some very good developments happening in machine learning and artificial intelligence which could potentially have huge impact on organization." [SU]*

> *"Today, every discussion about changes in technology, business, and society begins with data. In its exponentially increasing volume, velocity and variety, data is becoming a new natural resource and if not well analysed it will not make sense to the society. Organisations need to explore their existing data and make sense from them to keep operating in business." [SU]*

7 Conclusion

Both institutions used descriptive big data analytics to leverage on their big data. TUK also analysed its big data through diagnostic data analytics technique while SU also preferred predictive big data analytics technique. Arising from the findings from the two institutions the most used tools for big data analytics were Java and SQL. The institutions also used tools such as SPSS, R-Programming and NoSQL to get meaning out of their big data. The mean of proficiency in the various tools for big data analytics showed that the staff of the two universities were proficient in SQL followed by Java tools. However, they had inadequate skills on the use of tools like Cassandra, Talend, Rapidminer and MongoDB.

It also emerged that the two institutions use descriptive analytics. This was the most popular technique in both institutions. There was variance in the use of some techniques where SU applied predictive and TUK diagnostic techniques, SU used rules and algorithms to detect the patterns. They also employed statistical analysis, data mining and machine learning to get meaning from data. On the other hand, TUK employed diagnostic analytics to examine their big data. This is an approach where they just explore the different types of big data they generate and examine causes of events and behaviour based on the big data. The application of the two different big data analytics techniques could be so because SU has invested on big data analytics and has already established a lab that undertakes the project. The limitation of variety of techniques applied may also be indirectly related to low level of information and communication technology skills available in the institutions.

8 Recommendations

Based on the findings and conclusions of this study, the authors make the following recommendations:

8.1 Embrace more tools of big data analytics

Both TUK and SU need to embrace more big data analytics tools. They should particularly explore open source tools such as Hadoop and cloud-based analytics systems which bring significant cost advantages on the analysis and storage of large amounts of data.

8.2 Adopt diverse techniques of big data analytic

The two academic institutions should adopt diverse big data analytics techniques. For example, they can adopt techniques like learning analytics and mobile analytics techniques to enable them to get insight from their big data. Similarly, they can adopt advanced analytics techniques such as text analytics, machine learning, data mining, statistics, and natural language processing. The techniques can be used to analyse previously untapped data sources independently or together with their existing enterprise data to gain new insights resulting in better and faster decisions.

8.3 Invest on big data analytics and training on big data analytics

Both academic institutions should invest in enhancing their ICT tools, skills and capabilities. Investing on big data analytics significantly increases value addition on services or operating profits. Also investing on training personnel on big data analytics builds skills for the analytics. Skilled employees across the spectrum of data analytics roles are in short supply. So, aggressive actions to address this problem are critical.

9 References

Ajit-Kumar, R. (2016). Impact of Big Data Analytics on Healthcare and Society.'. *J Biom Biostat*, *7*(300), 2.

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential. *Vikalpa*, 38(4), 1-11.

Basu, A. (2013). Five pillars of prescriptive analytics success. *Analytics Magazine*, 8-12.

Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-Supervised Learning (Adaptive Computation and Machine Learning)*. MIT Press Cambridge, MA.

Clarke, J. A., Nelson, K. J., &Stoodley, I. D. (2013). *The place of higher education institutions in assessing student engagement, success and retention: A maturity model to guide practice*. Paper accepted for the 15th International Conference on the First Year in Higher Education. Wellington, New Zealand.

Creswell, J. W., & Plano -Clark, V. L. (2015. *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage

Cuzzocrea, A., Song, I. Y. & Davis, K. C. (2011). Analytics over large-scale multidimensional data: the big data revolution! *Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP*, New York, 101-104.

Eckerson, W. W. (2007). Predictive analytics. *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, *1*, 1-36.

Evans, J. R., & Lindner, C. H. (2012). Business analytics: the next frontier for decision sciences. *Decision Line*, *43*(2), 4-6.

Fan, W. & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM IGKDD Explorations Newsletter*, *14*(2), 1-5.

Harris, H., Murphy, S., & Vaisman, M. (2013). *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*. O'Reilly Media.

Kothari, C. R. (2014). *Research methodology: Methods and techniques*. New Age International, New Delhi.

Labrinidis, A. & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the twelfth VLDB Endowment*, 2032-2033.

Laney, D. (2012). 3D data management. Available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed 02 October 2017.

Loshin, D. (2013). *Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Elsevier, Amsterdam.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. London, McKinsey Global Institute

Miller, S. (2014). Collaborative Approaches Needed to Close the Big Data Skills Gap. *Journal of Organization Design,* 3(1), 26-30.

Mugenda, O. M. & Mugenda, A. G. (2012*). Research methods: quantitative and qualitative approaches*. Acts Press, Nairobi.

Patil, D. J. (2011). *Building data science teams.* O'Reilly Media, Inc.

Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*(1), 3.

Riabacke, M. Danielson, M. and Ekenber, L. (2012). State-of-the-Art Prescriptive Criteria Weight Elicitation. *Advances in Decision Sciences.*

Sagiroglu, S. & Sinanc, D. (2013). Big data: A review. *Proceedings of Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.

Shvachko, K., Kuang, H., Radia, S. & Chansler, R. (2010). The hadoop distributed file system. *Proceeding* of *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th symposium on* (pp. 1-10). IEEE.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. (2015). How 'big data 'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234-246.

Zikopoulos, P. & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, Pennsylvania.

About the Authors

Lucy Kibe is a Graduate Assistant in the Department of Information and Knowledge Management, the Technical university of Kenya. Prior to joining the academics, she worked at the America Reference Centre where she was in charge of processing information materials. She has authored journal articles, a book chapter and conference papers. Her current research interests include big data, cloud computing, bibliometrics, records management, indigenous knowledge, and knowledge sharing. She holds MSc in Information and Knowledge Management from the Technical University of Kenya and BSc in Information Sciences (first class honours) from the same university.

Tom Kwanya is an associate professor in the Department of Information and Knowledge Management at the Technical University of Kenya. He is currently also serving as the Director of the School of Information and Communication Studies. Prior to joining academics fulltime in 2013, he worked as a consultant on public information and knowledge management. He has authored several refereed journal articles, book chapters and conference papers. He has also edited two books and authored two monographs. His current research interests include organisational knowledge management, indigenous knowledge management, technology in information and knowledge centres, big data, and Internet of Things. Prof Tom Kwanya is also a research fellow in the Department of Information Science, University of South Africa.

Ashah Owano holds a PhD and a MA in Library and Information Studies from Moi University and the University of London (University College London) respectively. She is a librarian by profession and is currently a sessional lecturer at the Technical University of Kenya. Dr. Owano previously worked as the Resource Centre Manager at the National Museums of Kenya and an Information Designer at Egerton University. She has vast experience in information industry; was member of the editorial committee for the *Journal of East Africa Natural History* and Steering Committee member of the Biodiversity Heritage Library (Africa). Her research interests are in knowledge management and the use ICTs to access information and knowledge. Dr. Owano has worked with researchers in the documentation of Kenya's indigenous knowledge, including the Suba community and has collaborated with researchers in various multidisciplinary research projects.