

Big Data Analytics in Higher Education: A Review

Julius Murumba¹, Elyjoy Micheni²

^{1,2} Department of Management Science and Technology, Technical University of Kenya

ABSTRACT

Big Data provides an opportunity to educational Institutions to use their Information Technology resources strategically to improve educational quality and guide students to higher rates of completion, and to improve student persistence and outcomes. This paper explores the attributes of big data that are relevant to educational institutions, investigates the factors influencing adoption of big data and analytics in learning institutions and seeks to establish the limiting factors hindering use of big data in Institutions of higher learning. The study has been conducted through a desk search and reviewed sources of literature including scientific research journals and reports. The paper is based on desk research. The sources of literature that were reviewed included scientific research articles and journals, conference reports and theses. Online journals found on the internet were also examined with the search being broadened by Google Scholar where the following keywords were used “big data”, “developing countries”, “education systems” and “clustering”. The paper concludes that Big Data is important since it offers Universities opportunities to their Information Technology resources strategically to improve educational quality and guide students, colleges and universities see value in analytics; and therefore recommends that these institutions carry out investments in analytics programs and in people to have relevant data science. This is because Big Data can afford educational institutions opportunities to shape a modern and dynamic education system, in which every individual student can have the maximum benefit from, and can greatly contribute towards improving the quality of education

Keywords: Big Data, Analytics, Higher Education

Date of Submission: 25 May 2017



Date of Accepted: 16 June 2017

I. INTRODUCTION

Information Technology is a key enabler of many of today's technological innovations that have led to improvements that benefit society in many areas including education, business, security and health care. Technological advancements in digital technologies have led to an increase in the number of people accessing and using digital devices worldwide. [1] States that digital connectivity in Africa is driven by greater ownership of mobile devices, and is the driving force for the technological revolution on the African continent. Technological innovations, increased affordability and use of digital devices have led to a situation where extremely large amounts of data are being generated through use of these devices worldwide. Big Data Analytics can therefore help institutions of higher learning make use of data that is generated in their ecosystems. A report by [2], indicates that smart phone ownership rates in emerging and developing nations are rising at an extraordinary rate, climbing from a median of 21% in 2013 to 37% in 2015; and that overwhelming majorities in many countries surveyed reported owning some form of mobile device, even if they are not considered smart phones. A good example is Kenya, where smart phones have become increasingly common, while food is unfortunately scarce [3]. This trend is expected to continue, with data being rapidly generated and collected at a tremendous rate and will therefore contribute to towards big data. This pool of data when harnessed has the potential of helping to solve the statistical tragedy that has characterized many developing countries, and crises that are related to unreliable data, particularly in Africa. Big Data refers to data sets that are large enough to pose significant challenges when using commonly available or traditional tools and infrastructures to collect, manage, process and analyze the data within a tolerable amount of time. [4] Points out that Big Data is concerned with emergent suit of technologies that can process mass volumes of data of various types at faster speeds than ever before. [5, 6] observe that the attributes of big data are best described as high-volume, high-velocity and high-variety information assets which by their nature demand cost-effective, innovative forms of information processing that can enable enhanced insight and decision making. Big Data therefore changes the way we approach data analysis, inspiring entirely new families of information services and necessitating new processing models and knowledge representations. Big Data consists of extensive datasets; primarily in the characteristics of volume, variety, velocity, and/or variability that require a scalable architecture for efficient storage, manipulation, and analysis [7]. [8] Defines big data as distribution of data systems across horizontally-coupled

independent resources to achieve the scalability needed for the efficient processing of extensive datasets. [9] Describes Big Data as data that is too big in terms of volume, and moves too fast in terms of its acquisition and analysis speed that normally exceeds the processing capacity of conventional database management systems. Big Data is identified by a number of fundamental characteristics which include the following 1) Volume: which refers to colossal amounts of data collected that present a challenge to conventional information technology structures to store, process analyze and present. 2) Velocity: refers to the rapid pace at which information is generated and flows within an organization. The velocity dimension is also concerned with the capability of an organization understanding and responding to events as they occur [10]. 3) Variety: which refers to data in diverse formats both structured and unstructured. This variety of data formats come from interaction between people and the Information Technology systems such as sensors and web applications. 4) Veracity: which refers to the quality and authenticity of the data, and looks at how data that is being stored is meaningfully mined to address the problem being analyzed 5) Value: refers to the relevance of the data to Big Data Analytics and extent to which the data is utilized to generate value of the insights, benefits and business processes within an organization [4, 8, 11, 12, 13]. Big data comprises large and complex data sets which are difficult to process using traditional data processing applications and common database management tools.

II. OBJECTIVES

The paper was guided by the following specific objectives

1. To highlight the attributes of big data that are relevant to educational institutions
2. Investigate the factors influencing adoption of big data and analytics in learning institutions
3. Establish the factors limiting adoption and use of big data by Institutions of higher learning.

III. METHODOLOGY

This paper is based on desk research. The sources of literature that were reviewed included scientific research articles and journals, conference reports and theses. Online journals found on the internet were also examined with the search being broadened by Google Scholar where the following keywords were used “big data”, “developing countries” and “education systems”.

IV. NEED FOR BIG DATA IN ACADEMIA

Big data analytics is actively being used in many business organizations worldwide in business intelligence and in areas such as marketing and financial forecasting. In recent years Big Data has started attracting the interest of academia. Many academic institutions are moving to cloud architectures and with the increased use of digital devices by users in these ecosystems is leading to a situation more data is being collected in these institutions than ever before, creating considerable opportunities for using Big Data to analyze and correlate information that enhance decision making. [14] Observed that it is absolutely important for universities to use Big Data Analytics in order to deliver the very best of learning environments for the good of society. Possibilities for using Big Data to link research to education exist, and are achievable when latest research practices and outcomes are better used to inform teaching and in undertaking research activities as part of education. [4] notes that research into big data in higher education, especially in developing countries is limited despite growing interests in exploring and unlocking the value of the increasing data within higher education environment This paper seeks to contribute to the theoretical understanding of Big Data and Analytics within educational institutions of developing nations.

Big Data presents to Institutions a good framework for efficiently utilizing the vast array of data in shaping the future of higher education [4]. [4] Citing [15] states that Big Data application in higher education is attributed to technological innovations and developments; which have catalyzed the growth of analytics in higher education. According to [16], Data Warehouses and Cloud Computing coupled with greater ownership of digital devices by end users in the educational ecosystem are making it possible to collect, manage and maintain massive amounts of data. These Information Technologies are important resources that when exploited by policy makers are useful in driving institutional strategy and policy making for the future. Information Technology makes available sophisticated platforms that provide computing power necessary for ploughing through massive amounts of data, and turning these masses into meaningful information. Data mining technologies when applied use descriptive and inferential statistics to derive patterns from the massive amounts of collected data for actionable information [17]. An illustration of Big Data Environment is presented by figure 2 below. Big Data Analytics is relevant in addressing a significant number of pressing issues for education systems [14], key among them are 1) increasing educator effectiveness; 2) harnessing insights from learning experiences; 3) delivering education for all that may also be tailored for individual learners needs; and 4) equipping students with relevant skills for their future

careers. Institutions of higher learning are faced with pressure to improve quality of learning outcomes while cutting costs and to build new value relationships with students. This is achievable when these institutions have the capability of leveraging the data collected during the enrollment stages, finance sourcing and instructional process to support greater learning outcomes. Big Data technologies have been revolutionized by innovations in database technologies, computer hardware, especially memory and storage capacities, and increased availability of network bandwidth. According to [19], The Hadoop ecosystem which includes Pig, Hive, Mahout, and RHadoop is a good example of the Big Data technologies that can be deployed in educational systems. Hadoop and NoSQL databases are capable of enabling analysis of large heterogeneous datasets at unprecedented speeds. [20] Explain that Hadoop is an open-source framework for distributed computing that enables processing of large datasets through horizontal scalability. The Apache Hadoop project for instance develops open-source software for reliable, scalable and distributed computing. In fact, organizations such as Yahoo, Google, and Face book have invested in the Apache Hadoop Project with a view of addressing their Big Data needs. The Apache Hadoop software library framework allows distributed processing of large data sets across clusters of computers through simple programming models, and is designed to scale up from single machines to multiple machines each offering local computation and storage [13, 19].The Education sector is nowadays becoming more technology oriented. The digital revolution in developing countries is leading to development of new technologies such as ubiquitous computing devices and the Massive Open Online Courses all of which are radically transforming the mode and accessibility to teaching and learning [4]. These Massive open online courses (MOOC) are generating huge amounts of data that are relevant for Big Data Analytics [21]. [22] Underscores this by stating that the era of cloud and mobile computing is opening up many opportunities for revolutionizing education. Massive open online courses (MOOC) are a new phenomenon in higher education in which free online courses, open to an unlimited number of students are offered by Institutions of higher learning. MOOC describes courses offered fully online where learners can obtain certificates upon completion of the courses they enrolled in, and offers a paradigm shift in the delivery of learning. The virtual classroom is another important tool on which distance learning programs are based upon, and on which instructors can share learning material through the webinar. These tools, together with other learning management systems generate a lot of data which can be stored and analyzed for use in forecasting [21]. In such cases Big Data Analytics is useful in enabling institutions using such technologies to derive value from these sources of data. Improvements in Internet accessibility mean that big data analytics can be applied to help reduce the costs of education and improve student performances by examining and offering customized and self paced learning solutions for learners [23]. [24] Argue that higher education institutions that focus on better instructional quality and the acquisition and retention of students and faculty historically have used data warehouses and business intelligence tools to report on and analyze data. Big Data analytics has the potential to positively impact all the major areas that are of importance for an institution of learning; in areas such as student enrollment and retention, integrated information management and reporting, operational cost management, regulatory compliance and research. According to report by [18], through analytics learning institutions can perform thorough analyses of student and learning data to make informed decisions on future course offerings in order to cater for the needs of potential and existing students.

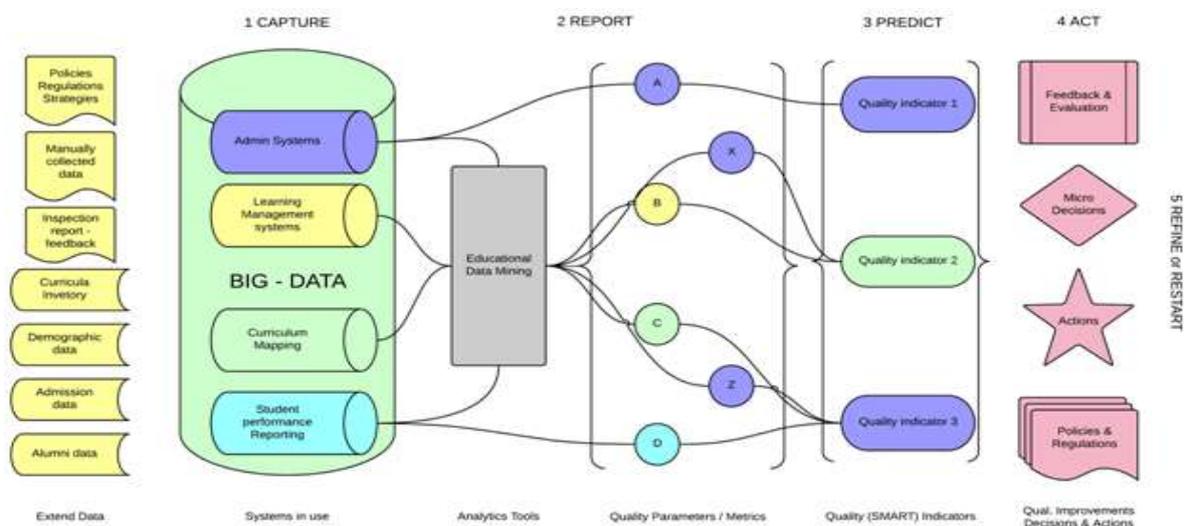


Figure 2: Analytics Driven Conceptual Framework: Adapted from [46]

V. BIG DATA ANALYTICS

According to [7], Information and Communication Technologies have become indispensable tools for supporting collaborative research work, and for this reason the critical role they play in supporting distributed teams can be enhanced with improvements in networking technologies and the development of tools and applications capable of supporting digital communications, access to and distribution of information. Information Technologies provide opportunities to the younger researchers and those who were not in major universities to participate in discussions with established researchers and those from prominent universities leading to an increased network of researchers and wider participation [12]. [13] States that new Information Technologies enable new communication structures that have radically reduced constraints of distance and time, and have also enabled creation of new and unique environments capable of supporting research and teaching. These new and emerging technologies afford university communities numerous opportunities to meet the challenge of the knowledge economy, offering new and novel possibilities for creating learning environments and applying knowledge in new ways in order to meet changing societal needs. The new Information Technology based infrastructures offer new possibilities for sharing of facilities and enables new forms of distributed organization. According to [13], the major components of the Information Technologies infrastructure in universities are: 1) Traditional packet switching networks and the newer point-to-point optical networks. 2). Middleware that enables integration of various technologies and makes it much easier to build university specific or inter-institutional virtual organizations in efficient, secure, and trustful ways. 3). Data and knowledge management services supported by vast networks of digital libraries, data sets, and archives providing content and sustainable knowledge management services. 4) High performance computation services capable of simulating complex phenomena such as galaxy formation or social-physical models of global warming. 5) User Interfaces and visualization services to support interaction between humans and the Information Technology environments. 6) Computer supported collaborative work to enable collaborative services to support and enable distributed teams to work together in ways similar to those in physical proximity. Various computer resources such as computer desktops, storage media, data and various input and output devices that are scattered across the university landscape can be connected via internetworking links through grid computing. Collaborative Research systems therefore can be presented as an amalgamation of grid computing technologies and research support systems in which Grid Computing functionalities are incorporated into Research systems. This way the collaborative Research Grid can be viewed as collection of computational resources on demand to match computational needs through generic service matchmaking on the Web.

VI. FACTORS INFLUENCING BIG DATA ANALYTICS IN EDUCATION

The decreasing costs of big data storage, open source software such as Apache Hadoop, NoSQL databases, network bandwidth and on-demand access to resources through cloud computing are bringing these complex technologies close to nearly everyone [23]. [30] Underscore this by stating that while the cost of Big Data and Analytic tools is coming down, they are becoming much easier to use. This is in turn opening up opportunities for use of these tools by enterprises and educational institutions to achieve better outcomes and more efficient use of resources. Expectations of good practices of accountability by stakeholders and increasing demands for evidence based guidelines to support decision-making are among the factors that are contributing to the emergence of Big Data in higher education. Therefore, finding appropriate technology platforms that will enable big data analytics and the relevant technological skills e.g. data scientists is the first step in providing a means to this end. This is because data can no longer be viewed as a back office accounts settling tool but rather a real time decision making tool that can be used by data scientists to derive useful information that would otherwise remain hidden in the terabytes of data [16]. Information Technology is reshaping education and expanding the education ecosystem beyond traditional lecture halls and classrooms to accommodate learners' who need flexibility in terms of time and distance through virtual and online learning systems [6]. The educational system is increasingly coming under pressure to respond to economic, political and social changes such as the need to increase the number of students in certain disciplines, and training graduates with skills and attributes required by industry [4]. For these institutions to be responsive to these changes, it is necessary that they use Information Technologies such as Big Data which has considerable opportunities in higher education. Educational data has in the past been gathered through traditional assessments mainly but is now increasingly being gathered through online educational systems, educational games and simulations leading to lots of data, which are in turn creating opportunities for Big data analytics. [17] states that in the context of today's technology enabled learning, the student is at the center of a network of information rich connections in the education system, and this is creating opportunities for Big Data analytics. [14] opines that institutions of higher education are operating in increasingly complex environments, and therefore it is imperative that they use Big Data in order to

deliver the best learning environments for the benefit of society. Possibilities exist for using Big Data to link research to education, by making good use of latest research practices and outcomes to inform teaching and to enable undertaking of research activities as part of education. In the course of learning processes, students generate massive volumes of potentially valuable information in areas such as course selection and registration, financial information, class participation, online resource usage, and textbook purchases, all of which are raw material for Big Data, and can be analyzed and redeployed to help administrators and students make better decisions. Institutions of higher education are collecting more data than ever, as a consequence of the digital revolution.

Many colleges and universities have demonstrated that analytics can help significantly advance an institution in such strategic areas as resource allocation, student success, and finance. Leaders of Institutions of higher education are looking into ways of initiating analytic programs in their institutions when they learn about transformations occurring at other institutions through big data analytics [31]. The McKinsey Global Institute Report on Big Data [33], states that the most developed regions such as Europe have had the biggest potential to create value through the use of big data. This is an encouragement for adoption and widespread use of big data in other regions especially developing nations. According to [32], the diffusion of big data provides an account of how Big Data moves from invention to widespread use and is facilitated by measures taken by service providers of key technologies needed to strengthen the resources and capabilities of academic institutions. [34] States that developing nations can tap into the many opportunities for big data that are available to create value from the massive amounts of data generated and in the long run help in their development. Big Data and analytics in higher education can be transformative, altering the existing processes of administration, teaching and learning and contribute to policy outcomes by helping address existing challenges facing educational systems [4, 35]. Successful Institutions have used big Data in the following ways 1) Creating a culture of completion and outplacement. 2) Reducing non productive credits. 3) Redesigning delivery of instruction. 4) Redesigning core support services such as human resources, academic services and finance to produce strategically useful data. and 5) Optimizing non core services and operations [17].

VII. BIG DATA IMPLICATION FOR EDUCATION IN DEVELOPING COUNTRIES

Many in development cite Big data as key to successfully implementing the Sustainable Development Goals. Digital learning is on the increase in developing nations, and is resulting in collection of massive data useful for analytics which can contribute to teaching and learning. Many students participate in social networks, online and mobile learning where new data are created. These new data are helping the students with the different background to correlate among themselves and is helping them understand core course concepts, and to expand their knowledge and skills [16]. These new computer supported interactive learning methods and tools often include intelligent tutoring systems, simulations and games which are capable of being used to create opportunities to collect and analyze student data from which patterns and trends can be discovered [36]. Big data concepts and analytics can thus be applied to various administrative and instructional applications, including recruitment and admissions processing, financial planning, donor tracking, and student performance monitoring [37].

According to [38] there is anticipation that in the developing world, more learning content will be delivered online through Virtual Learning Environments and supported by other digital systems including social media, in turn increasing the velocity of new student data. Tight budgets will however continue to impact higher education negatively in the developing nations leading to an increased focus on productivity and flexibility for institutions and students. According to [39], there are likely to be attempts to implement outcome based funding with major changes demanded in accountability, governance, core standards, completion rates and workforce development. Big data analytics has the potential to help administrators and teachers in developing nations make better judgements through data driven decision making supported by mining of large amounts of data [41]. [16] Noted that the use of data to guide instruction leads to significantly positive academic effects which can also be adopted in developing nations. Learning analytics is an area of research and application and is related to academic analytics, action analytics, and predictive analytics which have implication for academic systems. Learning analytics is related to educational data mining, and is an emerging research area with a series of computational and psychological methods which researchers can exploit to understand how students learn. According to [36], a key application of learning analytics is in identifying students at risk of failing a course or program of study, and in monitoring and predicting students learning performance and spotting potential issues early in time for interventions to be provided.

VIII. BIG DATA CHALLENGES IN EDUCATIONAL INSTITUTIONS

Analytics is a growing interest in many educational Institutions with many believing in its importance for the success of higher education. To improve the quality of learning outcomes, it is necessary that large volume of data generated by educational systems should be analysed effectively to facilitate appropriate responses to new challenges. Barriers to Big Data in Education generally include lack of executive vision, Inability to associate important business problems with big data solutions , Users or executives rooted in a old technologies, Cost , Lack of data warehouses and analytical tools, Data quality issues and simply leaving most data to be uncollected leading to no analyses. [31] Observed that widespread analytics use is limited mainly to the areas of enrollment management, student progress, and resource optimization. This limited scope of analytics is mainly due to barriers of affordability, data, culture, expertise, and communication. Utilizing the full potential in Big Data Analytics in education presumes existence of a functioning Big Data ecosystem in which there exist policies to support and facilitate the participating actors to collect, analyze, share, and distribute data and information. In reality achieving a well structured and functional big data ecosystem with clear incentives for all parties involved poses challenges in several areas. These are areas such as legislation, policy making, public administration and management [5]. This calls for development of workable models for big data. Big Data analytics in higher education also faces the challenge of finding the means to extract knowledge from the extensive data sets being generated daily and to distill the extracted data into usable information for administrators, students, instructors, and the public [40]. Pursuing big data is expensive, and as a consequence institutional leaders must be convinced that it will yield valuable results before they agree to invest in the technology. Therefore Information Technology leaders of these institutions should ideally demonstrate the importance of Big Data technologies in their Information Technology strategic plans, in addition to showing how the proposals they present address vital information issues [17]. This can help allay the fear that many big data projects don't have a tangible return on investment (ROI) that can be determined upfront [6]. Big Data also presents several analytical challenges that demand continual updating of tools and expertise, meaning that institutions need to have sufficient budgets to address these concerns. There also exist legitimate concerns about privacy, especially for data from online sources. This coupled with the digital divide in many nations presents obstacles to harnessing the power of Big Data for the benefit of the users of the educational systems [42]. Many challenges associated with the handling of Big Data are broadly due to its characterizing properties namely volume, velocity, veracity, variety and value. The challenge is in integrating the heterogeneous data sources and data types in this big data era because the data sources are diverse; for example there is need to integrate sensors data, cameras data, social media data, and so on and all of which are different in format, byte, binary, string, number, and so forth. Data Analytics requires communication with different types of devices and different systems including a means of extracting data from web pages [26].

IX. CONCLUSION

This paper has explored Big Data Analytics and its relevance in Educational systems with a view of helping educational institutions adopt Big Data Analytics. The paper has explored the attributes of big data that are relevant to educational institutions, the factors influencing adoption of big data and analytics in educational institutions and looked at the factors hindering use of big data in these Institutions. Big Data is far more than simply collecting information and generating reports. It is a strategic resource that can be used to improve educational quality. This paper recommends that educational institutions, particularly in developing countries be encouraged to make investments in analytics programs and in developing expertise in order to get value of big data. Big Data helps to cut costs and improve education by enabling administrators make decisions that are more specific and affords teachers valuable tools to choose from for a variety of learning. The Big Data approach to data management will help reduce difficulties associated with traditional data analysis; and this has the potential of enriching the education system with new learning ways, and making decision making by policy makers more efficient and targeted.

REFERENCES

- [1]. Juma, C. (2016). How can Africa master the digital revolution? Retrieved March 20, 2017, from World Economic Forum: <https://www.weforum.org/agenda/2016/04/how-africa-can-master-the-digital-revolution/>
- [2]. Pew Research Center. (2016, February 22). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Retrieved April 8, 2017, from <http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/>
- [3]. Ismail, N. (2016, September 8). Big Data in the developing world. Retrieved April 9, 2017, from Information Age: <http://www.information-age.com/big-data-developing-world-123461996/>
- [4]. Görnerup, O., Gillblad, D., Holst, A., & Bjurling, B. (2013). Big Data Analytics: A Research and Innovation Agenda for Sweden. The Swedish Big Data Analytics Network.

- [5]. Daniel, B. (2014). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*.
- [6]. Gartner. (2017). Big Data. Retrieved March 28, 2017, from <http://www.gartner.com/it-glossary/big-data>
- [7]. Gartner. (2017). Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest. Retrieved March 29, 2017, from <http://www.gartner.com/newsroom/id/3466117>
- [8]. NIST. (2015). Information Technology Laboratory. Retrieved March 17, 2017, from <https://www.nist.gov/el/cyber-physical-systems/big-data-pwg> ISO/IEC JTC 1. (2014). Big Data: Preliminary Report. Geneva: ISO.
- [9]. Daniel, B. K., & Butson, R. (2013). Technology Enhanced Analytics (Tea) In Higher Education. *International Conference on Educational Technologies*, (pp. 89-96).
- [10]. Sonka, S. (2016). Big Data Characteristics. *International Food and Agribusiness Management Review*, 19 (A), 7-12.
- [11]. Zikopoulos, P., Parasuraman, K., Deutsch, T., & Giles, J. C. (2012). Harness the power of big data The IBM big data platform. New York: McGraw Hill Professional.
- [12]. Elgendy, N., & Elragal, A. (2014). Big Data Analytics: A Literature Review Paper. P. Perner (Ed.): *ICDM*, 214–227.
- [13]. Mohebbi, A. (2015). Iterative big data clustering algorithms: a review. *Software: Practice And Experience*.
- [14]. Marsh, O., Maurovich-Horvat, L., & Stevenson, O. (2014). Big Data and Education: What's the Big Idea. *Big Data and Education conference*. UCL.
- [15]. Wagner, E., & Ice, P. (2012, July/August). Data changes everything: delivering on the promise of learning analytics in higher education. *EDUCAUSE Review*, 33–42.
- [16]. Drigas, A. S., & Leliopoulos, P. (2014). The Use of Big Data in Education. *International Journal of Computer Science Issues*, 11 (5), 58-63.
- [17]. Educause. (2013). *The Rise of Big Data in Higher Education*.
- [18]. Oracle Corporation. (2015). *Improving Higher Education Performance With Big Data*. Redwood Shores, CA 94065, USA: Oracle Enterprise Architecture White Paper.
- [19]. Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big Data Analytics for Security. *IEEE Computer and Reliability Societies*, 74-76.
- [20]. Luna, D. R., Mayan, J. C., García, M. J., Almerares, A. A., & Househ, M. (2014). Challenges and Potential Solutions for Big Data Implementations in Developing Countries. *IMIA Yearbook of Medical Informatics*, 36-41.
- [21]. Ray, S. (2013). Big Data In Education. *Gravity* (20).
- [22]. Kennedy, J. (2014). Characteristics of Massive Open Online Courses (MOOCs): A Research Review, 2009-2012. *Journal of Interactive Online Learning*, 13 (1).
- [23]. Ohri, A. (2015, March 6). Big Data Initiatives in Developing Nations. Retrieved April 10, 2017, from Big Data & Analytics Hub: <http://www.ibmbigdatahub.com/blog/big-data-initiatives-developing-nations>
- [24]. Stackowiak, R., Mantha, V., & Licht, A. (2015). *Improving Higher Education Performance with Big Data*. CA 94065, USA: Oracle Enterprise Architecture White Paper.
- [25]. Kurasova, O., Marcinkevičius, V., Medvedev, V., Rapčeka, A., & Stefanovič, P. (2014). Strategies for Big Data Clustering. *IEEE 26th International Conference on Tools with Artificial Intelligence* (pp. 740-747). IEEE.
- [26]. Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks*.
- [27]. Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., et al. (2014). A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE*, 267-279.
- [28]. Jain, A. K. (2011). *Data Clustering: 50 Years Beyond K-Means*. PRL.
- [29]. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *BioMed Central*, 1 (9).
- [30]. Segal, J., van Wyk, A., O'Flaherty, M., Simmons, M., Osinubi, F., & Yaiche, N. (2016). *Disrupting Africa: Riding the wave of the digital revolution*. PwC.
- [31]. Bichsel, J. (2012). *Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations*. Louisville, CO: EDUCAUSE Center for Applied Research.
- [32]. Micheni, E. M. (2015). Diffusion of Big Data and Analytics in Developing Countries. *The International Journal Of Engineering And Science (IJES)*, 2319 - 1805.
- [33]. McKinsey & Company. (2016). Retrieved March 20, 2017, from McKinsey & Company: <http://www.mckinsey.com/mgi/our-research>
- [34]. Apenteng, S. A. (2014). Big Data: A Tool for Development in Developing Nations Big Data: A Tool for Development in Developing Nations Big Data: A Tool for Development in Developing Nations Big Data: A Tool for Development in Developing Nations Big Data: A Tool for Development in. *International Journal of Scientific and Research Publications*, 4 (5).
- [35]. Baer, L. & Campbell, J. (2011). Game changers. *EDUCAUSE*. Retrieved March 24, 2016, from <http://net.educause.edu/ir/library/pdf/pub72034.pdf>
- [36]. Prakash, B. R., Hanumanthappa, M., & Kavitha, V. (2014). Big Data in Educational Data Mining and Learning Analytics. *International Journal of Innovative Research in Computer and Communication Engineering*, 2 (12), 7515-7520.
- [37]. Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16 (4), 9-20.
- [38]. Shacklock, X. (2016). *The Potential of Data Analytics in Higher Education*. London.
- [39]. United Nations Global Pulse. (2013). *Big Data for Development: A primer*.
- [40]. Western Interstate Commission for Higher Education (WICHE). (2015). Retrieved March 10, 2017, from Western Interstate Commission for Higher Education: <http://www.wiche.edu/info/publications/PI-2012PolicyInsights>
- [41]. Dede, C., Ho, A., & Mitros, P. (2016). Big Data Analysis in Higher Education: Promises and Pitfalls. Retrieved March 28, 2016, from <http://er.educause.edu/articles/2016/8/big-data-analysis-in-higher-education-promises-and-pitfalls>
- [42]. Spratt, S. (2015). *Ensuring Developing Countries Benefit from Big Data*. Brighton, UK: Institute of Development Studies.
- [43]. United Nations Global Pulse. (2013). *Big Data for Development: A primer*.
- [44]. Hervatis, V., Loe, A., Barman, L., O'Donoghue, J., & Zary, N. (2015). A Conceptual Analytics Model for an Outcome-Driven Quality Management Framework as Part of Professional Healthcare Education. *JMIR Medical Education*, 1 (2).
- [45].



Julius Murumba is a Doctor of Philosophy (Information Technology) student at Kibabii University, Kenya. He holds a Bsc and MSc (Information Systems), and is currently a Sessional Lecturer at the Technical University of Kenya. His research interests include Technology enhanced Learning, Cloud computing and Grid computing



Dr. Elyjoy M. Micheni is a Senior Lecturer in Information Systems and the Chairperson, Department of Management Science and Technology at The Technical University of Kenya. She holds a PhD (Information Technology) from Masinde Muliro University of Science and Technology, Master of Science (Computer Based Information Systems) from Sunderland University, (UK); Bachelor of Education from Kenyatta University; Post Graduate Diploma in Project Management from Kenya Institute of Management. .She has taught Management Information System courses for many years at University level. She has presented papers in scientific conferences and has many publications in referred journals. She has also co-authored a book for Middle level

colleges entitled: “Computerized Document Processing” Her career objective is to tap computer based knowledge as a tool to advance business activities, promote research in ICT and enhance quality service.