Tanzania Journal of Engineering and Technology 2022, **41(1)**:10-18
OPEN ACCESS articles distributed under Creative Commons Attribution Licence [CC BY-ND]
Websites: https://www.ajol.info/index.php/tjet/about; https://tjet.udsm.ac.tz

ISSN 1821-536X (**print**); ISSN 2619-8789 (**electronic**)
https://doi.org/10.52339/tjet.vi.768

Special Issue – First International Conference on Digital Transformation, 16 – 17, September 2021, Dar es Salaam, TANZANIA

Big Data Analytics Framework for Effective Higher Education Institutions

George Matto

ICT Department, Moshi Co-operative University, P.O. Box 474, Moshi, Kilimanjaro
Corresponding Author Email: gmatto2004@yahoo.co.uk

ABSTRACT

There has been an increased dependency on Information and Communication Technologies (ICTs) in undertaking various activities in Higher Education Institutions (HEIs) ecosystems. Because of that, huge volumes of data have increasingly been generated. There have been, for instance, considerable amounts of data generated through electronic platforms involved in students' admission and registration process, students' academic records management, teaching and learning data, curriculum related data, and several other administrative data. Analysis of data generated from these platforms stands to give students, lecturers, HEIs Management, policy makers and implementers, and other stakeholders useful insights that would help in improving HEIs' effectiveness. Unfortunately, literature have identified several challenges associated with existing big data analytics frameworks in HEIs. It was on this line that the present study, which was based on desk research, was carried out to propose an effective big data framework for analytics of such data. The proposed framework is composed of five stages; data collection, data pre-processing, data storage, data analytics, and data visualization. The stages were arranged systematically to address the identified challenges in the existing frameworks. Effective implementation of this framework will help HEIs to make a productive use of various data they generate. This will ultimately be beneficial to not only HEIs but also to aspired students, labour market, the government and the public at large.

ARTICLE INFO

First presented: **16-17 September, 2021**

Submitted: **22 February, 2022**

Revised: **29 April, 2022**

Accepted: **May 25, 2022**

Published: **15 July, 2022**

Keywords: *Higher Learning Institutions* (*HEIs*), *ICTs*, *Big data analytics framework.*

INTRODUCTION

The developments in Information and Communication Technologies (ICTs) have impacted all areas of our life. ICTs, according to Ejedafiru (2010), refer to technologies that transmit, store, create, display, share, or exchange information by electronic means. Matto (2015) pointed out that ICT is an umbrella term that includes any communication device or application, encompassing: radio, television, cellular phones, computer, and network hardware and software, satellite systems and so on, as well as the various services and applications associated with them, such as videoconferencing and distance learning. Today, ICTs have embedded themselves as commonplace entities in all aspects of life. One of the areas that have been impacted by ICTs is Higher Education Institutions (HEIs). Hanapiyah *et al*. (2018) pointed out

that the increasing introduction of ICTs in universities as well as other institutions of higher education is evident.

Information and Communication Technologies have been playing a vital role as a strong agent for change among several HEIs' roles and activities to optimize and streamline classroom, campus and institutional operations. Specifically, ICTs have been employed by HEIs to help in advertising academic programmes to prospective students so as to create awareness and engagement. Researchers, such as Cordero-Gutiérrez and Lahuerta-Otero (2020) contended that there has been an increased use of social media in advertising HEIs programmes which has maximized visibility and encourages student recruitment. ICTs have also been used by HEIs to admit students. According to Dubey and Pandey (2019) many HEIs have adopted the use of online admission systems because of the many advantages accorded by such systems. Tîrziu and Vrabie (2015) mentioned elimination of chaotic submission process, providing convenience for the appliers, no need to run in queues, providing transparency in the admission process, flexibility in timings, and making the admission process free from geographical boundaries as some of the advantages of online admission systems. Worldwide, there are various admission systems aiming at controlling quality in undergraduate admissions and widening access to higher education. And, in the case of Tanzania, all admissions to Universities are currently done through online systems (Mahundu, 2016).

The literature has further indicated that several other activities in HEIs ecosystem that were previously done manually are now done electronically via the use of ICTs. Students registration/enrolment processes for example, are done by the help of ICTs. Many HEIs are using Students' Information Systems to store and provide access to information about students, such as students demographics, courses information, instructor's information, class schedule, students class attendance, student's grades, and students billing and payment status, among many others (Saa *et al*., 2019; Pohekar, 2018). The ICTs are also used to improve the teaching and learning process where many HEIs have been developing or customizing Learning Management System (LMS) such as Blackboard, Moodle, and Massive Open Online Courses (MOOCs), and transfer the learning courses and the related data onto LMS (Basri *et al*., 2018). Additionally, HEIs have been using Library Automation Systems to provide an easy method for retrieving library information resources to readers unlike browsing through shelves (Msagati, 2016).

Because of the increasing dependency on ICTs in undertaking activities in HEIs' ecosystems, there have been increased generations of huge volumes of data. Such data are streamlined into databases, data warehouses and other storages in servers and other electronic platforms within HEIs premises, offsite data centers or in cloud. Since HEIs activities are ongoing such data keeps on growing. Hanapiyah *et al*. (2018) pointed out that nowadays HEIs are collecting more data than ever before. And, because the data come from different sources, they also in different formats. Some, such as those stored in relational databases, are structured while others, such as social media data, are either unstructured or semi structured. Ashaari *et al.* (2020) describes data that contains greater variety, arriving in increasing volumes and with more velocity as big data.

As Ashaari *et al.* (2020) explained, big data has been brought about by the increased data with a 3Vs model, that is, the increase of Volume, Velocity, and Variety. Volume means collection of massive data scale becomes increasingly huge. Velocity refers to the speed and dynamic nature of the data collection process and collection process and how to generate these data in real-time. And, Variety means many types of data including structured, semi structured and

unstructured data such as audio, video, text, and traditional structured data (Zulkarnain and Anshari, 2016; Ajah and Nweke, 2019). Literature has, however, added two more Vs in big data. These are Veracity referring to the trustworthiness of the data, and Value referring to how useful the data is in decision making (Hrehova, 2018). With the increasing areas of use of big data, additional and more Vs have increasingly begun to be added.

Big data, as equally as other massive datasets, contain useful insights (trends, patterns, and correlations) that when extracted may help to bring more accurate decision-making and better performance (Zulkarnain and Anshari, 2016). The process of uncovering such insights from large amounts of raw data is called big data analytics (Murumba and Micheni, 2017). Big data analytics has been used in multi-disciplinary areas. It is used in healthcare, for example, to predict epidemics, cure disease, improve quality of life and avoid preventable deaths. Government agencies are using big data analytics to fight fraud, waste and abuse. Big data is also used in agriculture, mining, security, communications and in several other areas (Memon, 2017).

Since HEIs generate huge amounts of data, they need to employ big data analytics to create better outcomes for students by helping instructors and HEIs management to answer the why questions around their institutional academic performance. The 'why' questions are such as why did a student not graduate? why did a student drop or fail a course? why did a student not master a particular concept or skill? (Tarmizi, 2019). If HEIs can be able to identify red flags around performance and behaviour, they can intervene before a poor assessment score results in a student failing a course or not graduating. This predictive capability is critical to intervention and is a cornerstone for effective HEIs.

Mining of educational data to extract useful insights such as predicting students' performance (Saa *et al*., 2019; Nahar, et al., 2021; Hooshyar *et al.*, 2020), discovering and decreasing dropout rates (Pal, 2012), discovering general students' behavior (Kularbphettong and Tongsir, 2012) and predicting instructor's as well as institutional performance (Tarmizi, 2016) have remained under a significant study over the past few years. The increased generation of data in recent years has boosted its importance. Unfortunately, there is lack of big data frameworks to guide on how the mining of such patterns can be achieved. It is on this line that the present study was carried out to propose a framework for analytics of such data. Specific objectives of the study were twofold; to establish challenges of the existing frameworks, and to propose an improved framework.

The remainder of this paper is organized as follows: Section 2 presents the methodological underpinnings used to carry out this study. Section 3 presents the challenges of the existing frameworks and the proposed framework, which are the basic findings of this research. And, section 4 concludes the paper and puts forward recommendations.

METHOD AND MATERIALS

This study was based on desk research. Documentary research as well as literature review were used to obtain data that informed the study. A total of 36 different literature, deemed important in enriching the study, including scientific research journal articles, conference proceedings and theses were reviewed. Most of these literatures were obtained online. In connection, a total of twelve (12) existing big data analytics frameworks were surveyed. Data analysis was done by using a content analysis method. Content analysis in this study involved thematic analysis in which, as per Ward*et al.* (2009), literature with similar themes (that is, similar challenges in the existing frameworks) were grouped together and analysed to obtain results. Results of the

analysis guided the development of the proposed big data analytics framework for effective HEIs.

RESULTS AND DISCUSSIONS

Challenges in the Existing Frameworks

With regard to big data framework in higher education, the study observed three kinds of challenges. First challenge was lack of frameworks that are designed specifically to extract patterns from educational data (Kim *et al*., 2014; Song and Ryu, 2015; Drogkaris and Gritzalis, 2015; Rajagopalan and Vellaipandiyan, 2013; Jimenez-Marquez *et al.*, 2019; Munshi and Yasser, 2017; Matto, 2019). The second challenge was limited sources of input data in the existing frameworks (Shamsuddin *et al*., 2019; Ashaari *et al.*, 2020; Otoo-Arthur and van Zyl, 2020; Mago and Khan, 2021; Al Hadwer *et al.*, 2019; Murumba and Micheni, 2017). And, the third challenge associated with existing frameworks was that most of them were introduced with no clear guideline on how the big data storage resources could be shared amongst different applications (Tekiner and Keane, 2013; Murumba and Micheni, 2017; Alsheikh, 2019). Below is a further discussion on each of the three observed weaknesses.

Concerning the first challenge, the study found that most of the existing big data analytics frameworks do not focus on extracting patterns from educational data. It was observed that most of them were focused on other areas such as healthcare, government, business, and security and deference. A framework by Kim *et al*. (2014), for example, focuses on u-healthcare systems that provide healthcare services based upon the analysis of the big data of vital signs. Similarly, Song and Ryu (2015) proposed a big data analysis framework on the same healthcare area. Drogkaris and Gritzalis (2015) proposed a big data framework for privacy preserving in e-government environments. Big data analytics framework in e-government is proposed also by a study by Rajagopalan and Vellaipandiyan (2013).

Another study by Jimenez-Marquez *et al.* (2019) proposed a big data framework for analyzing social media content. Specifically, the framework aims at knowing what customers are saying about the business through reviews in social media content. Moreover, a framework by Munshi and Yasser (2017) focuses on extracting patterns from the massive amounts of data evolving from smart grid meters used for monitoring and control purposes. Matto (2019) proposed a framework for extracting crime patterns from the growing volumes of data. This framework is aimed at assisting the Tanzania Police Force (TPF) and other law enforcement agencies in the country in improving strategies for crime detection and prevention.

For those studies that attempted to propose big data frameworks for extraction of patterns from educational data, the present study discovered that most of them were not exhaustive with regard to input data consideration. Shamsuddin *et al*. (2019), for example, proposed a big data analytics framework for smart universities implementations. In their proposed framework, the authors did not consider any input data. Similarly, Ashaari *et al.* (2020) proposed a conceptual framework for big data analytics technology capability and data-driven decision making in Malaysian HEIs without any input data consideration. This is considered a challenge because the type of input data normally determines the choice of the type of pattern mining method, thus if the type of input data is not known it makes it difficult to decide on the type of mining method(s) to be employed (Matto, 2019).

It was further revealed that there are studies that proposed big data frameworks with input data consideration. Unfortunately, the input data sources considered are not exhaustive with regard to the activities in HEIs' ecosystems. A study by Otoo-Arthur and van Zyl (2020), for example, proposed

a big data analysis framework which is made up of four layers: big education data source, information management, machine learning and analytics, and dashboards and visualization. The authors proposed various sources of big data in the first layer of their framework. The proposed sources were camera data, mobile app data, e-learning data, machine and sensors, administrative data, transaction and usage logs, email and messaging and social media data. Other essential sources of data such as programme advertisement data as well as admission processes data were, however, not included in their proposed framework. The framework proposed by Mago and Khan (2021) indicated the following input data sources: academic records, learning management systems, audio and video data, data through online learning platform, academic records, and text data. It was important, however, to include more other data sources to make the input data exhaustive. Similar challenge was observed in Al Hadwer *et al.* (2019) and Murumba and Micheni (2017).

The third challenge with the existing framework was lack of consideration on how the big data storage resources can be shared amongst different applications. Although big data analytics frameworks look at and across different dimensions of the problem, Tekiner and Keane (2013) contended that almost all current big data approaches are designed without coherent linkage or integration. Specifically, such frameworks lack deliberation on how the big data storage resources could be shared amongst different applications that would come under the big data framework. This was observed, for example, in Murumba and Micheni (2017) and Alsheikh (2019).

Proposed Framework

In order to address the identified challenges in the existing frameworks, this research proposes a more effective big data analytics framework for HEIs. In order for it to be exhaustive in terms of input data sources and more analytical in terms of processing and outputting the collected data, the framework comprises of five main stages. These are: data collection stage, data pre-processing stage, data storage stage, data analytics stage, and data visualization stage. These stages are arranged systematically to address the identified challenges in the existing frameworks. Figure 1 shows this proposed framework, and subsequent descriptions describes it more.

Data Collection Stage

All activities that generate data in the HEIs ecosystem are considered in this framework as potential sources of data. Consequently, the framework considers that data can be generated via HEIs academic programmes advertisement in various online platforms. They can also be generated from activities related to admission processing (this includes data for both admitted and non-admitted applicants), and from students' registration process. Other sources of data are those that come from Students Information Systems (this includes data such as students' academic and personal details, timetabling, accommodation, billing and payments data), Learning Management Systems (curricula and courses details), Library Management Systems as well as other relevant data related HEIs' administrative issues. These data are normally stored in HEIs' databases, data warehouses, flat files, social media and other platforms. Because of this, such data will be in different formats and structures which necessitates them to be pre-processed before being stored and analysed to obtain actionable information.

Data Pre-processing Stage

This stage involves three main activities: extraction of data, transformation of extracted data and loading transformed data into a centralized repository. Extraction involves obtaining data from different sources. During extraction, data cleaning such as filtering unnecessary and irrelevant

data, filling the missing values and annotation have to be done. Transformation involves integration of data from various sources, aggregating, splitting or formatting them. This stage is essential for transforming the data into schema-specific as per the demand of the big data warehouse. The load stage is involved with loading the transformed data into HEIs' big data warehouse.

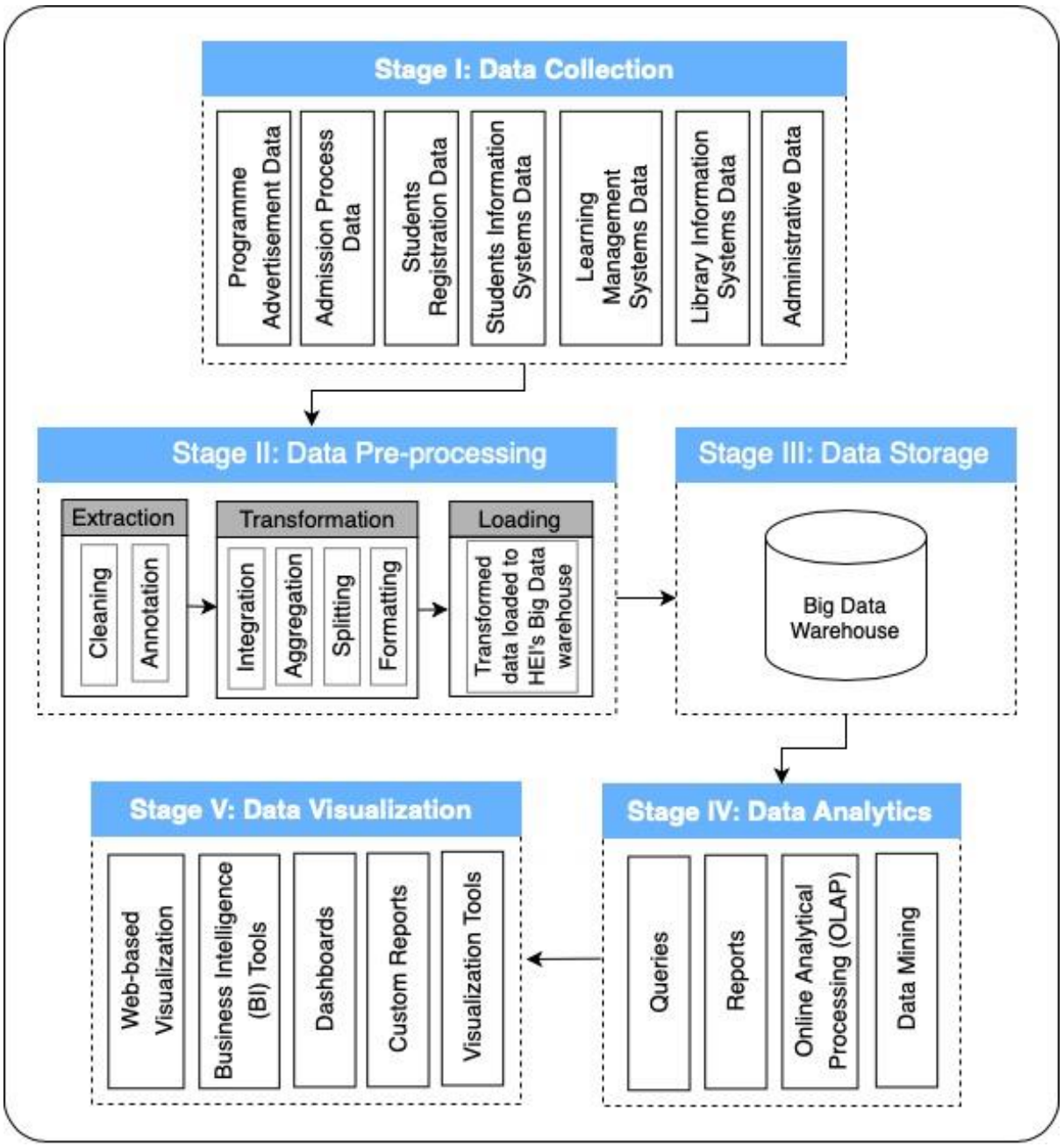

Figure 1: The proposed big data analytics framework for HEIs.

Data Storage Stage

As part of addressing the third challenge of existing frameworks as identified in this study, this stage is designed purposely to allow storage resources or data repository to be shared amongst different applications. After been extracted and transformed, data

from different sources as described in the data collection stage are loaded into the

designated big data warehouse in which different applications can easily access and do the analysis.

Data Analytics Stage

Once data have been transformed into a usable form in a big data warehouse, it has to be analyzed to generate actionable information. However, with the growing diversity in the nature of data, managing and analyzing diverse data set is becoming a very complex process. Thus, multiple analytics methods such as descriptive, diagnostic, predictive, prescriptive and cognitive analytics, that may suit the purpose are proposed. Data analytics in this framework involve running relevant queries, producing reports, performing online analytical processing queries and data mining.

Data Visualization Stage

This is the final stage where the analyzed data is made available to users in a form that is interpretable and integrated into existing processes, and ultimately used to guide HEIs' decision making. Basing on user's interest, visualization can be done via web, Business Intelligence (BI) tools, dashboards, custom reports as well as other relevant visualization tools.

Effective implementation of this framework will help HEIs to make a productive use of various data they generate by generating useful insights that would help in improving HEIs' effectiveness in terms of programme advertisements, students' admission process as well as the entire teaching and learning processes. This will ultimately be beneficial to not only HEIs but also to aspired students, labor market, policy makers, the government and the public at large.

CONCLUSION AND RECOMMENDATIONS

This study proposed a big data analytics framework for effective HEIs. The framework is proposed to confront three main challenges that were observed in the existing frameworks. The challenges are: lack of frameworks that are designed specifically to extract patterns from educational data; limited sources of input data; and, lack of clear guideline on how the big data storage resources can be shared amongst different applications. The proposed framework confronts these challenges through five stages: data collection stage; data pre-processing stage; data storage stage; data analytics stage; and data visualization stage. It is envisaged that through proper implementation of this framework HEIs will be able to make productive use of various data they generate by generating useful insights that would help in improving HEIs' effectiveness in terms of programme advertisements, students' admission process as well as the entire teaching and learning proces. The study recommends that HEIs should consider having a centralized repository of data generated from their different sources in which analytics could be done. In connection, the study calls for experimentation of the proposed framework as part of framework validation and possible improvements.

REFERENCES

Ajah, I. A. & Nweke, H. F. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data and Cognitive Computing*, *3*(2), 32. MDPI AG. Retrieved from http://dx.doi.org/10.3390/bdcc3020032.

Al Hadwer, A., Gillis, D., & Rezania, D. (2019, March). Big data analytics for higher education in the cloud era. In 2019 *IEEE 4th international conference on big data analytics (ICBDA)* (pp. 203-207). IEEE.

Alsheikh, N. A. (2019). Developing an integrated framework to utilize big data for higher education institutions in Saudi Arabia. *International Journal of Computer Science & Information Technology (IJCSIT) Vol, 11.*

Ashaari, M. A., Amran, A., Ahmad, N. H., Bakri, H., & Nazri, S. (2020). Big Data Analytics Technology Capability and Data-

Driven Decision Making in Malaysian Higher Education Institutions: A Conceptual Framework. *IOP Conf. Ser.: Mater. Sci. Eng.* 874 012021.

Basri, W. S., Alandejani, J. A. & Almadani, F. M. (2018). ICT Adoption Impact on Students' Academic Performance: Evidence from Saudi Universities. Education Research international. 2018, article ID 1240197, 9 pages.

Cordero-Gutiérrez, R. & Lahuerta-Otero, E. (2020). Social media advertising efficiency on higher education programs, *Spanish Journal of Marketing – ESIC*, *24(2)*: 247-262. https://doi.org/10.1108/SJME-09-2019-0075.

Drogkaris, P. & Gritzalis, A. (2015). A privacy preserving framework for big data in e-government environments. In International Conference on Trust and Privacy in Digital Business (pp. 210-218). Springer, Cham.

Dubey, P. and Pandey D. K. (2019). A Study on Use of ICT in Online Admission Process in Higher Educational Institutions in Chhattisgarh. International *Journal of Computer Sciences and Engineering – JCSE*, *7(3)*: 82 – 86.

Ejedafiru, E. F. (2010). Lack of ICT infrastructure as a barrier to resource sharing in Nigerian libraries. *Library Philosophy and Practice.*

Hanapiyah, Z. M., Hanafi, W. N. W. & Daud, S. (2018). Issues, challenges and opportunities of big data management in higher education institutions in Malaysia. *Indian J. Sci. Technol.*, *11*: 1-6.

Hooshyar, D., Pedaste, M., & Yang, Y. (2020). Mining educational data to predict students' performance through procrastination behavior. *Entropy*, *22(1)*: 12.

Hrehova, S. (2018). Brief overview of the concept of big data. In *3rd EAI International Conference on Management of Manufacturing Systems*. European Alliance for Innovation (EAI).

Jimenez-Marquez, J. L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J. L. & Ruiz-Mezcua, B. (2019). Towards a big data framework for analyzing social media content. *International Journal of Information Management*, *44*: 1-12.

Kim, T. W., Park, K. H., Yi, S. H. & Kim, H. C. (2014). A big data framework for u-healthcare systems utilizing vital signs. In 2014 *IEEE International Symposium on Computer, Consumer and Control*: 494-497.

Kularbphettong, K. & Tongsiri, C. (2012). Mining educational data to analyze the student motivation behavior. *International Journal of Information and Communication Engineering*, 6(8): 1032-1036.

Mago, B. & Khan, N. (2021). A Proposed Framework for Big Data Analytics in Higher Education. *International Journal of Advanced Computer Science and Applications*, 12(7).

Mahundu, F. G. (2016). E-governance: A sociological case study of the central admission system in Tanzania. *The Electronic Journal of Information Systems in Developing Countries*, 76(1): 1-11.

Matto, G. (2015). Factors Persuading ICTs Use among University Students in Africa and Applicability of the UTAUT Model: A Literature Survey. *Journal of Information Engineering and Applications*, 5(11): 9 – 15.

Matto, G. (2019). *Effective Mining of Crime Patterns from Growing Volumes of Data using Improved FP-Growth Algorithm*, Doctoral Dissertation, Nelson Mandela-African Institution of Science and Technology.

Memon, M. A., Soomro, S., Jumani, A. K., & Kartio, M. A. (2017). Big Data Analytics and its applications. *arXiv preprint arXiv*:1710.04135.

Msagati, N. (2016). Awareness and Use of OPAC by Distance Learners: The Case of the Open University of Tanzania. *Library Philosophy and Practice* (e-journal): 1482.

Munshi, A. A. & Yasser, A. R. M. (2017). Big data framework for analytics in smart grids. *Electric Power Systems Research*, *151*: 369-380.

Murumba, J. and Micheni, E. (2017). Big Data Analytics in Higher Education: A Review. *The International Journal of Engineering and Science (IJES)*, 6(6): 14 – 21.

Nahar, K., Shova, B. I., Ria, T., Rashid, H. B. & Islam, A. S. (2021). Mining educational data to predict students' performance. *Education and Information Technologies*, 1-17.

Otoo-Arthur, D. & van Zyl, T. L. (2020). A Scalable Heterogeneous Big Data

Framework for e-Learning Systems. In *IEEE 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*: 1-15.

Pal, S. (2012). Mining educational data to reduce dropout rates of engineering students. *International Journal of Information Engineering and Electronic Business*, 4(2): 1.

Pohekar, D. (2018). Role of ICT on Universities Administrative Services and Management. *International Research Journal of Engineering and Technology (IRJET)*, 5(11): 266 – 271.

Rajagopalan, M. R. & Vellaipandiyan, S. (2013). Big data framework for national e-governance plan. In *IEEE 2013 Eleventh International Conference on ICT and Knowledge Engineering*, 1-5.

Saa, A. A., Al-Emran, M. & Shaalan, K. (2019). Mining student information system records to predict students' academic performance. In *International conference on advanced machine learning technologies and applications*, 229-239. Springer, Cham.

Shamsuddin, N. T., Abdul Aziz, N. I. Che Cob, Z., Ab Ghani, N. L. & Drus, S. M. (2019). Big Data Analytics Framework for Smart Universities Implementations. *Springer Nature Switzerland.* SYMINTECH 2018, LNEE 565: 53–62.

Song, T. M. & Ryu, S. (2015). Big data analysis framework for healthcare and social sectors in Korea. *Healthcare Informatics Research*, 21(1): 3-9.

Tarmizi, S. S. A., Mutalib, S., Hamid, N. H. A., & Rahman, S. A. (2019). A review on Student attrition in Higher Education using Big Data Analytics and Data Mining Techniques. *International Journal of Modern Education and Computer Science*, 10(8), 1.

Tekiner, F. & Keane, J. A. (2013, October). Big data framework. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 1494-1499.

Tîrziu, A. M. & Vrabie, C. (2015). Education 2.0: E-Learning Methods. *5th World Conference on Learning, Teaching and Educational Leadership, WCLTA 2014*. *Elsevier*. *Procedia - Social and Behavioral Sciences*, 186(2015): 376 – 380.

Ward, V., House, A., & Hamer, S. (2009). Developing a framework for transferring knowledge into action: a thematic analysis of the literature. *Journal of health services research & policy*, 14(3), 156-164.

Zulkarnain, N. & Anshari, M. (2016). Big Data: Concept, Applications, & Challenges. *IEEE International Conference on Information Management and Technology (ICIMTech).*