International Journal of Information and Communication Technology Education
Volume 17 • Issue 3 • July-September 2021

Applying Big Data Analytics in Higher Education:
A Systematic Mapping Study

Adel Alkhalil, University of Ha'il, Saudi Arabia

Magdy Abd Elrahman Abdallah, University of Ha'il, Saudi Arabia & New Valley University, Saudi Arabia

Azizah Alogali, Saudi Standards, Metrology, and Quality Organization, Saudi Arabia

Abdulaziz Aljaloud, University of Ha'il, Saudi Arabia

ABSTRACT

Higher education systems (HES) have become increasingly absorbed in applying big data analytics due to competition as well as economic pressures. Many studies have been conducted that applied big data analytics in HES; however, a systematic review (SR) of the research is scarce. In this paper, the authors conducted a systematic mapping study to address this deficiency. The qualitative and quantitative analysis of the mapping study resulted in highlighting the research progression over the last decade, and identification of three major themes, 12 subthemes, 10 motivation factors, 10 major challenges, three categories of tools and support techniques, and 16 models for applying big data analytics in higher education. This result contributes to the ongoing research on applying big data analytics in HES. It provides a better understanding of the level of contribution to research as well as identifies gaps for future research direction.

KEYWORDS

Big Data, Business Intelligence, Data Analytics, Data Mining, Decision Support, Higher Education, Machine Learning, Systematic Review

INTRODUCTION

The last few decades have witnessed major developments in information and communications technologies, resulting in a significant increase the amount of "big data" (Tulasi, 2013). The big data (BD) generated from integrated technologies and systems shared by different organizations has also increased (Kaisler et al., 2013). This is due to applying artificial intelligence (AI) and Internet of things (IoT) technologies within systems across different sectors. Researchers' attention is now on how to exploit and implement BD to enable organizations to acquire new knowledge in response to emerging opportunities and challenges (Sin & Muthu, 2015).

Education institutions seeking to develop and to progress must employ big data analytics (BDA) to exploit creative opportunities and to access and develop ideas. The process of mining education data is typically referred to as educational data mining (EDM; Sin & Muthu, 2015). According to the International Educational Data Mining Society, EDM is defined as "an emerging field targeting the development of methods to explore data sets unique to the education settings" (International Educational Data Mining Society, 2011, p.601). As such, it is an interdisciplinary field that relies on

DOI: 10.4018/IJICTE.20210701.oa3

the application of machine learning, data mining (DM), statistics, recommender systems, psycho-pedagogy, and information retrieval to large volumes of educational data.

Higher education sectors (HESs) play a vital role in a nation's overall social and economic development. In turn, several factors contribute to establishing a quality HES, including goal-based processes, curriculum relevance in terms of discipline-specific subjects to meet business and industry needs, and the effective delivery of teaching and learning activities (Rajni & Malaya, 2015). HES are also recognized as an important contributor to technology-driven social advancement (Al-Shaya et al., 2012). The HES consists of several integrated dimensions: social, intellectual, cultural, psychological, human, and scientific, which all have a role in it realizing its goals and objectives (Basfar et al., 2011). HES services are framed around three core elements: inputs, processes, and outputs. Inputs include the students, management, faculty members, employees, materials, infrastructures; operations include teaching, training courses, curriculum development; and outputs include quality of graduates, training programs, scientific projects, research publications, conferences, and reputation (Namor, 2012). Further, big data analytics assists universities to accurately measure their key performance indicators (KPIs) and predict their future positions, thus leading to more informed decision-making and strategy selection.

The aim of this research is to explore how huge volumes of data generated from different technologies assist HEIs with their decision-making processes. More specifically, how universities collect and analyze relevant data to ensure accurate measurements of KPIs, and thus make informed decisions. The lack of a systematic mapping study in the research arena has motivated us to conduct a mapping study. In sum, this paper makes the following contributions to the field:

- identifies the research frequency, progression, and type over the last decade
- maps the challenges of applying BDA within HES domains
- analyses the tools and theoretical models to support the application of BDA in the HES

These contributions provide evidence for researchers to find solutions for specific issues in applying BDA in HEIs. They also support HEIs and BDA providers to better understand the existing models, frameworks, and tools, as well as the existing gaps and challenges.

This paper is organized as follows: section 2 discusses the background and similar studies; section 3 highlights the implemented research method; section 4 presents the results of the mapping study; and section 5 concludes the paper, including suggestions for future research.

BACKGROUND

Big Data and its Potential

The potential benefits of BDA are increasingly recognized by governments and major institutions across different business and industry sectors. Not long ago, data were confined to the category of organized folders, files, tables, and databases. The entry of new, quickly-transmitted, and structured or unstructured sources of information such as emails and videos, data from social media platforms (e.g., YouTube and Facebook), and chat messages (e.g., WhatsApp), has led to the formation of huge and unstructured datasets (Statistics Center, 2019, p. 4) Big data are measured in units such as terabytes, petabytes, and zettabytes, leading to the new challenge of how to take advantage of the data.

Processing and Analyzing Big Data

Data analysis is the science that forms the foundation for computer science, technical engineering, and software, and is widely taught in HES. Researchers describe data characteristics differently. Shah et al. (2015) identified three data characteristics: volume, velocity, and variety. Liao et al. (2014) identified four characteristics: volume, velocity, variety, and variability. Gandomi and Haider (2015) identified six characteristics: volume, velocity, variety, veracity, variability, and value. Sivarajah et al. (2017) identified seven characteristics: volume, variety, veracity, value, velocity, visualization, and variability. Current data processing technologies (e.g. databases and data warehouses) are largely insufficient to manage the amount of data currently generated around the world. Such large data volumes must be analyzed frequently and in a timely manner to optimize potential benefits. Hence, it is crucial for researchers to better understand the following types of BD. Structured data (data that can be processed, stored, and retrieved in a fixed format) is highly organized information that can be stored easily and smoothly and accessed from a database via simple search engine algorithms. Unstructured data (data that does not have any shape or structure) is disorganized making data processing and analysis difficult and time consuming. Semi-structured data (data that contains both the structured and unstructured formats), although not classified within a particular repository (database), contains information or dynamic tags that separate individual elements within the data (Turban et al., 2010; Witten et al., 2016).

With the influx of data and machines for conducting research, opportunities to extract data are ever-increasing. Indeed, data mining uses statistical, mathematical, and AI techniques to identify useful information and subsequent knowledge or patterns from a wide range of data (Turban et al., 2010; Witten et al., 2016). DM techniques can be grouped into three categories: association, prediction, and clustering (Turban et al., 2010). Prediction includes classification algorithms such as decision trees, artificial neural networks [ANNs], ANN/multilayer [MLP], rough sets, genetic algorithms and support vector machines, and regression such as linear/nonlinear regression, ANN/MLP and Support Vector Machines (SVM). Association includes correlation analysis (a priori algorithm) and sequence analysis. Clustering includes distant analyses, neural networks, and fuzzy logic (Turban et al., 2010).

The learning algorithms underpinning data extraction techniques are grouped as supervised or unsupervised. In a supervised learning algorithm, training data covers both the descriptive attribute and class attribute (Turban et al., 2010), whereas in unsupervised learning, training data covers only the descriptive attribute (Turban et al., 2010; Witten et al., 2016). Many software vendors provide powerful DM tools such as Statistical Package for the Social Sciences (SPSS), Statistical Analysis System (SAS), StatSoft and open-source DM tools, such as Weka (Turban et al., 2010).

Related Studies

Research and development on the utilization of BDA in HEIs has focused on two major areas: analyzing and predicting student performance and supporting management processes. In the context of analyzing and predicting student performance, nine studies of BDA in higher education (HE) are highlighted in this section. A systematic classification and mapping of the research utilizing BD in HE can help us to identify the research, its contributions, and its limitations, along with existing and emerging research. To systematically identify any existing study based on utilizing BD in HE, the authors executed the search string on various digital libraries and databases.

Jamiu et al. (2019) conducted a comprehensive systematic review on the use of Business Intelligence (BI) in HEIs to support decision-making processes. Enaro and Chakraborty (2019) conducted a study of research focusing on student performance analysis in HE and in predicting HES future performance. The study focused on one aspect of BD use with a limited number of studies included. Further, the study did not apply a systemic review of the research. Similar to Enaro and Chakraborty (2019), Saa et al. (2019) reviewed previous studies to analyze and predict student performance using DM techniques. The researchers identified four factors for predicting student performance and three DM techniques. Shuqfa and Harous (2019) also surveyed previous research and DM techniques used in the HE to predict student retention. The researchers focused on one aspect of BD use in HE with a limited number of studies included. Agrusti et al. (2019) reviewed previous studies in the area of predicting student retention in HE using DM techniques. Otoo-Arthur and Van (2019) conducted a systematic review of the framework and technologies to support BD applications in HEIs. Aldowah et al. (2019) conducted a systematic review to explore the potential of

DM in education to provide opportunities and solutions to various learning problems. The researchers identified the common DM techniques used in educational DM. Musa et al. (2018) reviewed studies to identify the success factors around using BI to assist decision makers to realize the benefits, and to reduce failure rates. Santi and Putra (2018) also reviewed the research on the use of BI to support decision-making in HEIs. The review explored three areas: tools, contributions, and applications.

The vast majority of these reviews covered research conducted up to 2018. There has been no mapping study in this area, nor a comprehensive study that covers all areas in using BD in HE. Therefore, the aim of this paper is to provide an up-to-date and comprehensive systematic review on the utilization of BDA in HEIs.

METHOD

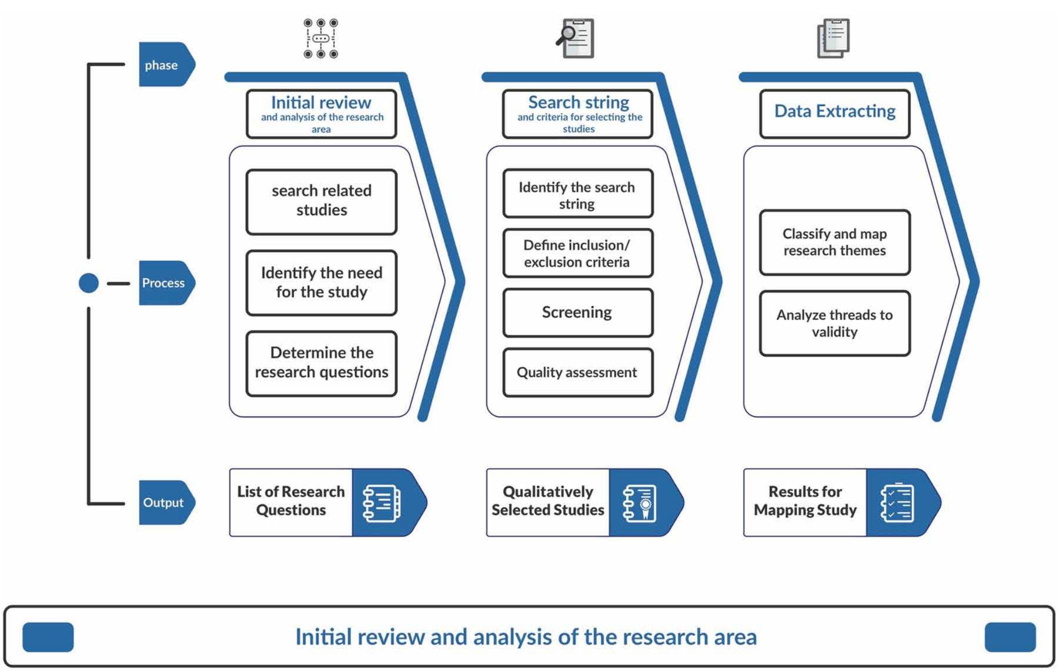
Mapping studies provide an overview of a research area and identify the frequency and type of research, and the results on a research topic. This mapping study followed the guidelines presented in Kitchenham et al. (2009) and consists of three phases, each with a number of steps (see Figure 1): (a) planning for the study, (b) conducting a search for primary studies, and (c) data extraction and analysis. These applications can reduce bias and threats to the research validity during vital phases of study planning, data collection, and results documentation.

Planning the Study

The first planning step was to identify the needs of the proposed mapping study (as identified and discussed earlier). This step resulted in specifying the research questions, which emerged in relation to three aspects: (a) what has not been addressed in previous similar studies, (b) the need to up-date answers to support continuous improvement in the research area, and (c) issue partly addressed but not explored compressively.

The following research questions (RQs) were formulated for this study:

Figure 1. Initial review and analysis of the research area



RQ1: What is the frequency and types of publications related to the application of big data analytics in HEIs over the last decade?
Objective: To highlight the growth and progression of the research on the application of big data analytics within HEIs.
RQ2: What type of research is published to support the application of big data analytics in HEIs?
Objective: To categorize the types of published research undertaken to address the gaps in the research area.
RQ3: What are the main themes of the research and how are they categorized?
Objective: To identify and categorize how big data analytics are applied and highlight to their contribution to the field?
RQ4: What factors motivate HEIs to implement big data analytics within their systems?
Objective: To identify the determinants and potential benefits underpinning the decision to implement big data analytics in HEIs.
RQ5: What are the critical challenges around the application of big data analytics in HEIs?
Objective: To identify the challenges and potential problems or hindrances related to the application of big data analytics in HEIs?
RQ6: What tools support the use of big data analytics?
Objective: To highlight the tool(s) that enables the automation and customization of processes for big data implementation.

Data Collection

Reputable databases and a well-defined search strategy were implemented to drive the data collection process. Specifically, a systematic and comprehensive three-step approach was adopted. Step one included searching the databases to identify potential articles. Step two included screening the Abstracts of the searched articles to remove duplicates and to assess their general suitability. Step three included applying the inclusion criteria to assess for quality literature. All articles meeting the inclusion criteria were subsequently included in t review.

Keywords (Search Strings)

PRISMA-compliant methodology guided the search process (Moher et al., 2010). First, keywords were identified and used to build a search string as follows: “big data” OR “business intelligence” OR “data mining” OR “large data sets” OR “knowledge discovery” AND “higher education” OR “tertiary education” OR universit* OR college* OR facult*. The identified search string was applied to five databases: Scopus, ScienceDirect, ACM, IEEE, and Web of Science.

Inclusion/Exclusion Criteria

The initial search results identified 1,080 papers. Duplicate articles were excluded. The inclusion criteria following were formulated to guide the selection process:

- Research includes the two main variables: BD and higher education.
- Research reflects results that support development of the higher education sector.
- Clear methodology applied
- Contributions of the study rigorously evaluated

The following exclusion criteria were also formulated:

- Did not share the same aim of this study
- Not related to educational use of big data in higher education
- Notes

- Workshops
- Book chapters and books

Screening

Before applying the criteria, the search results were screened by title and abstract. Each of the abstracts was read, and those matching the exclusion criteria were excluded. A total of 717 studies moved to the next round of filtering that examined the full text.

Qualitative Assessment

A quality assessment of the articles was performed to select only those articles closely connected to data management and analytics in HEIs, thus justifying their inclusion. Following the quality assessment, 79 articles were excluded from the review.

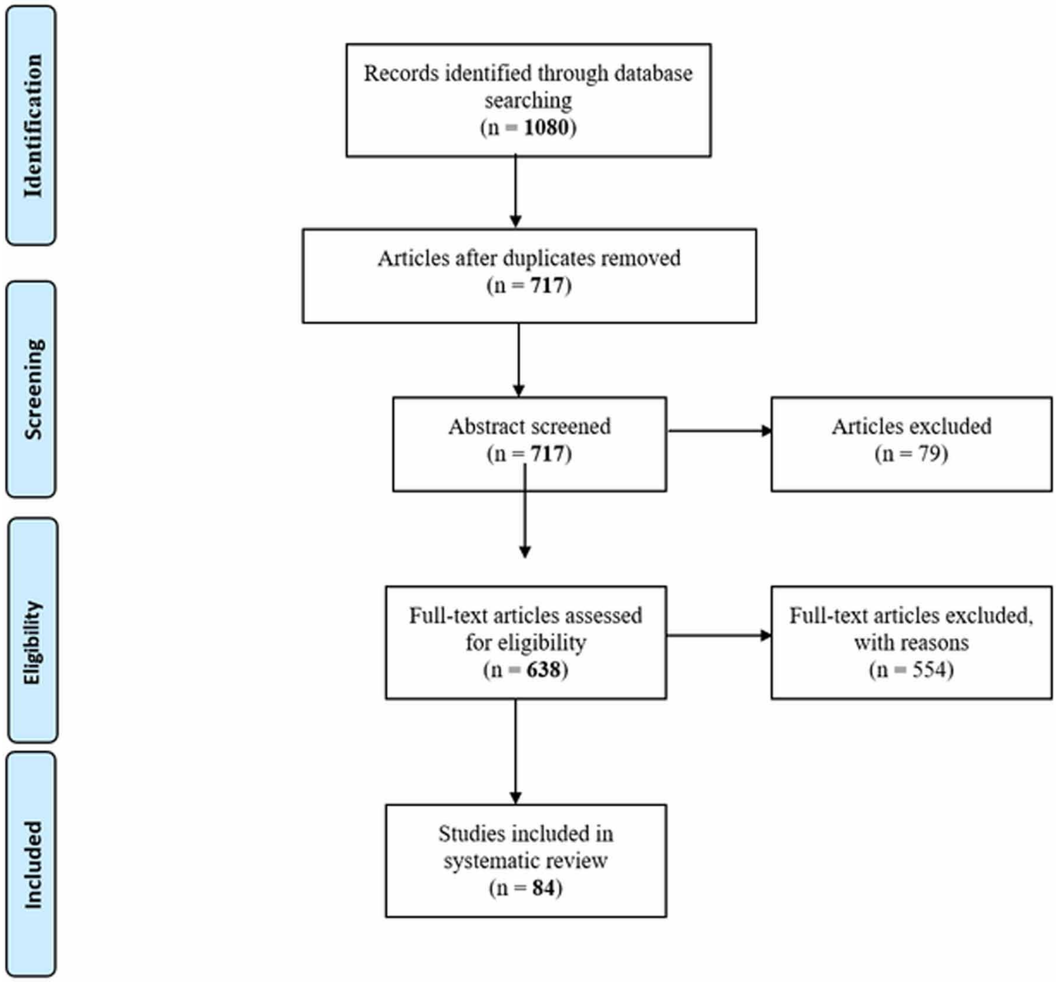
The remaining 638 studies were double reviewed by two researchers to further refine the number of studies. The researchers applied a value of 0 or 1 to each article for each inclusion criteria identified. The article was considered low quality when its value was 2 or below, medium quality when its value was 3, and high quality when its value was 4. A total of 84 items were rated high-quality and included for review (see Figure 2). Hence, a content analysis approach was applied at this stage. This is fundamentally about identifying the presence of particular words, themes, or concepts within the texts to later quantify and analyze for their relevance to the research questions (McAlister et al., 2017). To support the content analysis in this study, the text was analysed at the theme / concept levels and was coded according to its links to the any or all of the six research questions. The coding was based on the existence of the theme / concept rather than their frequency of use (McAlister et al., 2017). This was performed to break down the contents of the text into manageable themes and categories for quantitative analysis.

Inter-rater reliability is crucial to establishing trustworthiness in the research process in terms of reliability around the interpretation of data and the reported insights from the data analysis. Following each researcher's content analysis of the articles, a randomly selected sample of 10 articles was achieved to assess for inter-rater reliability. Specifically, a Table was created that included four elements: the page number and sentence / line from which the eligibility criteria was identified, the code name (the nominated eligibility criteria), the article title, and the name of the coder. The outcomes for each coder (researcher) were compared, with agreements and disagreements on the coding choices tallied for each article. Disagreements in the coding were identified and discussed, and changes were made to the coding assigned to the text based on consensus. Inter-rate reliability was calculated based on the number of agreed codes divided by the total number of codes. The desired agreement was no less than 90% agreement.

Documenting the Study

The selected studies were imported into Excel with attributes of reference identification, authors, year, and title. Minimizing threats to the validity of this research was attempted in all steps; namely, the search attempted to retrieve all possible and related studies and the scoring method ensured an unbiased selection of the studies. Issues around the availability of structured approaches to systematically producing and reporting results presented a potential threat to the validity of the research findings. This threat was minimized by developing structured tables with specific attributes on which the data extracted was based, and then statistical analysis was performed.

Figure 2. PRISMA flow chart(Moher et al, 2010)



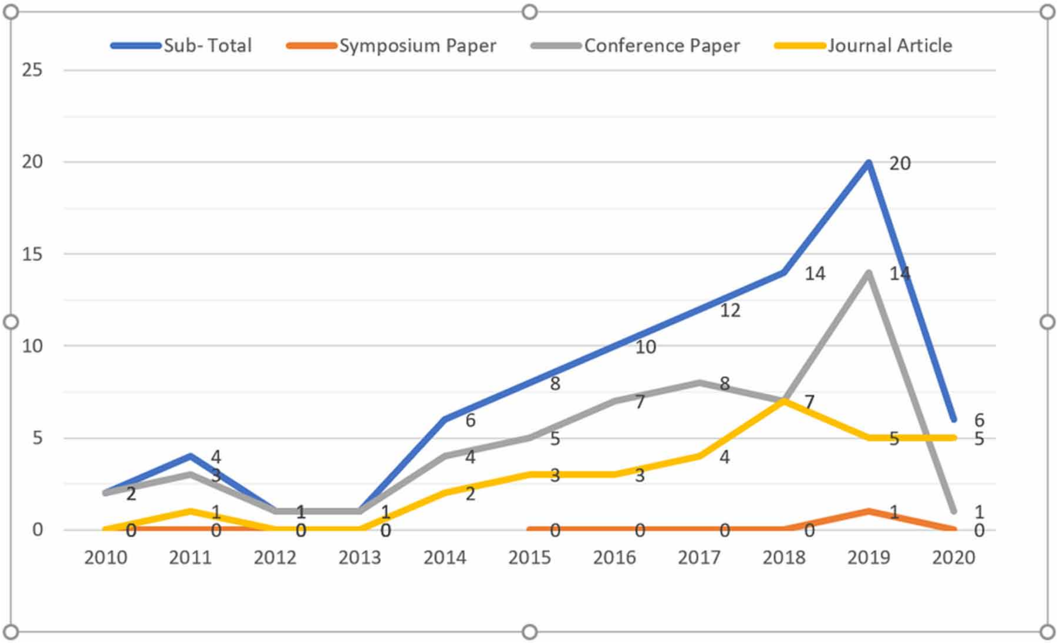
RESULTS AND THE ANALYSIS OF THE MAPPING STUDY

Frequency and Types of Published Research

This section aims to answer RQ1 and RQ2 by highlighting the frequency, progression, focus areas, and types of research publications. Figure 3 shows the results of data analysis to answer RQ1, showing the number of studies published on application of BDA in HEIs over the last decade. The type of artefact (journal article, conference paper, and symposium paper) is also provided.

Figure 3 shows that the number of studies conducted during the last decade was 84, including 30 journal articles, 53 conference papers, and one symposium paper. About two-thirds were published through scientific conferences, far outnumbering the articles in scientific journals. This can be explained by the fact that most journals require a detailed and clear methodology, empirical data collection, and validation of results, which were achieved in only one-third of the studies. This indicates that the maturity level of the research is still at an early stage, and comprehensive studies need to be conducted to further enhance the findings of the conference papers.

Figure 3. Number of types of research publications from 2010 to 2020



Further, Figure 3 shows the number of studies from 2010 to 2015, was limited to 22. Between 2015 and 2020, the number of studies gradually increased from 10 in 2016, 12 studies in 2017, 14 studies in 2018, peaking at 20 studies in 2019. In 2020, the number of studies decreased to 6.

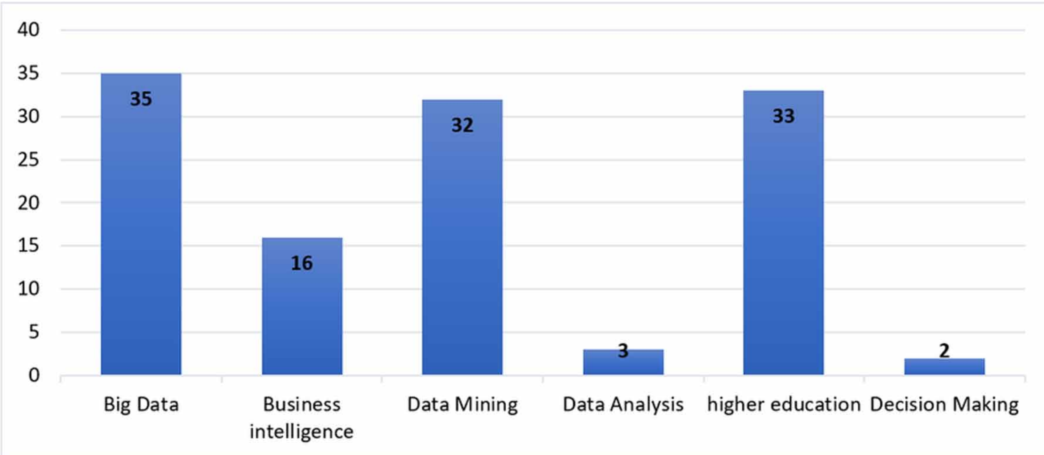
Figure 4 shows that most of the emphasized terms were in studies published from 2010 to 2020. For example, there were 35 iterations of the term ‘big data’. This indicates that the selected studies attempted to implement innovative BD technology into HE systems. The number of iterations of the term ‘higher education’, which is a fundamental variable in this study, was 33. This was followed by 32 mentions of ‘data mining’, which makes BD usable by identifying patterns and important relationships. ‘Business intelligence’ was used 16 times, ‘data analysis’ 3 times, and ‘decision-making’ 2 times.

This section discusses the findings for RQ2, the types of published research. Shaw (2002) classified research contributions into six distinct types (see Figure 5): specific solution, validation model, procedure or technique, evaluation report, descriptive model, and analytic model. Despite this diversity, most studies were evaluation reports, which represents less than half the types of published research. This was followed by procedure or technique studies (22.6%), descriptive model studies (16.6%), and analytic model studies (5.9%). Validation model and specific solution studies had an equal number of iterations (3.6%).

Based on the results for RQ1 and RQ2, the authors concluded that studies of the use of BDA in HEIs have received reasonable attention from the research community. However, the number of relevant publications emerged mostly as conference proceedings. Therefore, there is a need for comprehensive solutions and their evaluation, and studies to validate previously proposed models and frameworks in terms of journal-based research.

As shown in Figure 5, most published studies are evaluation reports (e.g., Apraxine & Stylianou, 2017), followed by procedure or technique studies (e.g., Rodzi et al., 2015), then descriptive model studies (e.g., Ijab et al., 2019). This signifies the need for more practical research outcomes and automated tools that can support HE to adopt BDA. These findings motivated the researchers to investigate the studies to develop patterns in the research area, explain processes, and develop tools and frameworks.

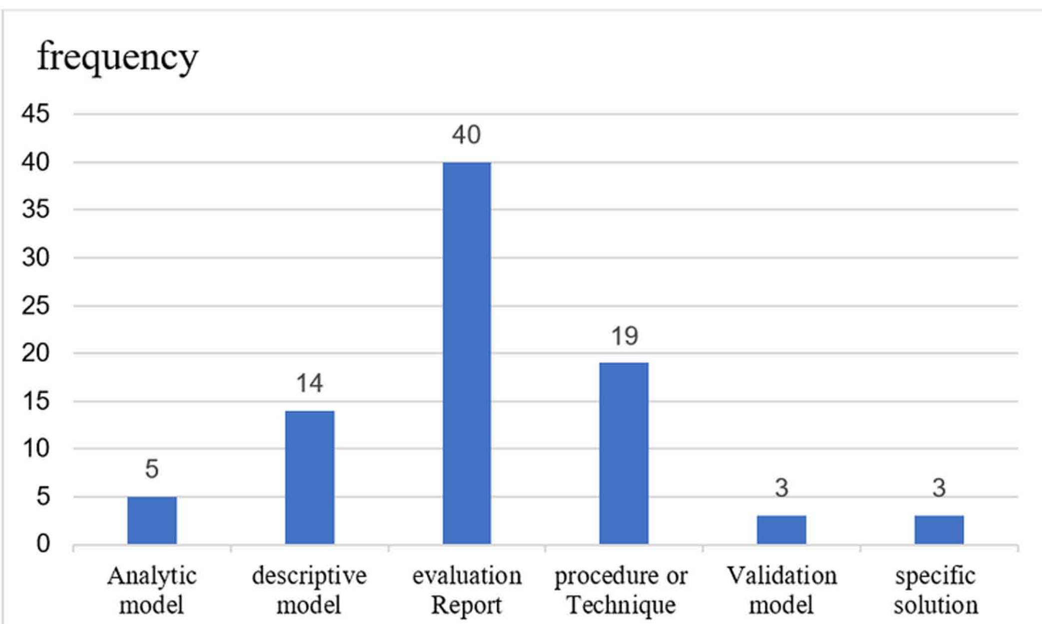
Figure 4. Keywords for studies from 2010 to 2020



Research Themes, Motivations, and Challenges

The taxonomy shown in Table 1 presents a systematic mapping and representation of the main research themes around the use of BDA in HEIs to emerge from the 84 studies. It also highlights the challenges identified around the application BDA. The analysis shows the three main themes were educational quality, decision-making process, and information management. Within these three themes, 12 subthemes emerged: five in educational quality, four in decision-making process, and three in information management. These findings highlight the level of contribution in each HES aspect. The

Figure 5. Types of research



authors also mapped the findings of the research teams with the challenges identified in the research areas. This helps to address the specific challenges for each part of HE system.

Improving educational quality is the main reason for applying BDA in HE, with about half of the studies focused on this area. In turn, improving student performance is identified as the main area of focus to enhance educational quality. In particular, the predictive feature of BDA is widely used to predict student performance. This allows universities to plan around specific courses for students and to predict and limit the number of student dropouts. Within this theme, a number of challenges were identified. Large volume of data and data variety were identified in more than half of the studies. Teaching and learning is also associated with improving education quality in HEIs using BDA. Specifically, it supports universities to identify deficiencies in their teaching methods. For example, Nengah et al. (2019) used a clustering DM method to analyze student tuition using a fuzzy C-means algorithm. Further, e-learning is given focus when using BDA to enhance education quality (Falakmasir et al., 2019; Zhang et al., 2015). For example, Falakmasir et al. (2019) showed how to apply BI and online analytical processes (OLAP) in e-learning to monitor and analyze learner behavior and performance in e-learning environments. They also revealed how BI and OLAP can be used to evaluate the courses contents and their effectiveness in the learning process. Large volume of data (23.8%) and lack of organizational resources and skills (23.8%) are the main challenges to applying BDA to enhance teaching and learning in HEIs (see Table 1).

Quality assurance and accreditation processes also received research attention for improving educational quality using BDA (e.g., Liu & Chang, 2016; Pérez et al., 2018; Skalka et al., 2013). For example, Skalka et al. (2013) developed the DEK (Data–Evaluation–Knowledge) model, an information support system to support universities to meet quality assurance requirements, which divides the problem of internal guidelines for quality assurance implementation. Lack of organizational resources and skills (33%) followed by large volume and variety of data (25% each) are the main challenges to applying BDA in quality assurance and accreditation processes in HESs (see Table 1).

The integration of machine learning and DM technologies supports the emergence of smart universities (smart campus) by applying BDA within their systems. For example, Shamsuddin et al. (2019) developed a conceptual framework to create a smart campus through BDA. The aspect to receive the least research attention is resource utilization. Huan and Bo (2018) developed a MapReduce parallel K-means cluster algorithm to support universities with BDA for resource management. Large volume of data, and concerns about privacy are discussed in Shamsuddin et al., (2019), Tong and Yi (2013), and Yang (2018) as the main challenges around utilizing BDA for smart campuses.

Improving decision-making processes in HEIs was associated with four subthemes: overall implementation, support for academic affairs management, student support, and service management. The implementation of BDA to support different HEI activities was the most popular research aspect. Due to the generality of this subtheme, several challenges were identified in the selected studies regarding the overall implementation (Table 1). Support for decision-making processes related to academic affairs was the main focus in other studies. For example, Boulila et al. (2018) identified a BI solution to support decision making for different users involved (students and staff). It includes gathering data from different sources, utilizing different operations (extract, transform, and load), proposing different solutions to describe academic procedures, and visualizing the outcomes. In this regard, data integrity and implementation complexity were discussed as the main challenges in 23.1% of studies. Further, enhancing decision-making processes to support student career pathways also received research attention. For example, Elhassan and Klett (2015) studied the alignment of learning and competencies, human resource strategy, and market dynamics. They create a bridge between HEIs and businesses for improving graduate employability and identify HE adaptation to market demands as an emerging strategy. Large volume of data was a main challenge in this subtheme (33.3%). Other studies focused on decision-making processes in HEIs by identifying the deficiencies in their systems. For example, OPREA (2017) proposed a model for use in HEIs to analyze activities

Table 1. Mapping the research themes with implementation challenges

Themes	Educational Quality				Decision-Making Process				Information Management				
Challenges	**Student performance**	**Teaching and learning**	**Quality assurance**	**Resource utilization**	**Smart campus**	**Support for academic affairs**	**Overall application**	**Student support**	**Identify deficiencies**	**Financial information**	**Collaborate**	**Library management**	**%**
Large volume of data	SR48, SR11, SR56, SR74, SR3, SR26 29.4%	SR52, SR13, SR38, SR14, SR35 23.8%	SR44, SR9, SR47 25%	SR70, SR71 50%	SR16, SR59 66.6%	SR75, SR1, SR73 15.4%	SR2, SR50, SR78, SR80 17.9%	SR25, SR80, SR63 33.3%	SR4, SR42, SR23 24.8%	SR60 25%		SR28, SR29 50%	26
Lack of resources or org. skills	SR36 5.9%	SR20, SR55, SR51, SR14, SR35 23.8%	SR77, SR58, SR39, SR47 33.3%	SR_12, SR_71 50%		SR46 7.7%	SR19, SR22, SR45, SR80 14.3%	SR_63 11.1%	SR_42 14.8%	SR60 25%			16.3
Data integrity		SR40, SR8, SR57 14.3%				SR76, SR82, SR64 23.1%	SR15, SR21, SR31, SR67, SR68, SR78, SR80 25%	SR84, SR63 22.2%					13.1
Un-structured data	SR36 5.9%	SR7 SR69, SR83 14.3%	SR47 8.3%				SR53, SR62 7.1%		SR4 14.8%				6.5
Variety of data	SR37, SR6, SR11, SR6 23.5%	SR83 4.8%	SR10, SR9, SR47 25%			SR1 7.7%	SR2, SR65 7.1%	SR63 11.1%	SR23 14.8%				10.6
Privacy concerns	SR48, SR26, SR66 11.8%				SR49 33.3%	SR18, SR24 15.4%	SR34, SR80 7.14%					SR41 25%	6.5
Data timeliness	SR81 5.9%	SR33 4.8%					SR17, SR68 7.1%						3.3
Setting data standards	SR63 5.9%		SR47 8.3%				SR80 3.6%	SR_63 11.1%		SR70 25%			4.1
Complexity	SR72, SR5 11.8%	SR54 4.8%				SR18, SR24, SR82 23.1%	SR17 3.6%		SR43 14.8%	SR79 25%	SR30 100%	SR61 25%	8.9
Data analytics (costs)		SR14, SR35 9.5%				SR73 7.7%	SR32, SR80 7.14%	SR63 11.1%					4.9

to identify the gaps around making informed decisions. Large volume of data was also the main challenge in this subtheme (24.8%).

Information management is the third theme identified by researchers to support HE systems, three subthemes identified. First, some researchers focused on supporting HEI process for managing and storing financial data. For example, Yang (2018) explained how BDA can be implemented to address these issues, however further research with automated tools is required to provide HEIs comprehensive support in this regard. Large volume of data, lack of resources and skills, standards and implementation complexity were identified in Dong (2017), Sihui and Xueguo (2016), and Utomo et al. (2017) as the main challenges in this area. Second, collaboration among HEIs using BDA is increasingly important. For example, Ken (2019) proposed a BD intelligent sharing system based on multi-layer architecture. It includes an algorithm of BD intelligence and uses statistical characteristics to shares the creative achievements of students beyond the university audience. Implementation complexity is identified by Wu (2020) as the main challenge to achieving this feature. Third, library management was examined by some researchers. For example, Liu (2018) discussed a filtering algorithm to conduct DM for book borrowing in libraries. Large volume of data was the main challenge in this subtheme (33.3%).

Algorithms, Tools and Techniques

A crucial question in HE around the translation of data into meaningful results is regarding data patterns (Witten et al. 2016). This has resulted in a new generation of automatic data processing using complex methods and algorithms—namely, DM (Turban et al., 2010). The articles in this review have been classified into three independent or combined categories of DM: prediction, clustering, and association rules, as illustrated in Figure 6.

Figure 6. Classifications of reviewed articles

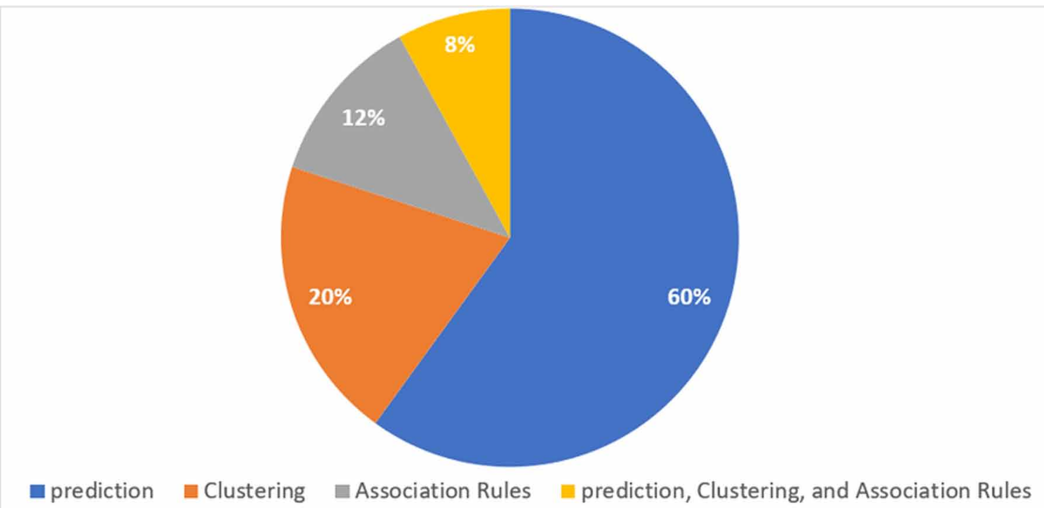


Table 2 shows 15 of 26 articles fell into the prediction category; with all studies focused on predicting student performance. Two articles investigated student dropout rates, applying the naïve Bayas and logistic regression algorithm for this purpose. Three studies used DM algorithms to predict financial outcomes. Table 2 illustrates that the remaining four articles tested the application of different algorithms to predict: teacher morality, teacher ability, research status, and student admission into university. Lastly, one study used DM algorithms to predict factors affecting academic performance and another study use the algorithms for predicting improvements in HEI management (see Table 2).

The clustering category DM algorithm method was used in eights studies overall (see Table 2). They included clustering: (a) student personality characteristics and learning status in distance education and HEIs resource utilization; (b) student tuition outcomes and faculty performance; (c) factors affecting HES retention/attrition rates and academic performance; and (d) improving HES e-learning and management processes. Lastly, the association rules DM algorithm was utilized in five studies: faculty evaluation, help in HES evaluation systems, library book management systems, improving HES e-learning and HES management processes.

This review also focused on the most commonly used open-source platforms in the literature, including Hadoop (see Figure 7). Nine articles used Hadoop to analyze BD in HEI-related contexts. Weka was the second most popular platform, used in five research studies.

KEY FINDINGS AND THEIR IMPLICATIONS

In sum, this study found that research interest in how HEI's use BD within their operations is strong. The main domains of research interest to emerge were on HEI's uses of BD for insights on educational quality, to shape decision-making process, and for information management. Regarding the previous research findings specifically, this study found that the application of BDA to support teaching and learning in HEIs presents the most challenges. In terms of methodology, this study found that previous research typically employed predication tools (e.g., classification algorithms and regression algorithms) to identify patterns in BD. Moreover, despite the research proposing up to 16 frameworks for analyzing BD, this study found that a framework for integrating BD into HEIs remains inadequate. Lastly, this study found that the research remains at an immature stage overall, with the majority of publications being conference papers rather than the outcomes of formal empirical research.

The implications of these findings are related to two outcomes in particular. As previously mentioned, a key affordance for HEIs associated with effective applications of BDA is improved decision making. Hence, the finding in this study that the research on HEIs' uses of BDA remains at an immature stage overall has implications for the extent to which HEIs can rely on current BDA models and frameworks to support enhanced decision-making processes. The study findings also have implications for the cost-effectiveness of BDA applications by HEIs. It may be reasonably argued that the research needs to target a higher level of maturity to inform best-practice applications of BDA to optimize the cost-effectiveness outcomes.

Study Limitations

The main limitation of this study is associated with the basic challenge of trying to summarize a research field according to a set number of studies. This study acknowledges that the aims and methodologies of a research investigation will vary from study to study. Nonetheless, as attempt has been made to generalize the results despite substantial heterogeneity.

CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK

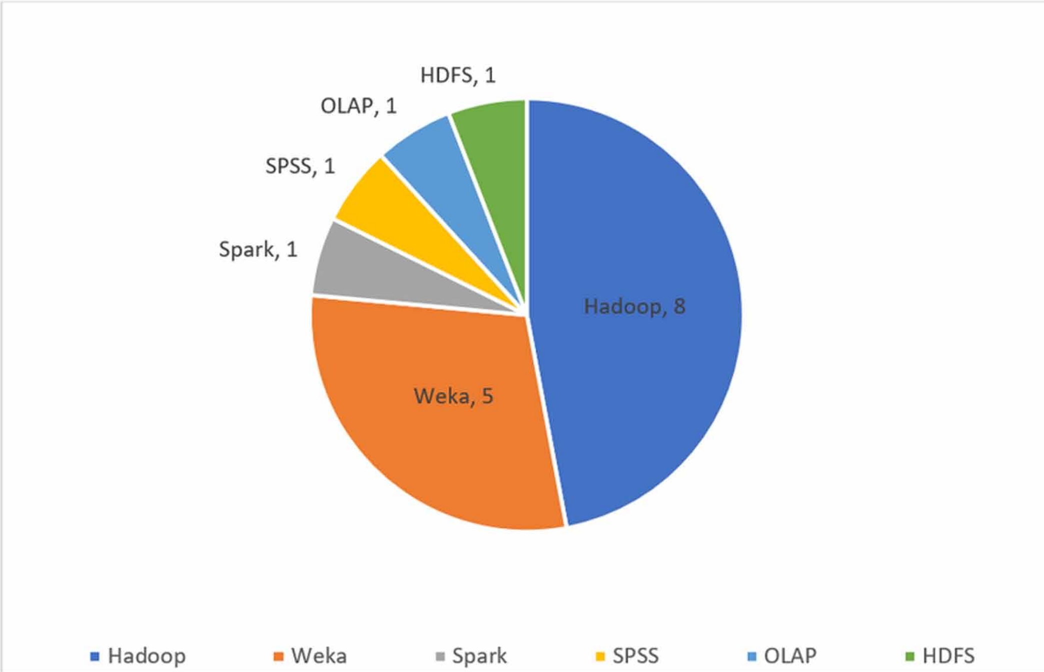
This rapidly developing research field requires a regular survey to measure the maturity level of the field. This has motivated the authors to write this paper. A systematic mapping study was conducted of 84 studies from 2010-2020 on applying BDA in the HES. Three major themes and 12 subthemes around the utilization of BDA in HEIs were identified. The implementation challenges were also mapped for each of the themes, showing that applying BDA to support teaching and learning in HEIs has the most challenges. Moreover, most studies employed predication tools, particularly classification algorithms and regression algorithms to identify patterns in BD, and most studies used supervised learning algorithms. There were 16 frameworks proposed, yet the research around developing a framework for integrating BD into HEIs remains inadequate. Most of the proposed frameworks were not validated, or only one method was used for validation, such as in-depth interviews. The findings of this systematic mapping study are expected to contribute to ongoing research and development to support the application of BDA in HEIs. Further, this study helps practitioners (providers of BDA, information technology, and quality assurance officers in HEIs) to better understand the role of

Table 2. Data mining (DM) algorithms of reviewed articles

Studies	DM Algorithms				Output
	Prediction		Clustering	Association rules	
	Classification	Regression	Outlier analysis		
	Naive Bayes, Random Forest, Decision Tree, Artificial Neural Network), and Support Vector Machine.	Logistic regression, Support Vector Machine (SVM), and Regression trees	K-means	A priori algorithm, Quick a priori based on classic a priori algorithm	
SR26, SR5 SR65, SR6, SR84, SR11	✓	✓			Predict student performance
SR71, SR36	✓	✓			Predict student dropout
SR79, SR78, SR23	✓	✓			Tuition fees, HEI financial status, and grant loan bursaries
SR57	✓				Predict teacher morality levels and teacher ability
SR3	✓				Predict candidate research focus
SR75	✓				HEI admission
SR76	✓				Improvement strategy for competency in Education for Sustainable Development (ESD)
SR55			✓		Student personality characteristics
SR12			✓		Performance evaluation of HES resources utilization
SR83			✓		Student tuition
SR14				✓	Faculty evaluation
SR8				✓	Help in HEI teaching evaluation systems
SR28				✓	Library books management system
SR52		✓	✓		Faculty performance
SR68		✓	✓		Factors affecting HEI retention/ attrition
SR73	✓		✓		Factors affecting academic performance
SR40			✓	✓	Improving HEI e-learning
SR4	✓	✓	✓	✓	Improving HEI management

existing models, frameworks, and tools to develop practical and corresponding solutions to support BDA applications in HEIs. Notwithstanding these contributions, based on the findings, this study recommends that greater emphasis be placed on conducting empirical research studies on HEIs applications of BDA. This is fundamental to building a stronger evidence base around best practice and to findings solutions to the current issues confronting HEIs in their application of BDA to improve operational outcomes._

Figure 7. Most used open-source platforms in the literature



REFERENCES

Abe, K. (2019, August). Data mining and machine learning applications for educational big data in the university. In *2019 IEEE International Conference on Dependable, Autonomic and Secure Computing*, Fukuoka, Japan. doi:10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00071

Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, *24*(2), 1527–1543. doi:10.1007/s10639-018-9839-7

Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University dropout prediction through educational data mining techniques: A systematic review. *Journal of E-Learning and Knowledge Society*, *15*(3), 161–182. doi:10.20368/1971-8829/1135017

Akhmetov, B., Izbassova, N., & Akhmetov, B. (2012, December). *Developing and customizing university business intelligence cloud*. In *2012 International Conference on Cloud Computing Technologies, Applications and Management* (pp. 229-233). doi:10.1109/ICCCTAM.2012.6488104

Al Rashdi, S., & Srinivas, S. (2016, December). Business intelligence maturity in educational institutions-sultan Qaboos university: A case study. In *International Conference on Intellectual Capital and Knowledge Management and Organisational Learning* (pp. 9-17). Academic Conferences International Limited.

Al Rashdi, S. S., & Nair, S. S. K. (2017). A business intelligence framework for Sultan Qaboos University: A case study in the Middle East. *Journal of Intelligence Studies in Business*, *7*(3), 73–95. doi:10.37380/jisib.v7i3.278

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13–49. doi:10.1016/j.tele.2019.01.007

Anshari, M., Alas, Y., Sabtu, N. P. H., & Hamid, M. S. A. (2016). Online learning: Trends issues and challenges in the big data era. *Journal of e-Learning and Knowledge Society, 12*(1), 38-84. 10.2991/saeme-17.2017.83

Apraxine, D., & Stylianou, E. (2017, April). Business intelligence in a higher educational institution: The case of University of Nicosia. In *2017 IEEE Global Engineering Education Conference* (pp. 1735-1746), Athens, Greece: IEEE. doi:10.1109/EDUCON.2017.7943085

Attaran, M., Stark, J., & Stotler, D. (2018). Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation. *Industry and Higher Education*, *32*(3), 169–182. doi:10.1177/0950422218770937

Barakat, M. (2007). *The role of Jordanian universities in achieving comprehensive security* [Paper presentation]. *Conference on The Role of the University Personality in Achieving Comprehensive Security*, Amman, Jordan.

Basfar, H., Amer, T., & Amer, R. (2011). *Professional development for faculty members in university education*. Amedical Institution for Publication and Distribution.

Bermudez, R. S., Manalang, J. O., Gerardo, B. D., & Tanguilig, B. T. (2011, August). Predicting faculty performance using regression model in data mining. In *2011 Ninth International Conference on Software Engineering Research, Management and Applications* (pp. 68-72), Baltimore, MD: IEEE. doi:10.1109/SERA.2011.29

Boulila, W., Al-Kmali, M., Farid, M., & Mugahed, H. (2018). A business intelligence-based solution to support academic affairs: Case of Taibah University. *Wireless Networks*, 1–8. doi:10.1007/s11276-018-1880-3

Calvo, P., Arroyo, J. C. T., & Delima, A. J. P. (2020). Higher education institution (HEI) enrollment forecasting using data mining technique. *International Journal in Advanced Trends in Computer Science and Engineering, 9*(2), 2060-2064.10.30534/ijatcse/2020/179922020

Cao, J., Li, Z., Luo, Q., Hao, Q., & Jiang, T. (2018, August). Research on the construction of smart university campus based on big data and cloud computing. In *2018 International Conference on Engineering Simulation and Intelligent Control* (pp. 351-353), Changsha, China: IEEE. doi:10.1109/ESAIC.2018.00088

Chen, S. (2019). Reform and practice of big data driven higher education management and decision. *Journal of Physics: Conference Series*, *1345*(2), 1–5. doi:10.1088/1742-6596/1345/2/022020

Cho, R., & Kim, D. (2019, June). The factors affecting university retention/attrition by big data analytics. In *HEAD'19, 5th International Conference on Higher Education Advances* (pp. 991-1001), Valencia, Spain: Editorial Universitat Politècnica de València.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage Publications.

Daniel, B., & Butson, R. (2014, September). Foundations of big data and analytics in higher education. In *International Conference on Analytics Driven Solutions: ICAS2014* (pp. 39-47). Ottawa, Canada: Academic Press.

Dong, J. (2018, May). Analysis and designing of educational big data mining system for higher education. In *2018 8th International Conference on Social science and Education Research*. Xi'an, China: Atlantis Press. doi:10.2991/sser-18.2018.39

Dyng Anh, T. V., & Shi, W. (2017, July). Trend of higher education abroad policy analysis based on regression analysis and data mining. In *2017 International Conference on Sports, Arts, Education and Management Engineering*. Shenyang, China: Atlantis Press.

Elhassan, I., & Klett, F. (2015, December). Bridging higher education and market dynamics in a business intelligence framework. In *2015 International Conference on Developments of E-Systems Engineering* (pp. 198-203). Dubai, UAE: IEEE. doi:10.1109/DeSE.2015.22

Enaro, A. O., & Chakraborty, S. (2019, December). A review on the academic performance analysis of higher education students using data mining techniques. In *2019 International Conference on Contemporary Computing and Informatics* (pp. 256-260). Singapore: IEEE. doi:10.1109/IC3I46837.2019.9055632

Falakmasir, M. H., Moaven, S., Abolhassani, H., & Habibi, J. (2010, June). Business intelligence in e-learning: Case study on the Iran university of science and technology dataset. In *The 2nd International Conference on Software Engineering and Data Mining* (pp. 473-477). Chengdu, China: IEEE.

Fan, C., & Han, S. (2016, July). Importance of central task in universities under the big data environment: Taking research synthesis of Hong Kong as an example. In *2016 2nd International Conference on Social Science and Higher Education*. Sanya, China: Atlantis Press. doi:10.2991/icsshe-16.2016.44

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Gartner, I., & Beyer, M. (2011). Gartner says solving 'Big Data' challenge involves more than just managing volumes of data. Gartner Special Report Examines How to Leverage Pattern-Based Strategy to Gain Value in Big Data, Gartner.

Guadamuz, A., & Cabell, D. (2014). Data mining in UK higher education institutions: Law and policy. *Queen Mary Journal of Intellectual Property*, *4*(1), 3–29. doi:10.4337/qmjip.2014.01.01

Hamed, M., Mahmoud, T., Gómez, J. M., & Kfouri, G. (2017). Using data mining and business intelligence to develop decision support systems in Arabic higher education institutions. In *Modernizing Academic Teaching and Research in Business and Economics* (pp. 71–84). Springer. doi:10.1007/978-3-319-54419-9_4

Hanandeh, F., Al-Shannag, M. Y., & Alkhaffaf, M. M. (2016). Using data mining techniques with open source software to evaluate the various factors affecting academic performance: A case study of students in the faculty of information technology. *International Journal of Open Source Software and Processes*, *7*(2), 72–92. doi:10.4018/IJOSSP.2016040104

Hasan, N. A., Miskon, S., Ahmad, N., Ali, N. M., Hashim, H., Syed, N., Abdullah, R. A. A., & Maarof, M. A. (2016). Business intelligence readiness factors for higher education institution. *Journal of Theoretical and Applied Information Technology*, *89*(1), 174–189. https://www.researchgate.net/profile/Nooradilla_Abu_Hasan2/publication/306208607_Business_intelligence_readiness_factors_for_higher_education_institution/links/5965b290aca27227d78c6b69/Business-intelligence-readiness-factors-for-higher-education-institution.pdf

Hegde, V., & Prageeth, P. P. (2018, January). Higher education student dropout prediction and analysis through educational data mining. In *2018 2nd International Conference on Inventive Systems and Control* (pp. 694-699). Pichanur, India: IEEE. doi:10.1109/ICISC.2018.8398887

Huan, J., & Bo, R. (2018). Application of the big data grey relational decision-making algorithm to the evaluation of resource utilization in higher education. *International Journal of Enterprise Information Systems*, *14*(2), 43–55. doi:10.4018/IJEIS.2018040103

Ida, M. (2011, July). XBRL extension for knowledge discovery in higher education. In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 4, pp. 2120-2123), Shanghai, China: IEEE. doi:10.1109/FSKD.2011.6020064

Ijab, M. T., Wahab, S. M. A., Salleh, M. A. M., & Bakar, A. A. (2019, December). Investigating big data analytics readiness in higher education using the technology-organisation-environment (TOE) framework. In *2019 6th International Conference on Research and Innovation in Information Systems* (pp. 1-7). Johor Bahru, Malaysia: IEEE.

International Educational Data Mining Society. (2011). Retrieved from https://www.educationaldatamining.org/

Jamiu, S. M., Abdullah, N. S., Miskon, S., & Ali, N. M. (2019, September). Data governance support for business intelligence in higher education: A systematic literature review. In *4th International Conference of Reliable Information and Communication Technology* (pp. 35-44). Johor-Malaysia, Malaysia: Springer.

Jha, S., Jha, M., & O'Brien, L. (2018, December). A step towards big data architecture for higher education analytics. In *2018 5th Asia-Pacific World Congress on Computer Science and Engineering* (pp. 178-183). Momi Bay, Fiji: IEEE. doi:10.1109/APWConCSE.2018.00036

Jun, W. (2016, November). Study on process management innovation of scientific projects in universities in context of big data. In *Proceedings of the13th International Conference on Innovation and Management* (pp. 1247-1250). https://d1wqtxts1xzle7.cloudfront.net/

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *46th Hawaii International Conference on System Sciences* (pp. 995-1004). doi:10.1109/HICSS.2013.645

Knauf, R., Sakurai, Y., Tsuruta, S., & Takada, K. (2010, October). Empirical evaluation of a data mining method for success chance estimation of university curricula. In *2010 IEEE International Conference on Systems, Man and Cybernetics* (pp. 1127-1133). Istanbul, Turkey: IEEE. doi:10.1109/ICSMC.2010.5642356

Kumaran, S. R., Othman, M. S., & Yusuf, L. M. (2015, October). Applying theory of constraints (TOC) in business intelligence of higher education: A case study of postgraduates by research program. In *2015 International Conference on Science in Information Technology* (pp. 147-151). Yogyakarta, Indonesia: IEEE. doi:10.1109/ICSITech.2015.7407794

Kun, W. (2019, June). Design of big data intelligent sharing system for the creative achievement of college students in internet + age. In *2019 International Conference on Robots & Intelligent System* (pp. 352-355). Haikou, China: IEEE.

Lara, J. A., De Sojo, A. A., Aljawarneh, S., Schumaker, R. P., & Al-Shargabi, B. (2020). Developing big data projects in open university engineering courses: Lessons learned. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 22988–23001. doi:10.1109/ACCESS.2020.2968969

Li, M., & Li, Y. (2019, October). Analysis and Implementation of key technologies for teaching quality evaluation system in universities based on data mining. In *2019 12th International Conference on Intelligent Computation Technology and Automation* (pp. 456-460). Xiangtan, China: IEEE. doi:10.1109/ICICTA49267.2019.00104

Li, S., & Ni, J. (2015, November). Evolution of big-data-enhanced higher education systems. In *2015 8th International Conference on Internet Computing for Science and Engineering* (pp. 253-258). Harbin, China: IEEE. doi:10.1109/ICICSE.2015.53

Liao, Z., Yin, Q., Huang, Y., & Sc Sheng, L. (2015). Management and application of mobile big data. *International Journal of Embedded Systems*, *7*(1), 63–70. doi:10.1504/IJES.2015.066143

Liu, W. F., & Chang, W. D. (2016, August). Research on how to improve the quality of education in colleges and universities under the background of big data. In *Proceedings of the 2016 International Conference on Management Science and Management Innovation*. Guilin, China: Atlantis Press. doi:10.2991/msmi-16.2016.94

Liu, X. (2017). The application of data mining technology in the teaching evaluation in colleges and universities. *Journal of Computational and Theoretical Nanoscience*, *14*(1), 7–12. doi:10.1166/jctn.2017.6115

Liu, Y. (2018). Data mining of university library management based on improved collaborative filtering association rules algorithm. *Wireless Personal Communications*, *102*(4), 3781–3790. doi:10.1007/s11277-018-5409-y

Liu, Y. (2019). C4. 5 crossover algorithm based on mapreduce in university big data analysis. *Journal of Physics: Conference Series*, *1176*(3), 1–8.

Ma, Y. (2017, March). Construction of college personnel files management informatization under the background of big data. In *Proceedings of the 2017 5th International Conference on Machinery, Materials and Computing Technology*. Beijing, China: Atlantis Press. doi:10.2991/icmmct-17.2017.256

Matsebula, F., & Mnkandla, E. (2016, November). Information systems innovation adoption in higher education: Big data and analytics. In *2016 International Conference on Advances in Computing and Communication Engineering* (pp. 326-329), Durban, South Africa: IEEE. doi:10.1109/ICACCE.2016.8073769

Matsebula, F., & Mnkandla, E. (2017, September). A big data architecture for learning analytics in higher education. In *2017 IEEE AFRICON* (pp. 951-956). Cape Town, South Africa: IEEE.

McAlister, A., Lee, D. M., Ehlert, K. M., Kajfez, R. L., Faber, C. J., & Kennedy, M. S. (2017). Qualitative coding: An approach to assess inter-rater reliability. American Society for Engineering Education, 1-9.

Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 55462–55470. doi:10.1109/ACCESS.2020.2981905

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336–341. doi:10.1016/j.ijsu.2010.02.007 PMID:20171303

Moreira, F., Ferreira, M. J., & Cardoso, A. (2017, July). Higher education disruption through IoT and big data: A conceptual approach. In *4th International Conference on Learning and Collaboration Technologies* (pp. 389-405). Vancouver, Canada: Springer. doi:10.1007/978-3-319-58509-3_31

Musa, S., Ali, N. B. M., Miskon, S. B., & Giro, M. A. (2018, July). Success factors for business intelligence systems implementation in higher education institutions–a review. In *International Conference of Reliable Information and Communication Technology* (pp. 322-330). Kuala Lumpur, Malaysia: Springer.

Namor, N. (2012). *The efficiency of faculty members and their impact on the quality of higher education: A case study of the faculty of economic sciences* [Master's Thesis]. Mentori University.

Natek, S., & Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications*, *41*(14), 6400–6407. doi:10.1016/j.eswa.2014.04.024

Ni, L. (2019, January). *Reform and innovation of college English writing teaching under the background of big data*. In *International Conference on Arts, Management, Education, and Innovation*, Seoul, South Korea.

Nobre, T., Trigo, A., & Sanches, P. (2014, June). SBIAES—business intelligence system for analysis of access to higher education: The case of the Polytechnic Institute of Coimbra. In *2014 9th Iberian Conference on Information Systems and Technologies* (pp. 1-6). Barcelona, Spain: IEEE.

Nukoolkit, C., Chansripiboon, P., & Sopitsirikul, S. (2011, August). Improving university e-learning with exploratory data analysis and web log mining. In *2011 6th International Conference on Computer Science & Education* (pp. 176-179). Singapore: IEEE. doi:10.1109/ICCSE.2011.6028611

Ong, V. K. (2015, July). Big data and its research implications for higher education: Cases from UK higher education institutions. In *2015 IIAI 4th International Congress on Advanced Applied Informatics* (pp. 487-491). Okayama, Japan: IEEE. doi:10.1109/IIAI-AAI.2015.178

Oprea, C., Popescu, D. M., Petrescu, A. G., & Barbu, I. (2017). A data mining based model to improve university management. *Journal of Science and Arts*, *17*(2), 285–313. https://www.icstm.ro/DOCS/josa/josa_2017_2/a_09_Oprea.pdf

Osuwa, A. A., Katende, J. O., & Osuwa, A. A. (2019, July). Perception of Smart Campus big data analytics in university. In *2019 Third World Conference on Smart Trends in Systems Security and Sustainablity* (pp. 48-58). London, UK: IEEE. doi:10.1109/WorldS4.2019.8903933

Otoo-Arthur, D., & Van Zyl, T. (2019, August). A systematic review on big data analytics frameworks for higher education-tools and algorithms. In *Proceedings of the 2019 2nd International Conference on E-Business, Information Management and Computer Science* (pp. 1-9). doi:10.1145/3377817.3377836

Otoo-Arthur, D., & Van Zyl, T. (2019). Analytical visualization of higher education institutions' big data for decision making. *Science*, *15*, 1–9.

Pérez-Pérez, Y. M., Rosado-Gómez, A. A., & Puentes-Velásquez, A. M. (2018, September). Application of business intelligence in the quality management of higher education institutions. In *International Meeting on Applied Sciences and Engineering* (pp. 1-8). San José de Cúcuta, Colombia: IOP Publishing. doi:10.1088/1742-6596/1126/1/012053

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20. doi:10.24059/olj.v16i3.267

Ragab, A. H. M., Noaman, A. Y., Al-Ghamdi, A. S., & Madbouly, A. I. (2014, June). A comparative analysis of classification algorithms for student's college enrollment approval using data mining. In *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments* (pp. 106-113). New York, NY: Association for Computing Machinery. doi:10.1145/2643604.2643631

Rajni, J., & Malaya, D. (2015). Predictive analytics in a higher education context. *IT Professional*, *17*(4), 24–33. doi:10.1109/MITP.2015.68

Renaud, J., Britton, S., Wang, D., & Ogihara, M. (2015). Mining library and university data to understand library use patterns. *The Electronic Library*, *33*(3), 23–39. doi:10.1108/EL-07-2013-0136

Ribeiro, C. E., & Ribeiro, A. I. J. T. (2015). Methodology for generation of business intelligence platforms to university evaluation committees. In *Proceedings of the Annual Conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective-Volume 1* (pp. 187-194). Brazilian Computer Society.

Rodzi, N. A. H. M., Othman, M. S., & Yusuf, L. M. (2015, October). Significance of data integration and ETL in business intelligence framework for higher education. In *2015 International Conference on Science in Information Technology* (pp. 181-186). Yogyakarta, Indonesia: IEEE. doi:10.1109/ICSITech.2015.7407800

Saa, A. A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technology*. *Knowledge and Learning*, *24*(4), 567–598. doi:10.1007/s10758-019-09408-7

Saad, N. (2012). Sources of financing higher education in the Kingdom of Saudi Arabia and Britain: A comparative study. *Arab Studies in Education and Psychology*, *24*, 263–287.

Salaki, R. J., & Mogea, T. (2020). Agile analytics: Adoption framework for business intelligence in higher education. *Journal of Theoretical and Applied Information Technology*, *98*(7), 1032–1042.

Santi, R. P., & Putra, H. (2018, October). A systematic literature review of business intelligence technology, contribution and application for higher education. In *2018 International Conference on Information Technology Systems and Innovation* (pp. 404-409). Padang, Indonesia: IEEE. doi:10.1109/ICITSI.2018.8696019

Segooa, M. A., & Kalema, B. M. (2018, March). Leveraging big data analytics to improve decision making in South African public universities. In *2018 IEEE 3rd International Conference on Big Data Analysis* (pp. 9-13). Shanghai, China: IEEE. doi:10.1109/ICBDA.2018.8367462

Senthil Kumar, J., Venkataraman, V., Meganathan, S., & Meena, V. (2017). A data mining approach to classify higher education sector data using Bayesian classifier. *International Journal of Mechanical Engineering and Technology*, *8*(9), 95–103.

Shah, T., Rabhi, F., & Ray, P. (2015). Investigating an ontology-based approach for big data analysis of inter-dependent medical and oral health conditions. *Cluster Computing*, *76*(1), 351–367. doi:10.1007/s10586-014-0406-8

Shamsuddin, N. T., Aziz, N. I. A., Cob, Z. C., Ab Ghani, N. L., & Drus, S. M. (2018, November). Big data analytics framework for smart universities implementations. In *International Symposium of Information and Internet Technology* (pp. 53-62). Bandung, Indonesia: Springer, Cham.

Shaya, A., Amer, T., & Amer, R. (2012). *Higher education and future challenges*. Dar Al-Zahra.

Shuqfa, Z., & Harous, S. (2019, November). Data mining techniques used in predicting student retention in higher education: A survey. In *2019 International Conference on Electrical and Computing Technologies and Applications* (pp. 1-4). Aurak, UAE: IEEE. doi:10.1109/ICECTA48151.2019.8959789

Sihui, D., & Xueguo, X. (2016, June). Research on tacit knowledge mining of university libraries based on data mining. In *2016 13th International Conference on Service Systems and Service Management* (pp. 1-4). Kunming, China: IEEE.

Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytic—literature review. *ICTACT Journal on Soft Computing, 5*(4), 1035-1049. http://ictactjournals.in/paper/IJSC_V5_I4_paper6_1035_1049.pdf

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. doi:10.1016/j.jbusres.2016.08.001

Skalka, J., Drlík, M., & Švec, P. (2013, March). Knowledge discovery from university information systems for purposes of quality assurance implementation. In *2013 IEEE Global Engineering Education Conference* (pp. 591-596). Berlin, Germany: IEEE. doi:10.1109/EduCon.2013.6530165

Stefanova, K., & Kabakchieva, D. (2017, June). Educational data mining perspectives within university big data environment. In *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 264-270). Madeira Island, Portugal IEEE. doi:10.1109/ICE.2017.8279898

Tang, L., Zhou, C., He, L., & Zhang, S. (2017, June). Research on evaluation of morality and ability of teachers in universities based on the perspective of data mining. In *2017 International Conference on Management, Education and Social Science*. Qingdao, China: Atlantis Press. doi:10.2991/icmess-17.2017.41

Tastimur, C., Karakose, M., & Akin, E. (2016, September). Improvement of relative accreditation methods based on data mining and artificial intelligence for higher education. In *2016 15th International Conference on Information Technology Based Higher Education and Training* (pp. 1-7). Istanbul, Turkey: IEEE. doi:10.1109/ITHET.2016.7760736

Trakunphutthirak, R., Cheung, Y., & Lee, V. C. (2019, January 27). A study of educational data mining: Evidence from a Thai university. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 734–741. doi:10.1609/aaai.v33i01.3301734

Tulasi, B. (2013). Significance of big data and analytics in higher education. *International Journal of Computers and Applications*, *68*(14), 66–79. doi:10.5120/11648-7142

Turban, E., Sharda, R., & Delen, D. (2010). *Decision support and business intelligence systems* (10th ed.). Pearson.

Utami, N. W., Sukajaya, I. N., Candiasa, I. M., & Dewi, E. G. A. (2019, October). The implementation of data mining to show UKT (students' tuition) using fuzzy c-means algorithm: (case study: Universitas Pendidikan Ganesha). In *2019 International Conference on Advanced Computer Science and information Systems* (pp. 101-106). Nusa Dua, Indonesia: IEEE. doi:10.1109/ICACSIS47736.2019.8979933

Utomo, M. N. Y., Permanasari, A. E., Tungadi, E., & Syamsuddin, I. (2017, November). Determining single tuition fee of higher education in Indonesia: A comparative analysis of data mining classification algorithms. In *2017 4th International Conference on New Media Studies* (pp. 113-117). Yogyakarta, Indonesia: IEEE. doi:10.1109/CONMEDIA.2017.8266041

Villegas-Ch, W., Palacios-Pacheco, X., & Luján-Mora, S. (2019). Application of a smart city model to a traditional university campus with a big data architecture: A sustainable smart campus. *Sustainability*, *11*(10), 2857–2889. doi:10.3390/su11102857

Villegas-Ch, W., Palacios-Pacheco, X., Ortiz-Garcés, I., & Luján-Mora, S. (2019). Management of educative data in university students with the use of big data techniques. *Sustainability*, *11*(19), 145–234. http://hdl.handle.net/10045/94687

Viloria, A., Naveda, A. S., Palma, H. H., Núñez, W. N., & Núñez, L. N. (2019, October). Using big data to determine potential dropouts in higher education. In *First International Conference on Emerging Electrical Energy, Electronics and Computing Technologies 2019*. Melaka, Malaysia: IOP Publishing.

Wang, D., & Shao, J. (2018). Application of data mining in analysis of college students' academic performance. *Journal of Advanced Oxidation Technologies*, *21*(2), 45–53.

Wang, X. (2015, July). Information service in the big data era and development strategies for university libraries. In *2nd Annual International Conference on Social Science and Contemporary Humanity Development*. Wuhan, Hubei, China: Atlantis Press.

Wang, X. (2019, April). Construction of financial management information system model in colleges and universities based on data mining algorithm. In *2019 11th International Conference on Measuring Technology and Mechatronics Automation* (pp. 623-627). Qiqihar, China: IEEE. doi:10.1109/ICMTMA.2019.00143

Wen, S., Rong, T., Jian, W., Hongmei, L., & Jianye, L. (2014, April). Applying data mining to improve faculty evaluation for educational organizations. In *Proceedings of the 2014 International Conference on Engineering Technology and Applications* (pp. 183-190). Tsingtao, China: CRC Press.

Weng, S. S., Liu, Y., Dai, J., & Chuang, Y. C. (2020). A novel improvement strategy of competency for education for sustainable development (ESD) of university teachers based on data mining. *Sustainability*, *12*(7), 2679–2704. doi:10.3390/su12072679

Williamson, B. (2018). The hidden architecture of higher education: Building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education*, *15*(1), 1–26. doi:10.1186/s41239-018-0094-1

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, J. (2020). Design of university evaluation system based on big data. *Journal of Physics: Conference Series*, *1437*(1), 1–5. doi:10.1088/1742-6596/1437/1/012065

Yan, W. (2011). Application research of data mining technology about teaching quality assessment in colleges and universities. *Procedia Engineering*, *15*, 4241–4245. doi:10.1016/j.proeng.2011.08.795

Yang, C. Y., Liu, J. Y., & Huang, S. (2020). Research on early warning system of college students'behavior based on big data environment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*(W10), 659–665. doi:10.5194/isprs-archives-XLII-3-W10-659-2020

Yang, L. (2018, March). Research on optimization of university financial informational construction in the era of big data taking x university as an example. In *2nd International Conference on Culture*, *Education and Economic Development of Modern Society*. Moscow: Atlantis Press. doi:10.2991/iccese-18.2018.233

Yang, P., & Yang, M. (2019, July). Research on the management model of university students academic early warning based on big data analysis. In *2019 International Conference on Communications, Information System and Computer Engineering* (pp. 639-642). Haikou, China: IEEE. doi:10.1109/CISCE.2019.00148

Yi, T. (2010, July). Apply data mining to predict teachers' position settlement: A case study of university. In *2010 3rd International Conference on Computer Science and Information Technology*. Chengdu, China: IEEE.

Yulianto, A. A., & Kasahara, Y. (2018, July). Implementation of business intelligence with improved data-driven decision-making approach. In *2018 7th International Congress on Advanced Applied Informatics* (pp. 966-967). Yottori, Japan: IEEE. doi:10.1109/IIAI-AAI.2018.00204

Zhang, G., Li, J., & Hao, L. (2015). Research on cloud computing and its application in big data processing of distance higher education. *International Journal of Emerging Technologies in Learning*, *10*(8), 142–189. doi:10.3991/ijet.v10i8.5280

Zhao, B., & Fu, L. (2017, September). Reform of teaching mode in universities based on big data. In *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 453-458). Changsha, China: Springer. doi:10.1007/978-981-10-6388-6_38

Adel Alkhalil has received his PhD degree from Bournemouth University, UK under supervision of Dr. Reza Sahandi and Dr. David John. In September 2016, Dr. Alkhalil joined the faculty of computer science and engineering at the University of Hail, Saudi Arabia as an assistant professor. His research interests include software evolution for mobile and cloud computing systems, decision support systems, knowledge-based systems, and big data analytics.

Magedy Abd Elrahman has a PhD in Educational leadership & management. His research interests focus on strategic planning in higher education, Leadership in Educational Management, and Tertiary Education KPIs.

Azizah Alogali, a PhD holder in Educational Leadership and masters graduate in Education Policy, Higher Education Administration and Program Evaluation. She has worked as a professor in Saudi university and also as a Saudi government consultant in quality assurance and Organizational development.

Abdulaziz Aljaloud has a PhD in Computer Science, specialized in Data Science. his research fields include Artificial Intelligence, Data Science, Machine learning, and their applications to education & learning. he has over 3 years of experience in higher education & quality assurance.