



A Comprehensive Review on Various Data Science Technologies Used for Enhancing the Quality of Education Systems

Olfat M. Mirza^(✉)

Department of Computer Science, College of Computers and Information Systems,
Umm Al-Qura University, Makkah, Saudi Arabia
ommirza@uqu.edu.sa

Abstract. Education is one of the major sources for determining the growth of country with high economic development. But the challenges facing by the education systems are poor decision-making ability, high difficulties in adapting new curriculums, inefficient teaching, and training. These factors could inherently affect the performance of education sectors in terms of increased unemployment, reduced workforce, and dissatisfaction outcomes. In order to solve these problems, this research work aims to deploy the data science technologies for improving the learning strategies in education systems. Here, the data mining techniques are mainly used to extract the relevant or useful information from the data and is widely used for solving the higher education problems. Also, this work investigates some of the challenges associated to the deployment of big data in education systems, which includes consequentialism, scientism, privacy, and security. Moreover, operating characteristics and features of the cyber security model are assessed and validated in Sect. 5. Finally, the overall paper is summarized with its obtainment and future work.

Keywords: Data Science Technologies · Education System · Data Mining · Data Warehouse · Big Data · Business Intelligence · Academic Skills

1 Introduction

In the present days, education [1] is considered as one of the key element used for improving the progression rate of country with the profitable financial and social incomes. But, the poor decision making [2, 3] could degrade the performance and efficiency of the entire system with reduced accuracy levels. Hence, it is more essential to implement the data science technologies for enhancing the overall growth and efficiency of higher education systems. For this purpose, there are various data science methods have been used in the conventional works, which supports to accomplish an efficient decision making by providing the suitable solutions for the given problems. Moreover, the deployment of data science technologies such as data mining, data warehouse, business intelligence, and big data are extensively used in many application systems [4, 5]. But, the use of

these techniques in an education system could increase the accuracy of decision making with simple decisions for solving the complex problems. The sample data science technologies [6–8] used for enhancing the higher education sector is illustrated in Fig. 1.

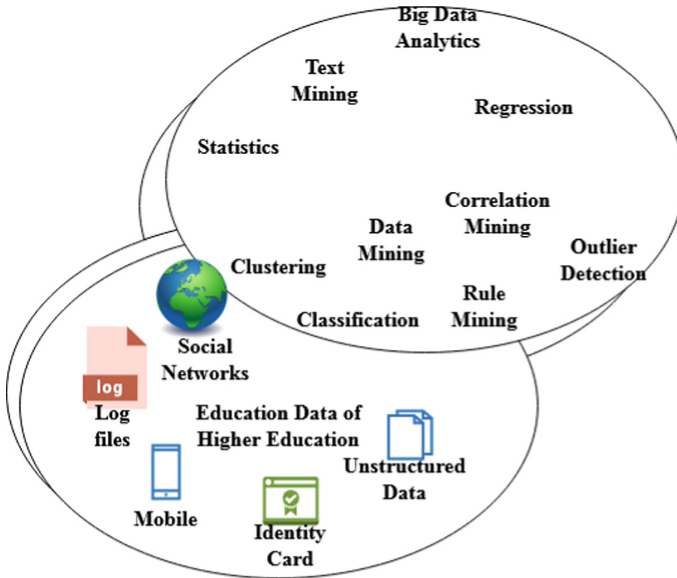


Fig. 1. Data science technologies used in higher education system.

The key objectives behind this research work are as follows:

- To investigate the working operations and features of various data science technologies used in an education system.
- To study how the data science methodologies could support improving the learning strategy of education sectors.
- To analyze the efficiency and benefits of using data science methodologies such as data mining, data warehouse, business intelligence, and big data.

The rest of sections structuralized in this paper are as follows: Sect. 2 discusses about the conventional works developed with the data science technologies for education systems. Section 3 presents the clear description of each data science methodology used for improving the learning strategy of higher education institutions. Then, Sect. 4 presents the results and discussion of the existing methodologies used in an education system. Finally, the overall paper is concluded with its obtainment and future scope in Sect. 5.

2 Related Works

This section some of the existing works related to the deployment of data science technologies used in an education system. Also, it discusses about the purpose of various methodologies helps to improve the learning and monitoring skills.

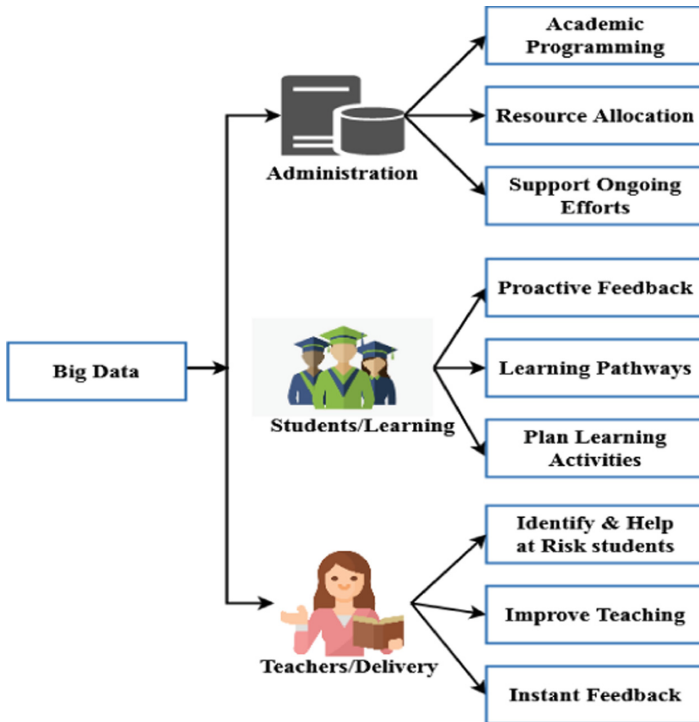


Fig. 2. Use of Big data in education.

Daniel [9] presented a clear review on big data methodologies used for improving the education system, and this work intends to create the awareness on various problems facing by the educational researchers. Also, the reconceptualization based big data model has been discussed for enhancing both the quality of instructions and research. Moreover, the big data could be more useful for improving the educational system, which is graphically represented in Fig. 2.

Moreover, the different types of issues related to the deployment of big data for education research are as follows: technical, ontology, conception, epistemology, digital divide, methods, data analysis, ethics and privacy. *Chweya, et al.* [10] discussed about the benefits of using big data and IoT in education system for boosting the smart training with quality materials. Typically, the IoT and big data are considered as the developing technologies and plays a vital role in education system. Here, the highlights related to the linking of these two technologies were presented, which helps to improve the training strategies of both learning and teaching. *Aldowah, et al.* [11] conducted a detailed review on Educational Data Mining (EDM) and Learning Analytics (LA) model for improving the quality of higher education system by the following ways:

1. Computer supported behavior analysis
2. Computer supported predictive analysis
3. Computer supported visualization analysis

4. Computer supported learning analysis

Moreover, the data mining techniques could provide the suitable solutions for solving certain problems related to the learning factor, which includes the types of collaborative learning, self-learning, monitoring the student's learning ability, evaluation of materials used for learning, and analysis of social networks. *Moscoso-Zea, et al.* [12] suggested the data warehousing model for providing opportunities to improve the education systems. Also, it aims to investigate the effectiveness of using data warehousing model according to the specifications. *Bhanti, et al.* [13] deployed an e-governance system with the help of data warehousing model for an education system. Here, the main intention of using the data warehousing methodology is to regulate and control the authorities and students in a distributed environment based on the factors of efficiency, accessibility, and transparency. *Tulasi, et al.* [14] recommended the big data analysis model for enhancing the students' higher education system based on the factors of resource management and efficient decision making. The key factor of this work was to improve the success rate of students by finding the risks at the earlier stages. Moreover, the innovative models are mainly deployed in an educational institution for solving the complex issues as well as transforming the education policies.

Ahmad, et al. [15] investigated about the performance of various classification based data mining techniques used for accurately predicting the academic performance of students, which includes the types of Naïve Bayes (NB), Decision Tree (DT), and rule based approaches. Moreover, this framework comprises the stages of data gathering, integration, transformation, pattern extraction, and classification. This work mainly concentrates on extracting the most useful patterns by using the education data mining models, where the hidden information is also extracted from the educational dataset. *Sin and Muthu* [16] utilized the concept of learning management system with big data technologies for monitoring the students' online activities. Typically, there are different types of techniques have been used in the present days for improving the educational systems, which includes the regression models, nearest neighbour, clustering, and other learning-based classification techniques. In addition to that, this work objects to perform the skill estimation and behaviour detection processes by using the predictive learning models. *Hasan, et al.* [17] suggested the business intelligence models for education systems, which covers the broad factors of technology, organization, and social. Here, some intelligent solutions have been provided for obtaining growth in educational system.

From the survey, it is identified that the existing works could highly utilized the big data and dimensionality techniques for enhancing the performance of education systems.

3 Data Science Methodologies

This section investigates some of the recent data science techniques used for improving the efficiency of education system, which includes the technologies of data mining models, big data models, business intelligence and data warehousing techniques. The detailed description of these methodologies is presented in the following subsections:

3.1 Data Mining Techniques used in Education System

Data mining is one of the most suitable methodology used for improving academic intervention system. Generally, the data mining techniques are highly depending on the processes of clustering, classification, valuation, and conception. The data mining techniques are mainly used to extract the relevant or useful information from the data and is widely used for solving the higher education problems. *Bhise, et al.* [18] suggested the educational data mining technology for improving the better understanding of students. Here, the main purpose of using this technology is to extract the useful information from the education dataset for analysing the behaviours and trends of students related to education. *Kumar, et al.* [19] intended to predict the performance of students by using an educational data mining techniques. The key factor of this paper is to accurately predict the progress rate of students and to improve their performance in academics with the help of data mining models. Here, the different types of data mining techniques used for predicting the performance of students are listed in below:

1. Decision Tree (DT).
2. Artificial Neural Network (ANN).
3. Random Forest (RF).
4. Support Vector Machine (SVM).
5. Naïve Bayes (NB).
6. Linear regression.
7. K-Nearest Neighbor (KNN).
8. Random tree.

Also, various attributes that affect the performance of prediction accuracy are investigated in this work, which includes internal attributes, external attributes, family income, gender, GPA, SPM grades, academic information, and internal assessment Fig. 3.

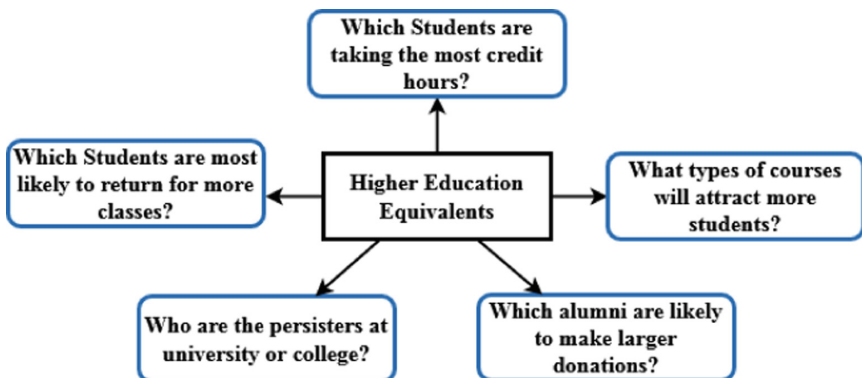


Fig. 3. Use of data mining techniques in higher education.

3.2 Big Data Models

Big data is one of the popular data science technology used in different types of applications such as organizational sectors, educational environment, healthcare systems, and transportations for efficiently processing the large number of data. The key dimensionalities of big data are variety, velocity, volume, complexity and veracity. In paper [20], the big data model has been utilized for enhancing the process of education, where the university education environment is taken for analysis. *Manohar, et al.* [21] utilized the big data analytics technique for predicting the performance of students based on the probability measure. Here, the big data methodology is mainly used to assess the success factor of students with respect to the parameters of past performance, subject choice, demographic, and credit scores. Moreover, it helps to provide the solutions for the problems of data fog state, data cleaning, refining, and lack of data governance. *Johnson, et al.* [22] suggested the big data technique for improving the education strategy of higher institutions by enhancing the better understanding of students. Also, it investigates some of the challenges associated to the deployment of big data in education systems, which includes consequentialism, scientism, privacy, and security. *Li, et al.* [23] presented an adaptive big data based learning methodology for supporting students to cultivate their learning abilities. This framework comprises the modules of prediction, display, adaptive, intervention, content delivery, and learning module. Here, it is stated that the big data is a kind of advanced learning paradigm used for improving the standard of education by capturing the learners’ activities. The major benefits of using the big data technology in education systems are as follows:

- It helps to acquire an efficient data learning and storage.
- It enables the diversification of test types.
- It increases the visualization of knowledge growth.
- It permits timely supervision and personalization.

The typical flow of using big data methodology in an educational system for improving the learning capabilities of students is illustrated in Fig. 4.

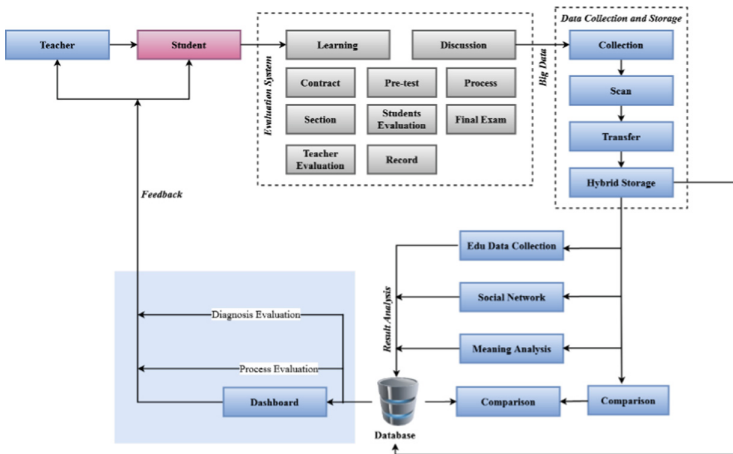


Fig. 4. Flow of analysis using big data technology.

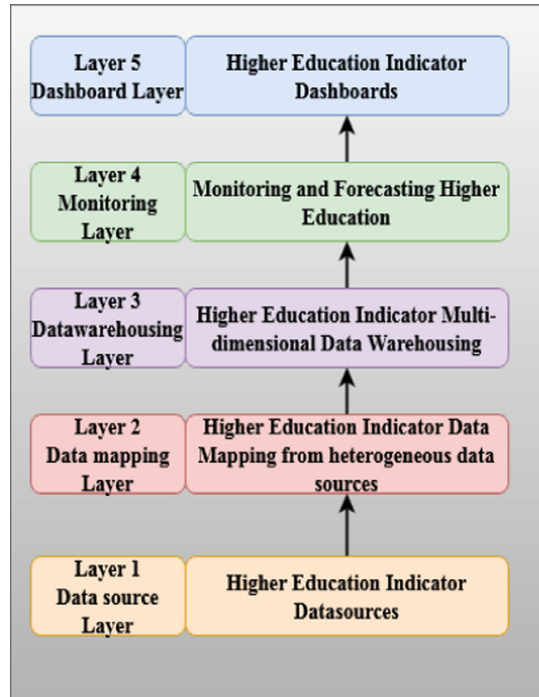


Fig. 5. BI model used in higher education system.

3.3 Data Warehousing Models

Typically, data warehousing is a kind of information model that is mainly used for making the strategic decisions based on business intelligence. Also, it performs the data operations of extraction, cleaning, transportation, and aligning for providing the complete data support to the end users. Taylor, *et al.* [24] utilized a project based data warehousing model for analysing the criticalities in an education system. The main purpose of this paper is to completely provide the data-driven support for enhancing the decision making process with better understanding [25]. Here, the key benefits of using the data warehousing techniques in an educational system are presented in detail, which includes:

- It enhances the process of critical thinking and deep learning.
- It increases the ability of students to solve real-world problems.
- It supports better understanding and growth of technical skills.
- It establishes teamwork activities among students with intelligent ideas.
- It efficiently increases satisfaction, joyfulness, and motivation.

In paper [26], the big data and data warehousing methodologies have been investigated for enabling an efficient data storage and accessing in an education systems. Also, it discussed about the different processing concepts of big data method for analysing the problematic features, which includes map reduce, no SQL, and new SQL.

Santoso, et al. [27] suggested the data warehouse tools for efficiently handling the academic data of universities. The main intention of this paper is to analyse that how the big data framework is incorporated with the data warehouse for enhancing the decision-making process. Based on this study, it is analysed that the big data incorporated with the data warehouse model could efficiently improve the learning ways of education system with better decision making.

3.4 Business Intelligence Models

The business intelligence model has been widely applied in many higher educational institutions for improving an academic decision making. The main purpose of using this technology is, it efficiently supports for the monitoring the analysing the recent trends of education indicators. Also, this technology has been extensively applied in the large-scale industries for providing support to intelligent decision making. *Khatibi, et al.* [28] suggested an advanced business intelligence model for monitoring the higher education pointers by incorporating the internal and external data sources. Figure 5 depicts the layered architecture of business intelligence model used in the higher education system. *Guster, et al.* [29] have stated that the business intelligence model is one of the most suitable option for providing an intelligent solution to the higher education system. In paper [30], the business intelligence models are utilized for increasing the technical efficiency of higher education institutions. The main aim of this work is to enhance both the system proficiency and management operations of the education institutions by efficiently trade the funds and resources. Moreover, various key performance indicators are utilized in this analysis for validating the business models used in the education systems, which includes the types of course work, graduation rate, student enrolment, exam assessment, and academic workloads.

4 Results and Discussion

This section evaluates the performance rate and efficiency of the data science techniques used in the education systems. Also, it examines the advantages and disadvantages of each technology based on its working operations and key characteristics. Figure 6 compares the performance efficiency of various data science technologies such as Business Intelligence (BI), Big Data, Modern Data Warehouse (DWH), Education Data Mining (EDM), and Data Warehouse (DWH). Based on the analysis, it is observed that the big data technology provides an increased efficiency for improving the learning strategy of higher education systems, which is moderately greater than the other models Table 1.

Table 2 and Fig. 7 shows the accuracy level various data mining techniques such as Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), K-Nearest Neighbor (KNN), and Neural Network (NN) used in the education systems. From the results, it is identified that the NB and KNN techniques outperform the other techniques with increased accuracy value.

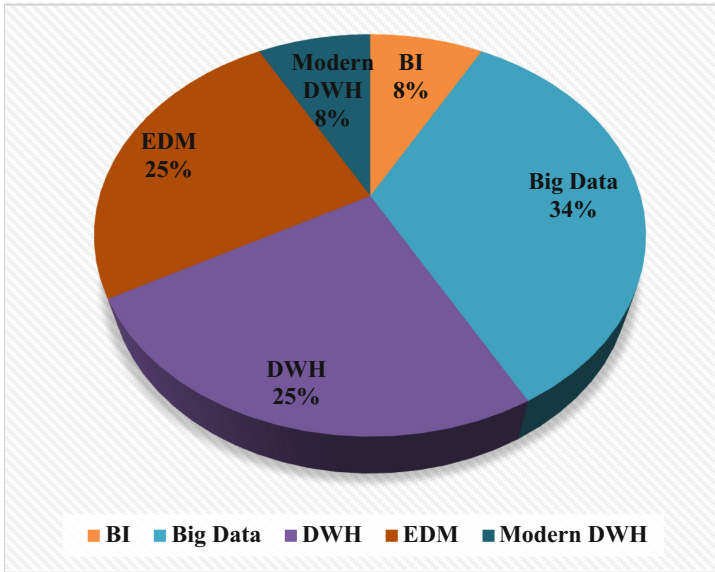


Fig. 6. Recent data science technologies support education sector.

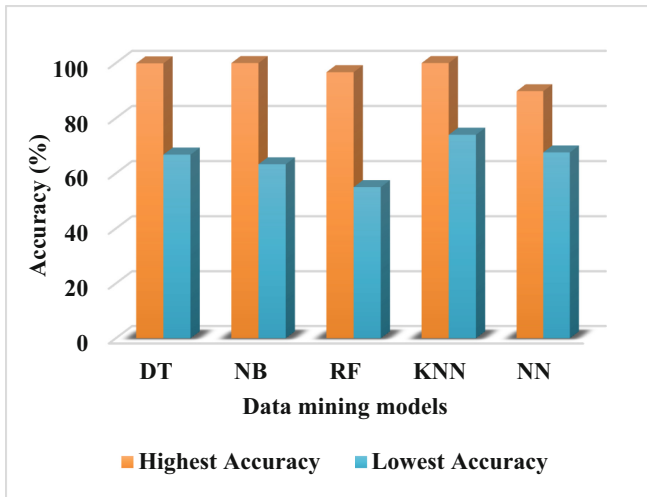


Fig. 7. Accuracy analysis of data mining models.

Table 1. Various data mining techniques used in an education system.

Methods	Description	Applications
Clustering Model	It is mainly used to split the whole data into the group of clusters	It helps to analyse the behaviour of new joining students. Also, it is used to estimate the similarities across various schools
Prediction	It is used to predict the value from the combination of data using the predictor variable	By using the prediction model, the behaviour of students can be analysed (For instance, off task activities, betting the system, and skidding)
Relationship based mining	It helps to discover the relationship among many variables	Based on the relationship mining, the curricular associations have been discovered with improved learning model
Data discovery	It acts as a component used for clustering or classification	The characteristics and behaviour of students have been analysed based on the discovery of data models
Data Distillation	The purpose of using data distillation to identify the suitable features of data	It helps to improvise the student learning process

Table 2. Highest and lowest accuracy of various data mining models.

Methods	Highest Accuracy	Lowest Accuracy
DT	99.9	66.8
NB	100	63.3
RF	96.7	55
KNN	100	74
NN	89.8	67.6

5 Conclusion

Due to the increased demand of data science technologies, which have been extensively used in many application systems such as education, transportation, healthcare and other large industries. Normally, there are various techniques have been used for improving the learning strategy of higher education system. Also, it helps to improve the students' learning ability by deploying an advanced model. In which, the data mining techniques

are highly depends on the processes of clustering, classification, valuation, and conception. The data mining techniques are mainly used to extract the relevant or useful information from the data, and is widely used for solving the higher education problems. Similar to that, the big data model is also one of the suitable for improving the level of adoption of education system. The main purpose of using this technology is, it efficiently supports for monitoring the analyzing the recent trends of education indicators. Also, this technology has been extensively applied in the large-scale industries for providing support to intelligent decision making. During the performance assessment of these techniques, the efficiency of all data science methodologies is validated and compared. Based on these results, it is evident that the proposed KNN and NB models could offer an increased accuracy, when compared to the other techniques.

References

1. Alom, B.M., Courtney, M.: Educational data mining: a case study perspectives from primary to university education in australia. *Int. J. Inform. Technol. Comput. Sci.* **10**(2), 1–9 (2018)
2. Agarwal, R., Dhar, V.: Big data, data science, and analytics: The opportunity and challenge for IS research. 3, *INFORMS*, 2014, pp. 443–448
3. Pandey, U.K., Bhardwaj, B.K.: Data Mining as a Torch Bearer in Education Sector. arXiv preprint [arXiv:1201.5182](https://arxiv.org/abs/1201.5182), 2012
4. Gokalp, M.O., Kayabay, K., Akyol, M.A., Eren, P.E., Koçyiğit, A.: Big data for industry 4.0: A conceptual framework. pp. 431–434
5. Bajpai, N., Biberman, J., Sharma, A.: *Information and Communications Technology in the Education Sector in India* (2019)
6. Sree, G.S., Rupa, C.: Data mining: performance improvement in education sector using classification and clustering algorithm. *Int. J. Innov. Res. Develop.* (ISSN 2278–0211), **2**(7), 101–106 (2013)
7. Dwivedi, S., Roshni, V.K.: Recommender system for big data in education, pp. 1–4
8. Drigas, A.S., Leliopoulos, P.: The use of big data in education. *Int. J. Comput. Sci. Issues (IJCSI)* **11**(5), 58 (2014)
9. Daniel, B.K.: Big Data and data science: a critical review of issues for educational research. *Br. J. Edu. Technol.* **50**(1), 101–113 (2019)
10. Chweya, R., Ajibade, S.S.M., Buba, A.K. and Samuel, M.: IoT and Big Data Technologies: Opportunities and Challenges for Higher Learning. *Int. J. Recent Technol. Eng. (IJRTE)* **9**(2), 909 (2020)
11. Aldowah, H., Al-Samarraie, H., Fauzy, W.M.: Educational data mining and learning analytics for 21st century higher education: a review and synthesis. *Telematics Inform.* **37**, 13–49 (2019)
12. Moscoso-Zea, O., Paredes-Gualtor, J., Luján-Mora, S.: A holistic view of data warehousing in education. *IEEE access* **6**, 64659–64673 (2018)
13. Bhanti, P., Kaushal, U., Pandey, A.: E-governance in higher education: concept and role of data warehousing techniques. *Int. J. Comput. Appl.* **18**(1), 875–895 (2011)
14. Tulasi, B.: Significance of big data and analytics in higher education. *Int. J. Comput. Appl.* **68**(14), 2013
15. Ahmad, F., Ismail, N.H., Aziz, A.A.: The prediction of students' academic performance using classification data mining techniques. *Appl. Math. Sci.* **9**(129), 6415–6426 (2015)
16. Sin, K., Muthu, L.: Application of big data in education data mining and learning analytics--a literature review. *ICTACT J. Soft Comput.* **5**(4), (2015)
17. Hasan, N.A., et al.: Business intelligence readiness factors for higher education institution. *J. Theor. Appl. Inf. Technol.* **89**(1), 174 (2016)

18. Bhise, R.B., Thorat, S.S., Supekar, A.K.: Importance of data mining in higher education system. *IOSR J. Human. Social Sci. (IOSR-JHSS)*, **6**(6) 18–21 (2013)
19. Satapathy, R., Cambria, E., Hussain, A.: Literature Survey. In: *Sentiment Analysis in the Bio-Medical Domain*. SC, vol. 7, pp. 21–38. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68468-0_2
20. Michalik, P., Štofa, J., Zolotova, I.: Concept definition for Big Data architecture in the education system, pp. 331–334
21. Manohar, A., Gupta, P., Priyanka, V., Uddin, M.F.: Utilizing big data analytics to improve education.
22. Johnson, J.A.: The ethics of big data in higher education, the international review of information. *Ethics* **21**, 3–10 (2014)
23. Li, Y., Zhai, X.: Review and prospect of modern education using big data. *Procedia Comput. Sci.* **129**, 341–347 (2018)
24. Taylor, E., Goede, R.: Using critical social heuristics and project-based learning to enhance data warehousing education. *Syst. Pract. Action Res.* **29**(2), 97–128 (2016)
25. Sedkaoui, S., Khelfaoui, M.: Understand, develop and enhance the learning process with big data. *Inform. Disc. Delivery* **47**(1), 2–16 (2018)
26. Ptiček, M., Vrdoljak, B.: Big data and new data warehousing approaches, pp. 6–10
27. Santoso, L.W.: Data warehouse with big data technology for higher education. *Proc. Comput. Sci.* **124**, 93–99 (2017)
28. Khatibi, V., Keramati, A., Shirazi, F.: Deployment of a business intelligence model to evaluate Iranian national higher education. *Social Sci. Human. Open* **2**(1), 100056 (2020)
29. Guster, D., Brown, C.G.: The application of business intelligence to higher education: technical and managerial perspectives. *J. Inform. Technol. Manage.* **23**(2), 42–62 (2012)
30. Alpar, P., Schulz, M.: Self-service business intelligence. *Bus. Inf. Syst. Eng.* **58**(2), 151–155 (2016)