Revue d'Intelligence Artificielle
Vol. 36, No. 4, August, 2022, pp. 589-595
Journal homepage: http://iieta.org/journals/ria

Big Data in Academia: A Proposed Framework for Improving Students Performance

Imran Rashid Banday[1], Majid Zaman[2*], Syed Mohammad Khurshid Quadri[3], Sheikh Amir Fayaz[1], Muheet Ahmed Butt[1]

[1] Department of Computer Sciences, University of Kashmir, J&K 190006, India
[2] Directorate of IT&SS, University of Kashmir, J&K 190006, India
[3] Department of Computer Science, Jamia Millia Islamia, New Delhi 110025, India

Corresponding Author Email: zamanmajid@gmail.com

https://doi.org/10.18280/ria.360411

Received: 16 April 2022
Accepted: 2 August 2022

Keywords:
big data, education, subject recommendation, heterogeneous sources, information technology, Kashmir university

ABSTRACT

The way people learn has radically changed as a result of information technology. As an informal method of learning, fragmented learning has become a popular way to learn new technology and expertise. Academic organizations generate a large amount of heterogeneous data, and academic leaders want to make the most of it by analyzing the large amount of data in order to make better decisions. The volume isn't the only issue; the organization's data structure (structured, semi structured, and unstructured) adds to the complexity of academic work and decision-making on a daily basis. As big data has become more prevalent in educational settings, new data-driven techniques to enhance informed decision-making and efforts to improve educational efficacy have emerged. Traditional data sources and approaches were previously too expensive to obtain with digital traces of student behaviour, which offer more scalable and finer-grained comprehension and support of learning processes. This study provides a fragmented learning solution for students in a data environment that can suggest subjects to them based on their geographical location, gender, and district of residence, among other factors. This suggested framework is expected to play a key role in directing the development of a society that values lifelong learning.

1. INTRODUCTION

In recent years, the IT sector has seen a tremendous growth in the volume of data created, owing mostly to Internet services, prompting a rethinking of the word database. Big Data is a new term for the description and management of massive amounts of structured and unstructured data produced by businesses, organizations, and social media settings [1].

Education systems like Universities operate in an increasingly complicated and competitive environment. To respond to national and global economic, political, and social developments, they must compete with other institutions. Furthermore, various stakeholders want Universities to develop appropriate solutions to these needs in a timely way. To address this issue, Universities must develop the appropriate judgments for coping with these rapid changes by examining large data sources that have been generated [2, 3]. The majority of Universities institutions spend a significant amount of money on information technology in order to construct a data warehouse system. Because of the rapid pace of technological advancements, as well as the emergence of new web-based learning modes results in the information and communication technology revolution. It has increased a competition among higher education institutions, new technological solutions to improve the quality of higher education have emerged from various IT solution providers such as SAP, Cisco, Microsoft, and others, who have provided a parallel universe of IT's [4, 5].

Higher education institutions, such as universities, must deal with massive amounts of data from various sources for accreditation purposes, including data generated from online transactions, videos, audios, images, emails, social media, click streams, logs, posts, search queries, social networking interactions, science data, mobile phone applications, and data stored on multiple operating systems [6]. To meet accreditation criteria, this data is normally kept, sorted, retrieved, and analyzed in a traditional format. Traditional database approaches and tools, on the other hand, are incapable of efficiently processing large amounts of data [7].

The growing number of data repositories in universities necessitates the management of internal data. The data's heterogeneity is owing to the various types of data, which include various streams, courses, and structures. The University of Kashmir has been collecting data and working on data automation and integration since 2002, resulting in a wide range of data repositories and hence Big Data analytics. The heterogeneity, redundancy, and inconsistency in the current set of data are the key reasons for implementing the big data idea at University of Kashmir [8, 9].

This paper is structured as: A brief overview of big data is provided in this section .A generalized review of literature is described in section 2. Levels of Kashmir University education system is defined in which academic data is analyzed and the various course structures are provided in section 3, while in section 4, a manual approach of experimental analysis has been provided. In section 5 many resolutions and issues using Big Data in University system are eloborated. Finally section 5 concludes the paper.

2. LITERATURE BACKGROUND

This section defines the various researches based on the academic data using Big Data.

Logica and Magdalena [10] uses big data in the academic environment in which the research study addresses the progress of Big Data technologies, as well as how they might be applied to e-Learning and their impact on the academic environment. In addition, for a consortium of colleges, authors developed a three-step system architecture based on real-world software solutions with the goal of analyzing, organizing, and accessing large data sets in the Cloud. Furthermore, the study concentrated the research on using the graphical Gephi tool to explore unstructured data.

Zakir et al. [11] analysis the big data in which the authors focuses on the numerous technologies that operate together as a Big Data Analytics system to help estimate future volumes, gather insights, take preemptive measures, and improve strategic decision-making. Furthermore, utilizing a set of data algorithms for huge data sets such as Hadoop and MapReduce, the authors examines the adoption, utilization, and impact of big data analytics on an enterprise's economic value in order to strengthen its competitive edge.

Ahmed [12] proposes a case study on Saudi Universities in order to accredited the Big Data. Using a descriptive study of Saudi institutions, this paper examines the elements required for Big Data deployment and provides insight into how to improve these aspects in higher education accreditation. Security difficulties, safeguarding privacy, technical ability, IT infrastructure, top management backing, and collaborative information-sharing initiatives are identified to be six major considerations for big data application in higher education assessment in Saudi institutions.

Chaurasia et al. [13] proposes a study on dig data analytics in order to generated academic excellence in higher education. The goal of this research is to offer a big data academic and learning analytics enabled business value model to describe the potential benefits and financial value of creating such analytics capabilities in higher education institutions. The study looked at 47 case studies from 26 HEIs to see if there was a link between the BDA's existing and projected advantages and the road to creating financial value for big data academic and learning analytics success in HEIs. The strain of meeting all legal and regulatory obligations, as well as competitiveness, had led HEIs to use BDA tools aggressively. The study discovered, however, that the application of risk and security, as well as predictive analytics, to higher education sectors is still in its early stages. The findings provide academic administrators fresh perspectives on how to develop BDA skills for HEI reform, as well as an empirical framework for a more in-depth examination of BDA execution.

Santoso [14] proposes a study in which authors looks at how big data technologies may be used in conjunction with a data warehouse to aid decision-making. They suggest Hadoop as a large data analytic tool to be used for data input and staging in this system. The report finishes with a discussion of potential directions for the creation and execution of a Big Data institutional initiative.

Dwivedi and Roshni [15] gives a recommender system for the Big Data in the field of Education. This study employs interactive filtering-based ranking approaches to suggest optional courses to students based on their overall grade point average. To produce a set of suggestions, authors have used the Mahout machine-learning library's item-based recommendation on top of Hadoop. The log-likelihood method is used to find patterns between grades and topics. To test the rating system, the Root Mean Square Error between the actual grade and the suggested grade is employed. Schools, colleges, and universities might utilize the findings of this study to offer alternate optional courses to students.

Sin and Muthu [16] proposes a generalized literature survey of Big Data in Education mining in which the top 45-google searches of a Scholarly articles search on "Educational Data Mining" identified two key themes in this study. The publications were mostly from educational data mining journals and conferences. "Introduction to the principles of utilizing Data Mining in Education" and "Prediction or assessment of Student Performance using Data Mining" were recognized as the two significant themes. Because the previous trend included papers on performance prediction, authors deduce that "Performance Prediction Using Data Mining" is the main topic.

Shah [17] proposes a study on Big Data in higher education in which the authors depict that at the executive leadership level of the chosen state institutions, it is necessary to investigate the application of big data business analytics in the decision-making process. Particularly in terms of how descriptive, predictive, prescriptive, decisive, and fundamental analytics, as well as data collecting, affect public university executive leadership decision-making in terms of student retention and success.

There is a lot of research on Big Data in Education, and we have highlighted some of the most significant publications in this study. We believe there is not a single research employing big data that can propose subjects to students based on criteria such as geography, gender, and past records in which a specific student succeeded.

3. LEVELS OF KASHMIR UNIVERSITY EDUCATION SYSTEM: ANALYZING ACADEMIC DATA

The University of Kashmir is one of Jammu and Kashmir's most prestigious educational institutions, offering sophisticated and broad learning as well as a community of scholars and individuals. Okoroma [18], Ramesh et al. [19], Sokout and Usagawa [20], Sahu et al. [21] described university as an advanced institution of learning that guides men and women to a high level of cognitive growth in the Arts, Sciences, Humanities, and Educational topics (2005). The Kashmir University education system has experienced considerable changes in terms of organization, Course structures, curriculum, and now offers undergraduate, graduate, and doctoral degrees.

Currently, there are around 45 postgraduate courses in progress, as well as a PhD programme. Approximately 9000 students from different course selection streams, both genders, and students from different geographic locations (districts) are enrolled in these 45 courses. The student population ranges in age from 21 to 30, with around 45 percent of the female students and the remainder male. Thus, around 9,000 individuals get postgraduate degrees each year, and analyzing this type of data necessitates a large quantity of resources to develop useful patterns from a large amount of data. Heterogeneity, redundancy, and data inconsistency are some of the key challenges we encounter while interpreting data. Because of the vast volume of data, several repositories emerge, giving rise to the Big Data notion.

3.1 Student performance: Analysis of the future

A few indicators that predicted a student's likelihood of success were also identified in research publications [22-25]. These were the following:

Past Success: If a student has a track record of getting good grades, it can be used as a solid predictor of future performance. Multiple study papers and polls have also demonstrated that married students perform better in school than single pupils. The research articles also stated that the older a student is, the greater the chances of achieving a higher GPA [26-29].

Subject Selection: Various studies have found that students who picked math and honors in high school were more likely to succeed in college and graduate degrees than those who chose other topics.

Other Elements: The research identified a number of other factors like learning environment, family issues and instabilities, parental habits and peer relationships that were good predictors of student performance. These included a student's success in online coursework and the ratio of credits attempted to credits finished [30-33].

As a result of these research, we will be more motivated to adopt data analytics at the university level in order to improve individual student performance. In order to test the effectiveness of the framework in the following part, we picked a group of students and used a manual technique to execute these types of approaches**.** Without regard to any specific trait or attribute, we selected a random sample of students.

3.2 Course structures

A student can choose any course to study a subject, and various courses have different subject streams. The University of Kashmir offers a variety of semester and yearlong course options. Each course has its own semesters, and each semester has its own set of subjects. As a result, managing this type of data will become an extremely time-consuming operation. Also, in this study we have presented one of the basic course structures of undergraduate subject (BCA) in which there are different semesters and each semester have different subject codes and subject names as shown in below table (Table 1). The primary criteria for choosing this course are that, it is regarded as a professional course and is strongly advised to the students.

Furthermore, we have provided a brief statistics in which it shows the total number of students enrolled in different streams and different subjects with individual list of boys and girls in each semester (Table 2).

Since this is only a rough estimate for the year 2016, we can see that there are approximately 1 lakh students enrolled in the odd semesters, with approximately 50,000 males and 37,000 females, and we have a data from 2002, handling this much data will be difficult, and we will need to use big data to work efficiently.

Table 1. Brief course structure (BCA stream)

id	Course_Code	Semester	Subject_Code	Display_code	Subject_Name	Credits	batch
11926	BCA	1	BCA16101CC	BCA16101CC	PROGRAMMING IN C/C++	6	2016
11924	BCA	1	BCA16102CC	BCA16102CC	COMPUTER SYSTEM ARCHITECTURE	6	2016
11388	BCA	1	CF116	CF116	COMPUTER FUNDAMENTALS	6	2016
11432	BCA	1	MC116	MC116	MATHEMATICS IN COMPUTING	6	2016
11387	BCA	1	PC116	PC116	PROGRAMMING IN C LANGUAGE	6	2016
11617	BCA	2	PIJ216	PIJ216	PROGRAMMING IN JAVA	6	2016
11616	BCA	2	DS216	DS216	DISCRETE STRUCTURES	6	2016
13263	BCA	3	CNS316	BCA316C7	COMPUTER NETWORKS	6	2016
13264	BCA	3	BCA317G	BCA317G	INTRODUCTION TO PROGRAMMING	6	2016
13262	BCA	3	OS316	BCA316C6	OPERATING SYSTEMS	6	2016
13106	BCA	4	PH417	PHY416C	PHYSICS	6	2016
13111	BCA	4	CPA416	COM416C7	CORPORATE ACCOUNTING COMMERCE	6	2016
13263	BCA	3	CNS316	BCA316C7	COMPUTER NETWORKS	6	2016

Table 2. Semester and gender distribution of students for the year 2016 in University of Kashmir

UG (Under Graduation)			
Semester/Year	Male appeared	female appeared	Total appeared
First semester	13905	13948	27853
3rd semester	14965	1565	30620
3rd year	16523	17072	33611
PG (Post Graduation)			
First Semester	1583	2006	3589
3rd semester	1237	1389	2626
BE (Engineering)			
First Semester	406	189	595
3nd semester	404	165	569
5th semester	458	189	647
7th semester	462	186	648
Total	**49943**	**36709**	**100758**

4. EXPERIMENTAL SETTING: MANUAL APPROACH

We have been experimenting with a manual technique on numerous PG students from varied backgrounds (Districts) with different streams in order to check their performance by recommending the best subjects based on their learning profiles, gender, district, and streams. Based on earlier data acquired from the University of Kashmir, this approach was manually applied. We obtained the academic records of around 2026 students from Kashmir University's examination department. The data's overall structure is displayed in the table below (Table 3) and its graphical representation is shown in below Figure 1.

Table 3. Student academic details before subject recommendation

Total Number of Students		2026	Overall Pass Percentage
Male		1123	62%
Female		903	74%
Professional Courses	BCA	516	64%
	BBA	498	66%
	B.Com Hons	554	68%
	B.Tech	459	71%
District	Srinagar	1304	72%
	Ganderbal	432	66%
	Budgam	290	65%

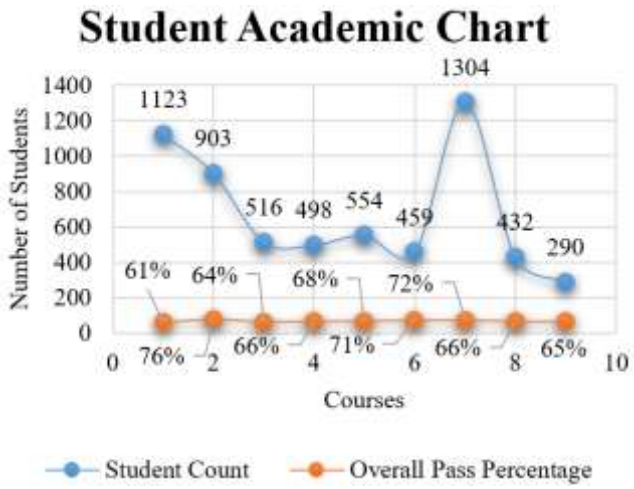

Figure 1. Graphical Representation of student academic details before subject recommendation

We manually analyzed the academic data of some selected streams (BCA, BBA, B.Tech, and B.com Hons) for the year 2016 based on the details provided above, and we saw the pass percentage of students based on three categories: District in which the student resides, Gender, and subject stream that a student has chosen at the under graduate level. As can be seen in the table (Table 3), the district of Srinagar has the highest pass percentage when compared to the other districts. The quantity of amenities Srinagar students receive in comparison to students from distant areas may be the cause of this. The amenities may include access to the internet, transportation options, libraries, etc. We also saw that females performed better than males, with nearly 74 percent of females passing in the year 2015, and that students who chose B.Tech as their subject stream had better results.

Now comes the question of what we can learn from the data. Is our prediction correct for every year? If yes, will it aid in improving the student's performance in each of the three categories? Will this unique idea assist students in receiving accurate subject recommendations? Based on the prior results generated, we recommend that these selected students pursue subjects for PG courses in order to answer these questions. According to the records, a female student from Srinagar district with the subject stream B.Com hons has an overall passing percentage of 81 percent, which is much higher than the overall result in B.com hons independent of these categories. Therefore, if a student belongs to the district of Srinagar and is female, we recommend that she choose M.com hons as her PG subject stream.

4.1 Experimental implementation

We were able to manually generate the results because the number of students in this study were small, and we were able to recommend the subjects to the students based on the results. Our trial yielded positive findings in general. The following is a tree structure for district Srinagar based on the categories that we shaped in this paper (Figure 2).

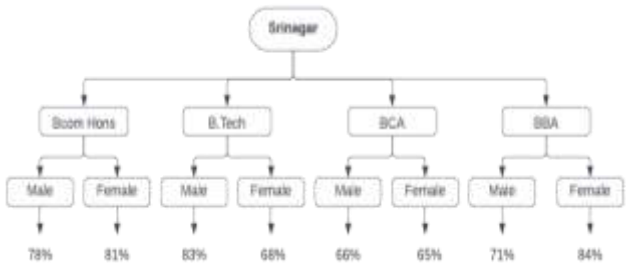

Figure 2. Tree representation for district srinagar with percentage gender distribution of subject streams

Based on past records, district, and gender, we can advise a best subject for a student to choose a professional subject in his or her PG course in the above figure. Assume a student from the district of Srinagar wishes to pursue a PG degree in any of the four streams. So, according to the aforementioned subject pass percentages based on gender and district, if a student is a boy from the district Srinagar, the best option for him is to pursue an M.Tech degree, and so on.

After the subjects were recommended to the students, our results finally proven to be the best, and the results provided some promising outcomes, as seen in the table below (Table 4). Here, the total pass rate for each course rises, demonstrating the importance of subject suggestion for each student's higher education.

Table 4. Student academic details after subject recommendation

Total Number of Students		1996	Overall Pass Percentage
Male		1099	69%
Female		897	76%
Professional Courses	MCA	502	72%
	MBA	498	69%
	M.Com Hons	545	74%
	M.Tech	451	76%
District	Srinagar	1298	77%
	Ganderbal	417	69%
	Budgam	281	71%

As can be seen, there were 2026 students on whom the experiment was conducted by offering subjects to students based on the above categories, and out of them only 1996 students chose the recommended subjects. The overall pass rate was found to have climbed significantly from 68 percent to 72 percent. Furthermore, the percentages for each district have been increased, and so on. The following is a graphical representation below (Figure 3).

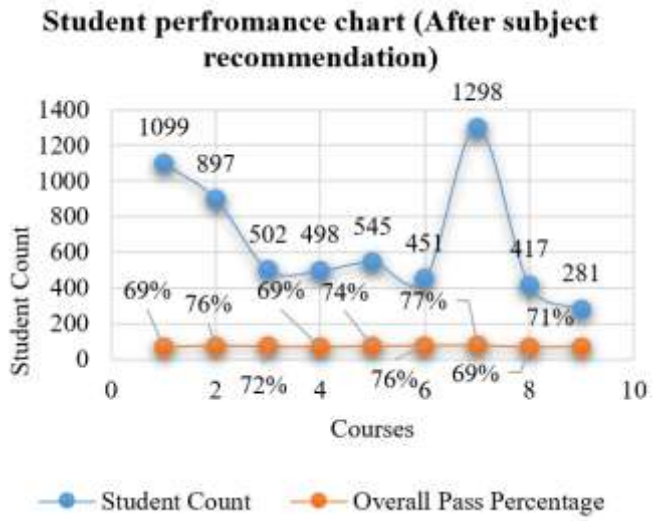

Figure 3. Graphical representation of student academic details after subject recommendation

The graph below (Figure 4) depicts the overall percentage rise following the implementation of the subjects recommended to students.

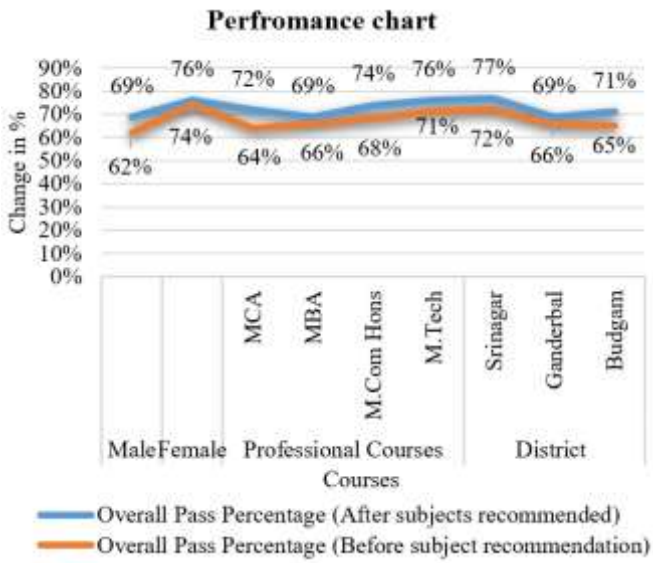

Figure 4. Performance analysis of percentages after and before subject recommendation

Thus, we can conclude with that there is a significant improvement in the students' performance if the subjects will be recommended to him/her based on the district he/she belongs and the gender. However, the biggest concern is that the experiment was carried out manually because the number of pupils were only approximately 2,000. Since the actual number of students at the undergraduate and graduate levels is 1,000 times higher, and our experiment was based on a small number of professional courses, the overall number of courses are roughly 80. Each course will have a different structure and stream, resulting in data heterogeneity and structural invariance. As a result, we will face the Big Data challenge, where such a large volume of data will be difficult to manage. Furthermore, with such a large amount of data, it is nearly impossible to predict how many students will apply for a certain stream.

5. RESOLUTION AND CHALLENGES: A BIG DATA ANALYTICS APPROACH

Higher educational institutions like universities and their affiliated colleges hold very huge quantity of data related to students, courses and staff. Analyzing this data can allow us to obtain insights which can enhance the operational effectiveness of educational organization. By doing statistical analysis on this educational big data, variables like student course selection, examination results, and career prediction of each student can be processed. Many elements of large data collection and its expansion in supporting and boosting student performance can be included in the big data framework, which has been accepted as a new technical method. Ashraf et al. [34], Zaman et al. [35] defines big data as a new generation of data analysis approach centered on obtaining, aggregating, and analyzing very huge amounts of data. Big data is defined as the utilization of massive amounts of data created every second across the Internet (Villars et al. 2011), which is then extracted to maximize its potential and value for the user [36, 37]. Because data is used at every stage of the educational process, it's intriguing to see how big data may help with topic selection and student achievement. Because data obtained since 2002 has mostly focused on automation, there is a lack of data integration, and the University of Kashmir thus supplies sophisticated data. The question becomes, "What do we intend to do with these data?" This study tries to find a strategy to improve student learning performance by selecting disciplines where a big data framework may be beneficial. As mentioned in the prior section, the model is anticipated to contribute by integrating multiple sources of knowledge [38-40].

The following are some of the challenges [41-43] faced while using data mining techniques to study the University's big data: The data at the university is heterogeneous, lacking in atomicity, and the main issue is that the data is inconsistent and redundant.

6. CONCLUSIONS

The existence of big data opens up exciting new research avenues and the opportunity to add new perspectives to existing educational threads. In education, all types of big data present opportunities and challenges. Big data methods are powerful tools for analyzing learner processes because of the sheer amount of micro level data available, but this power can lead researchers to overlook broader and conceivably more important patterns that cannot be analyzed at the individual level. This paper demonstrated a manual approach to handling academic data in which some patterns are generated and a particular subject is recommended to the student based on the patterns generated, which are based on three categories:

geographical location, previous subject details, and gender. Initially, a manual approach was used, and the experiment was conducted on approximately 2,000 students. It was discovered that after the subjects were recommended to the students, there was an overall increase in performance. Because this approach was tested on a smaller set of data, we will need to use a big data framework to generate new data-driven techniques to improve informed decision-making and efforts to improve educational efficacy.

REFERENCES

[1] Chen, J., Becken, S., Stantic, B. (2021). Harnessing social media to understand tourist mobility: The role of information technology and big data. Tourism Review, 77(4): 1219-1233. https://doi.org/10.1108/TR-02-2021-0090

[2] Beerkens, M. (2022). An evolution of performance data in higher education governance: A path towards a 'big data'era? Quality in Higher Education, 28(1): 29-49. https://doi.org/10.1080/13538322.2021.1951451

[3] Gao, P., Li, J., Liu, S. (2021). An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. Mobile Networks and Applications, 26(5): 2123-2126. https://doi.org/10.1007/s11036-021-01777-7

[4] Zheng, Y.F., Zhao, Y.N., Bai, X., Fu, Q. (2021). Survey of big data visualization in education. Journal of Frontiers of Computer Science & Technology, 15(3): 403-422. https://doi.org/10.3778/j.issn.1673-9418.2009014

[5] Dalwai, T., Mohammadi, S.S., Chugh, G., Somerville, A. (2021). Big data analytics and accounting education: A systematic literature review. Fourth Industrial Revolution and Business Dynamics, 159-174. https://doi.org/10.1007/978-981-16-3250-1_8

[6] Almutairi, M.M. (2021). Role of big data in education in KSA. International Journal of Information Technology, 13(1): 367-373. https://doi.org/10.1007/s41870-020-00489-7

[7] Daniel, B.K. (2017). Big data in higher education: The big picture. In: Kei Daniel, B. (eds) Big Data and Learning Analytics in Higher Education. Springer, Cham, 19-28. https://doi.org/10.1007/978-3-319-06520-5_3

[8] Ashraf, M., Zaman, M., Ahmed, M. (2020). An intelligent prediction system for educational data mining based on ensemble and filtering approaches. Procedia Computer Science, 167: 1471-1483. https://doi.org/10.1016/j.procs.2020.03.358

[9] Ashraf, M., Zaman, M., Ahmed, M. (2018). Using ensemble StackingC method and base classifiers to ameliorate prediction accuracy of pedagogical data. Procedia Computer Science, 132: 1021-1040. https://doi.org/10.1016/j.procs.2018.05.018

[10] Logica, B., Magdalena, R. (2015). Using big data in the academic environment. Procedia Economics and Finance, 33: 277-286. https://doi.org/10.1016/S2212-5671(15)01712-8

[11] Zakir, J., Seymour, T., Berg, K. (2015). Big Data Analytics. Issues in Information Systems, 16(2). https://doi.org/10.48009/2_iis_2015_81-90

[12] Ahmed, A.I. (2016). Big data for accreditation: A case study of Saudi universities. Journal of Theoretical and Applied Information Technology, 91(1): 130.

[13] Chaurasia, S.S., Kodwani, D., Lachhwani, H., Ketkar, M.A. (2018). Big data academic and learning analytics: Connecting the dots for academic excellence in higher education. International Journal of Educational Management, 32(6): 1099-1117. https://doi.org/10.1108/IJEM-08-2017-0199

[14] Santoso, L.W. (2017). Data warehouse with big data technology for higher education. Procedia Computer Science, 124: 93-99. https://doi.org/10.1016/j.procs.2017.12.134

[15] Dwivedi, S., Roshni, V.K. (2017). Recommender system for big data in education. In 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH), Hyderabad, India, pp. 1-4. https://doi.org/10.1109/ELELTECH.2017.8074993

[16] Sin, K., Muthu, L. (2015). Application of big data in education data mining and learning analytics-a literature review. ICTACT Journal on Soft Computing, 5(4): 1035-1049. https://doi.org/10.21917/ijsc.2015.0145

[17] Shah, T.H. (2022). Big data analytics in higher education. In: Information Resources Management Association (eds) Research Anthology on Big Data Analytics, Architectures, and Applications, pp. 1275-1293. IGI Global, Hershey. https://doi.org/10.4018/978-1-6684-3662-2.ch061

[18] Okoroma, N.S. (2008). Admission Policies and the Quality of University Education in Nigeria. Educational Research Quarterly, 31(3): 3-24.

[19] Ramesh, V., Parkavi, P., Ramar, K. (2013). Predicting student performance: A statistical and data mining approach. International Journal of Computer Applications, 63(8): 35-39. https://doi.org/10.5120/10489-5242

[20] Sokout, H., Usagawa, T. (2021). Using access log data to predict failure-prone students in Moodle using a small dataset. In SHS Web of Conferences, 102: 04001. https://doi.org/10.1051/shsconf/202110204001

[21] Sahu, R., Dash, S.R., Das, S. (2021). Career selection of students using hybridized distance measure based on picture fuzzy set and rough set theory. Decision Making: Applications in Management and Engineering, 4(1): 104-126. https://doi.org/10.31181/dmame2104104s

[22] Zaman, E.M., Quadri, S.M.K., Butt, E.M.A. (2012). Information integration for heterogeneous data sources. IOSR Journal of Engineering, 2(4): 640-643. https://doi.org/10.9790/3021-0204640643

[23] Ahalt, S., Kelly, K. (2013). The big data talent gap. UNC Kenan-Flagler Business School White Paper, 1-15.

[24] De La Fuente, J., Villar, M., Estrada-Peña, A., Olivas, J.A. (2018). High throughput discovery and characterization of tick and pathogen vaccine protective antigens using vaccinomics with intelligent Big Data analytic techniques. Expert Review of Vaccines, 17(7): 569-576. https://doi.org/10.1080/14760584.2018.1493928

[25] Mayer-Schönberger, V., Cukier, K. (2014). Big data: A revolution that will transform how we live, work, and think. American Journal of Epidemiology, 179 (9): 1143–1144. https://doi.org/10.1093/aje/kwu085

[26] Huda, M., Anshari, M., Almunawar, M.N., Shahrill, M., Tan, A., Jaidin, J., Daud, S., Masri, M. (2016). Innovative teaching in higher education: The big data approach. International Conference on New Horizons in

Education (INTE), Vienna. https://doi.org/10.13140/RG.2.1.1267.6087
[27] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Numerical and experimental investigation of meteorological data using adaptive linear M5 model tree for the prediction of rainfall. Review of Computer Engineering Research, 9(1): 1-12. https://doi.org/10.18488/76.v9i1.2961
[28] Kaul, S., Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Is decision tree obsolete in its original form? A Burning debate. Revue d'Intelligence Artificielle, 36(1): 105-113. https://doi.org/10.18280/ria.360112
[29] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Knowledge discovery in geographical sciences—a systematic survey of various machine learning algorithms for rainfall prediction. International Conference on Innovative Computing and Communications, pp. 593-608. https://doi.org/10.1007/978-981-16-2597-8_51
[30] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). Performance evaluation of GINI index and information gain criteria on geographical data: An empirical study based on JAVA and Python. International Conference on Innovative Computing and Communications, pp. 249-265. https://doi.org/10.1007/978-981-16-3071-2_22
[31] Fayaz, S.A., Zaman, M., Butt, M.A. (2021). An application of logistic model tree (LMT) algorithm to ameliorate prediction accuracy of meteorological data. International Journal of Advanced Technology and Engineering Exploration, 8(84): 1424-1440. https://doi.org/10.19101/IJATEE.2021.874586
[32] Fayaz, S.A., Zaman, M., Butt, M.A. (2021). To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: an empirical discourse of geographical data mining. Procedia Computer Science, 184: 935-940. https://doi.org/10.1016/j.procs.2021.03.116
[33] Fayaz, Sheikh Amir, S. Jahangeer Sidiq, Majid Zaman, and Muheet Ahmed Butt. "Machine Learning: An Introduction to Reinforcement Learning." Machine Learning and Data Science: Fundamentals and Applications (2022): 1-22.
[34] Fayaz, S.A., Zaman, M., Kaul, S., Butt, M.A. (2022). How M5 Model Trees (M5-MT) on continuous data are used in rainfall prediction: An experimental evaluation. Revue d'Intelligence Artificielle, 36(3): 409-415. https://doi.org/10.18280/ria.360308
[35] Zaman, M., Kaul, S., Ahmed, M. (2020). Analytical comparison between the information gain and Gini index using historical geographical data. International Journal of Advanced Computer Science and Applications, 11(5): 429-440. https://doi.org/10.14569/IJACSA.2020.0110557
[36] Mohd, R., Butt, M.A., Baba, M.Z. (2020). GWLM–NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction. Data Technologies and Applications.
[37] Zaman, M., Butt, M.A. (2012). Information translation: A practitioners approach. World Congress on Engineering and Computer Science (WCECS), San Francisco, USA.
[38] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). A super ensembled and traditional models for the prediction of rainfall: An experimental evaluation of DT versus DDT versus RF. In Communication and Intelligent Systems, Springer, Singapore, pp. 619-635.
[39] Altaf, I., Butt, M.A., Zaman, M. (2022). Disease detection and prediction using the liver function test data: A review of machine learning algorithms. International Conference on Innovative Computing and Communications, pp. 785-800. https://doi.org/10.1007/978-981-16-2597-8_68
[40] Altaf, I., Butt, M.A., Zaman, M. (2021). A pragmatic comparison of supervised machine learning classifiers for disease diagnosis. 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1515-1520. https://doi.org/10.1109/ICIRCA51532.2021.9544582
[41] Fayaz, S.A., Kaul, S., Zaman, M., Butt, M.A. (2022). An adaptive gradient boosting model for the prediction of rainfall using ID3 as a base estimator. Revue d'Intelligence Artificielle, 36(2): 241-250. https://doi.org/10.18280/ria.360208
[42] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall. International Journal of Advanced Technology and Engineering Exploration, 9(89): 511-524. https://doi.org/10.19101/IJATEE.2021.874647
[43] Fayaz, S.A., Zaman, M., Kaul, S., Butt, M.A. (2022). Is deep learning on tabular data enough? An assessment. International Journal of Advanced Computer Science and Applications, 13(4): 466-473.