Tracking Students' Progress using Big Data Analytics to enhance student's Employability: A Review

Anfal Al Yousufi, Vikas Rao Naidu, Karan Jesrani, and Vishal Dattana
Department of Computing, Middle East College, Muscat, Oman
[anfal@ mec.edu.om, vikas@mec.edu.om, kjesrani@mec.edu.om, vishal@mec.edu.om]
www.mec.edu.om

Abstract

The amount of data made available is enormous. Hence having a way to track both learners' progress and enable the institutions to identify and track their success rate, status, achievements, and weakness in comparison to other benchmarked institutions would allow them to be more proactive and progressive. Benchmarking would help check the performance rate of their learners, faculty and curriculum performance against similar ones for better insights into potential future enhancements. The end result would be to enable decision makers to detect, analyse, understand and predict the following: education progress, learners' behaviour, and course outcomes. This research paper will provide an insight on various aspects of Big Data Analytics (BDA), for the above purpose by reviewing various practices being followed. Finally, authors will provide recommendations for successful implementation of BDA to track students' performance and take necessary actions. This will be helpful in tracking their academic progress and hence enhance their employability skills by identifying the areas of improvements.

Keywords: Big Data Technology, Big Data Analysis, e-learning, decision making, employability.

1. Introduction

Big data technology helps the institution to track and identify their success rate and help analyze their business situation, key performance indicators (KPI) and academic accreditation. This would help them to provide evidence of their success rate and identify areas of improvements which would allow them to make more proactive and progressive decisions. Benchmarking would enable decision makers to detect, analyze and predict which would help them identify the learners' weaknesses, strengths and predicting who might fail or pass. As per Bichsel, "education needs investment in analytics professionals who contribute to defining the key questions throughout all process till developing data models to design and deliver alerts, recommendations, dashboards, and reports"

Being able to do analysis would help with clustering and grouping based on learning difficulties and interaction patterns, and on the other hand, Big Data Technology implementation would help with educational data mining and learning analytics. It would also help methods of education.

The rule of data in educational environments is evolving personalize education and enhance learning, teaching and assessments in comparison to the traditiona

and has the very potential to affect institutions, students and suppliers. Hence, making sure privacy is at the top priority while doing data analysis of the students results / Performance.

Modern universities employ student data to comprehend students' educational experiences. As a result, teachers are able to instruct students more effectively. In order to provide pupils with a well-rounded learning strategy, big data is also employed to change the educational system. (UpGrad, 2022).

The most advanced method for gathering, allocating, completing, and discovering large datasets is known as big data. Larger datasets are used. It uses a variety of patterns to anticipate data. Education and big data have a lot in common. This is due to the fact that educationalbig data offers a number of applications that support innovative teaching practices, effective educational administration, and effective research management (UpGrad, 2022).

Educational institutions must deal with student data much as businesses must deal with a lot of data pertaining to their customers, partners, staff, and financials. Huge amounts of data are created each year when thousands of students enroll in a variety of courses at various institutes. The course information, enrollment year, student ID, exam grades, and marks earned in each topic are all included in the student data. For the advancement of their jobs, analysis of this data can be quite helpful.

2. Literature Review

Educational institutes must treat the education process as an incomplete solution and thus should be treated as a research problem (Humboldt 1970). The methods for involving students in inquiry-based learning are interdependent (Healey and Jenkins 2009). Only blended learning can accomplish this (Gebre Yohannes et al. 2016) (Hasan et al. 2015) (Siddiqui et al. 2012), yet even that is insufficient. It is important to record information regarding learners and the learning process. EDM (Educational Data Mining) and Learning Analytics are a result of this (LA). Learning analytic

will be used in this case to monitor, gather, examine, and present the report data of individual students (Hadhrami 2017) and show continuous development of the undergraduates throughout the planned module (Shute, 2008), assist instructor in disseminating discovery, and assist administrators in forecasting future enrolment.

For comprehending and enhancing educational activity and the conditions in that takes place, analytics about learning means the process of measuring, collection, examination, and disseminating of information on students and their circumstances (Siemens & Gašević, 2012).

A number of studies in the learning area have lately been conducted with the goal of increasing the student's effort on the various educational activities. The improvement of the educational system has facilitated the finest possible growth of education in both students and teachers. Every aspect of the educational area, including design and implementation, is covered by the structural approach. The goal of this research was to make the education approach better for both teachers and students. The data from the Learning Management System is the primary focus of learning analytics (LMS).

Utilizing a learning management system has the advantage of aiding in the growth and enhancement of teaching and learning in an appropriate way. Learning analytics is becoming more and more important and well-liked. It is based on data from remote sources. There is also the theory that learning is dependent on LMS. It is typical for knowledge-embedding examined information on LMS to disregard tasks that originate from external sources, leading to the perception that only a minimal amount of a student's involvement in education is being considered.

The second claim shows that the system is designed with administrators' authority in mind rather than that of the pupils and teachers. Learning analytics therefore places less emphasis on individuals and more on educational measurement.

A project called Uatu has researched using data visualiser to ascertain the activities of the pupil involvement through the module content because the system has certain challenges identifying trackable contributions. The students collaborate to contribute to and keep information in the Google Drive sheets. These Google's documents provide a set of information on the students' work, including a list of edits and the content that each edit added. The study comes to the conclusion that no comments can be taken and obtained through the

data collecting. Students are therefore unable to provide appropriate group engagement. The end result demonstrates that locating the analytics can be challenging and that the system may be shut off.

According to Michael Wesch's observation, which he made after using a portion of the issue he was having to evaluate the development of the knowledge transfer services, the observer's goal was to establish the connection from inquiries to questions. Other strategy was by concentrating on teaching analytics and have the students interact with their own data. The system's efforts are explained to the pupils, along with how it will reflect on their actions in order to spot problematic behavior. As a result, learning analytics will help the researcher fully comprehend how the system is progressing. It focuses on how students perform, reflect, and respond in order to discover how students' behavior changes as a result of their comprehension.

Learning analytics, according to Hadhrami, are not a recent development; they have roots in various corporate intelligence and educational domains. The growing volume of high-quality data is a result of students taking part in various educational processes all across the world. Learning analytics' goals are to obtain knowledge, collect data, and adjust the educational system; as a result, they will help teachers use and develop their ways of transferring and seeking knowledge to give as many types as possible of problems that students need to practice and learn from (Hadhrami 2017).

Data from learning management systems are the primary focus of learning analytics (LMS). The information from the Learning Management System features showing figure of how many times each student has accessed the pages, student evaluation in accordance with the course pre-requirements, duration of students accessing the reading materials, duration of assessment completions, and participation in forum debates.

This cannot be sufficient in the rapidly altering world; visual analysis data employing video-streaming servers can bring additional advantages in addition to LMS, giving stakeholders more capability for better prediction, increasing pupil's achievement aspect, and optimizing learning and learner experience. The two analytics will be strengthened by blended learning, which will also favor research-based learning.

Dashboards offer real-time visualization that enhances student performance while saving teachers' time and efforts by giving students the

knowledge and tools they need to succeed. By giving learners and teachers the tools they need to assess their progress and attain the greatest outcomes, learning objectives are achieved more quickly. The benefit of learning analytics helps teachers improve instruction and learning so that students comprehend concepts clearly (Eckerson 2011). Additionally, learning analytics aids and assists educational establishments in enhancing student performance. The development of sound knowledge is based on the educational models that have been adopted by the learning analytics and data mining communities in educational institutions.

Students can be given numerous options in modern classrooms to learn content in various ways. Additionally, they have a variety of ways to provide access to the materials (Uskov 2016). With the help of different technical advancements like the Learning Management System (LMS) and the streaming server, these classes' learning environments are improved. These technological advancements make it easier to provide active classroom instruction, which is a component of blended learning.

LMS is used to provide a standardized condition for data communication, if it's for an evaluation or an activity for online students. High-quality video can be used for concurrent playback with a fair degree of latency thanks to streaming servers. Learning analytics is the process of analyzing and mining data that has been collected from the streaming server and LMS logs. Possible LA goals, which is surveillance, examination, future estimation, modifications, evaluation, check-up, adaptation, customisation, recommending points, and considerations, can be accomplished in this environment (Chatti et al. 2012).

In some ways, teaching is a creative endeavor where information is transferred to students through the arts. It is a teacher's responsibility to spread knowledge in the classroom using the most effective techniques. The teacher's job is to determine the needs of the students and any obstacles to learning they may be facing. Simulators are used in computer hardware instruction to help students comprehend it better by showing them how the hardware functions from the inside out (Hasan & Mahmood, 2012) (Siddiqui, Hasan, Mahmood, &Khan, 2012) (Kazmi, Hayat, Hasan, & Dattana, 2017). The learning experience for students is improved by employing blended learning, which requires the integration of both theoretical and practical concepts. help enhance the discovery learning process by allowing students to quickly perceive and understand examples. Students' learning experiences are improved by

integrating technology and multimedia applications, for instance GeoGebra for the analysis of statistical and mathematical side (GebreYohannes, Bhatti, & Hasan, 2016) (Bhatti, Hasan, Farsi, & Kazmi, 2017). Some tried learning that is based on activity, where the teacher watched the effects of teaching and learning, to improve student engagement and learning experiences (Hayat, Hasan, Ali, & Kaleem, 2017). The design of the learning environment promotes the finest teaching methods and maximizes student engagement (Bhatia & Naidu, 2017)(Naidu, Singh, Hasan, & Hadrami, 2017). Applications for interaction are used to deepen students' grasp of foundational ideas (Raweehi, Ansari, Udupi, & Naidu, 2017) (Farsi, Udupi, & Naidu, 2017). Instructors can quickly get used to and improve the process of transferring knowledge and getting knowledge (Naidu, Balushi, & Bhatia, 2017) by using games or tech-based education thanks to the important role that open-source tools play in education (Naidu, Singh, Harrasi, &Balushi, 2017). In their paper (Puka et al, 2020) present an experimental investigation of multidimensional relationships in online learning. Their efforts are directed at making sure the students are happy with the e-learning procedures. They used data gathered from the Bosnia and Herzegovina region to inform the survey they performed for their study. They employed the random sample method, selecting 11 BIH higher education institutions, and giving the students questionnaires. Confirmative factor analysis was used to examine and identify the data once the replies from the students were received. The well-structured mathematical models were operated to validate the model.

They have demonstrated in their research that the metacognitive strategies variable directly influences the satisfaction of the students while using e-learning, while the motivation, inclination, and goal setting variables, along with the environment structuring and communal scopes, indirectly affect the students' satisfaction. They have provided a plan of action for the advancement of e-learning based on the research they did. In their study, they offered eight hypotheses. Confirmatory factor analysis (CFA) and structural equation model analysis (SEM) have been utilized to generate the hypotheses and model their study. They have been using the Lisrel 8.8 program for analysis. The Cronbach's Alpha (CA) indicator was used to assess the reliability of the measurement scale. The SPSS 20 program's correlation analysis was used to examine the relationships between the variables.

The interaction between the learner and a specific instructional environment is the focus of the most

recent research. Soon, information about a learner from the time of his or her birth until the presentcan be made available in real time with the use of big data technologies and the internet of things (IoT) (Al- Emran, Malik and Al-Kabi, 2020). It will assist us in integrating the conventional performance and usage data gathered from the previous institutions and academic environments that each student has personally experienced, as well as the unique statistics of each student for various perspectives such as therapeutic, familiar, financial, spiritual, gender, liaison, sensitivity, psychosomatic, and so forth.

3.   Review methodology

This study uses a systematic strategy of literature review to accomplish the aforementioned goal (Baig et al.,2020). Finding the limits and research gaps in a certain topic through the examination of the literature forms thebasis of a good review. The process of examining, accessing, and comprehending the approach might be referred to as a systematic review. It details the pertinent research issues and field of study. Exploring and conceptualizing the existing research, identifying the themes, relationships, and gaps, and describing the appropriate future objectives are the main goals of performing a systematic review. The objectives of this investigation line up with the discovered causes. The Kitchenham and Charters (2007) techniques are used inthis study.

4.   BDA for students employability

Teachers will be able to better understand student behavior and performance by analyzing these data trails.Big Data makes it feasible to keep tabs on their behaviorin ways like:
- Exam question response times
- sources they pick out for research
- Questions are omitted.
- Questions that they've successfully respondedto

Students will receive feedback on their performance thatis considerably more enhanced thanks to the real-time analysis. Results may be considerably enhanced by the input. This is because pupils will be able to identify the subject areas in which they excel and those in which they fall short. There are other ways to measure a student's achievementoutside test scores. Big data and education work together to assist teachers gauge how long it takes a pupil to respond to a question. They can also tell what questions the kids struggle with.

Big Data aids teachers in monitoring their pupils' performance. The analysis aids in comprehending both the performance of an individual and a group. Teacherswill be able to identify the areas of interest among pupils thanks to the statistical analysis of individual grades. The grading scheme can be improved to emphasize the significant areas in which the student has excelled.

Using this method, teachers will be able to provide students with insightful feedback and help them determine the best career route. For instance, Ashford School in the UK uses big data technologies like Socrative, Nearpod, and Classroom Monitor to assess its students. They can monitor pupils' behavior andperformance in real time thanks to these programs.

Gaining a student's attention is one of the most intriguing and practical Big Data uses in education. There will always be those inattentive students who arepreoccupied with gazing at their phones or at others, regardless of how entertaining the lecture is. The use of student biometric information, such as pulse rate, facialexpressions, and things they touch during the lecture, has been proposed by big data professionals. Asmartwatch-like gadget or a camera mounted on the ceiling can both be used to record this data. This information may be used to assess each student's attention span. The instructor can then take theappropriate action to get the pupils' attention after receiving the data back.

Regardless of the enrollment at universities and colleges, big data education enables the development of customized curricula for each student. Big data use "blended learning" for this, which combines offline and online learning. As a result, the students have the choiceto follow the lessons they are interested in and learn at their own speed while still having access to offline instruction. Students will also be able to ask their lecturers for advice using these contemporary programs. Through MOOCs, this type of blended learning is accessible to students anywhere. For instance, almost
400 students registered for Andrew Ng's machine learning course at Stanford. But more than 100,000 students registered for this course when it was made available online through Coursera (UpGrad, 2022). In contrast to the Standard offline class, more students regarded the study material to be helpful and were able to understand the ideas.

Big data also aids in monitoring how graduates perform once they enter the workforce. Future students will find it easier to choose the right college and subject as a result of this.

Big data assists tutors in determining the following to decrease the amount of dropouts: How many students submitted their projects beyond the deadline?

- How many students attend each class, and how does that compare to other courses?
- What percentage of students in a specific course drop out?
- What is the decrease rate for a course in comparison to past years?

All of this data makes it easier for instructors and educational institutions to identify the precise causes of dropouts. As a result, they can offer better support for difficult learning aspects like academic writing. Big data also enables them to change the course curricula to make them more engaging.

5. BDT implementation

One of the biggest trends to note down is that big data is growing, and it is coming fast. The Educational Data Mining (EDM) is all boosting the learning outcomes by analysing the collected data. As follows business and scientific fields, educational research always finds ways to augment learning through this type of research (DataLab, 2013). In the past, collecting data to be perform the best practices was not cheap, but with the new shift, it became possible to collect huge amount of data efficiently. With that we can search on new patterns and algorithms to up notch education, this also helps model through developing statistical, machine-learning and data-mining methods.

Another implementation method is done for Learning Analytics is done for improving the learner's success rates which is done by analysing huge data sets and its contexts. It also unticks few points, such as, academic progression, prediction for future performance and detection of potential issues.

On the other hand, when implementing Big Data Technology, it falls around many challenges, especially within education, such as; security, privacy, ethics, timely analysis of data, efficient and effective data capture, storage, transfer, visualization.

6. Security

Big data technology security is the process of guarding huge sensitive chunks of data; including: first name, lastname, email address etc. Previously security had a lot of disasters where one had to have recovery plans and lots of policies such as password policy and firewalls as well as encryption and antivirus software. However, with BDT it promises for information security. This includes collecting, processing, analysing and extracting useful information from a huge dataset both structured and unstructured data.

Within security there comes variety of threats which includes log files, traffic data, user activity, physical security (Alguliyev, Imamverdiy, 2013). Not sufficient enough to secure advanced technology such as Big Data, Internet of Things (IoT), and cloud computing. Education Institutions lack adequate policies as well as access to data besides security the storage, transfer, and procession of huge sets of diverse educational data types

Some security requirements include: access control, threat filtering, activity monitoring, alerting. Data base environment because of the nature of big data and the analytics and the people that need to have access to it a lot of times permissions are granted on a very wide basis and that is because more people have access. With that, it increases the landscape for threats (endpoint threats, malware infections, account takeover, malicious insider use cases, or DDoS attack). In terms of activity monitoring, one needs to understand who is accessing what and if that is appropriate and does that violate some either regulation or data security standard within the organization. Then being able to get this information to the team that is responsible for securing it a lot of times, that means feeding it from the monitoring tools to some other monitoring mechanism. Hence, big data trend is not something to ignore especially when it comes to the challenges of securing.

Privacy

Data is key to unlocking lots of mysteries around us to finding cures and solutions that have plagued society forever. Privacy laws have been put around us to plan and protect us. We have this huge data around us, however we are able to use tiny fraction of it. Big Data in education requires transparency

that reveals the identity of the learners to inform decisions, e.g.: MOOCs "Massively Open Online Course". This shows the benefits to society what big data can yield which includes: Collect, centralize and analyse the learners' data. Privacy of data infuses both privacy and security controls into the data for data protection by default. This helps minimizes liability but still retaining the actual full value of data (Jeffrey Alan Johnson, 2014)

Ethics

Identifying the institution methods for preserving personal data privacy, individual consent, data ownership, and transparency. Data is the new oil energy, it is potential to outlining, defending and recommending concepts and how is the data used (Jeffrey Alan Johnson, 2014).

Lack of skilled professionals

Within this field and the new era of technology and big data, educators and learners need both to be trained to understand the system and to be able to use it and face any challenges around.

Data processing, storage and interoperability

Data comes in different forms, includes structured, semi-structured and unstructured. Hence it also comes with lots of challenges, includes; data integration, cleansing, and storage.

7. Conclusion and Future work

Big Data improves and revolutionizes the productivity of education outcomes using its technology all over the education system levels at teaching, learning, retention, administration, and reporting. This has helped organizations and institutions around the world to reduce the time spent on the selection process, it could also be used in recruitments.

BDTs in education are implemented for several purposes such as:
- Detection of learners dropping out risk rates
- Performance prediction
- Behaviors investigation
- Absences tracking
- Courses recommendations
- Instant assisting and assessments
- Visual analytics on learners' interaction

with a discussion forum
- Improvement of accessibility
- Research and development
- Evaluation and accountability
- Identification of learners at-risk of failing, andat the course level

Through Big Data Analysis, professors can identify areas where students struggle or thrive, understand the individual needs of students, and develop strategies for personalized learning allows students to choose their education path For example, Big Data Analysis may show that traditional, in-person learning methods lead to problemsin the performance of a student.

Acknowledgements

References

- A. Puška, E. Puška, L. Dragić, A. Maksimović, and N. Osmanović, "Students' Satisfaction with E-learning Platforms in Bosnia and Herzegovina," Technol. Knowl. Learn., 2020, doi: 10.1007/s10758-020-09446-6.
- Romero, C., and S. In their study paper, Ventura has put out a few futuristic concepts that they believe have an impact on the advancements in EDM/LA as well as future trends
- Bhatti, A. H., Hasan, R., Farsi, A. A., & Kazmi, S. I. (2017). Dynamic Technology Tool to Support Active Learning in Mathematics. 2017 International Symposium on Educational Technology (ISET). HongKong, China: IEEE.
- Bhatia, S., & Naidu, V. R. (2017). Flipped Teaching in Mathematics. International Conference on Education and New Learning Technologies. Barcelona, Spain: EDULEARN17.
- Chatti, M., Dyckhoff, A., Schroeder, U. and Thüs, H. (2012) "A Reference Model For Learning Analytics".
- International Journal Of Technology Enhanced Learning4 (5/6), 318
- Eckerson, W. (2011) Performance Dashboards. 1st edn. New York: Wiley
- GebreYohannes, H. M., Bhatti, A. H., & Hasan, R. (2016). Impact of multimedia in Teaching Mathematics. International Journal of Mathematics Trends and Technology (IJMTT), 39(1).
- Hadhrami, G. (2017) "Learning Analytics Dashboard To Improve Students' Performance And Success". IOSR Journal Of Research & Method In Education (IOSRJRME) 07 (01), 39-45
- Healey, M. and Jenkins, A. (2009) Developing undergraduate research and inquiry. York: HE Academy. York
- Humboldt, W.V. (1970) 'On the spirit and organisationalframework of intellectual institutions in Berlin', Minerva 8, pp. 242-267.

- Naidu, V. R., Balushi, H. A., & Bhatia, S. (2017). Effectiveness Of Free & Open Source Tools To Enhance Game Based Learning Experience In School Education. International Conference on Education and New Learning Technologies. Barcelona, Spain: EDULEARN.
- M. Al-Emran, S. I. Malik, and M. N. Al-Kabi, "A Survey of Internet of Things (IoT) in Education: Opportunities and Challenges," in Studies in Computational Intelligence, vol. 846, Springer Verlag,2020, pp. 197–209.
- Naidu, V. R., Singh, B., Harrasi, R. A., & Balushi, H. A. (2017). Technology Enhanced Learning Assisted By Free And Open Source Software. 4th International Conference on Education, Social Sciences and Humanities . Dubai, UAE: SOCIOINT.
- Raweehi, M. G., Ansari, W. A., Udupi, P. K., & Naidu, V. R. (2017). Interactive E-Book Developement using FOSS. Free and Open Source Software Conference, (pp. 67-70). Muscat, Oman.
- Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189. doi:10.3102/0034654307313795.
- Siddiqui, O., Hasan, R., Mahmood, S. and Khan, A. (2012) "Simulators As A Teaching Aid For Computer
- Architecture And Organization". 2012 4Th International Conference On Intelligent Human-Machine
- Systems And Cybernetics
- A. Puška, E. Puška, L. Dragić, A. Maksimović, and N. Osmanović, "Students' Satisfaction with E-learning Platforms in Bosnia and Herzegovina," Technol. Knowl. Learn., 2020, doi: 10.1007/s10758-020-09446-6.
- Carnegie Mellon University. What is EDM? - DataLab - Carnegie Mellon University. CMU - Carnegie Mellon University. https://www.cmu.edu/datalab/getting-started/what-is-edm.html
- Alguliyev, & Imamverdiyev. Big Data: Big Promises for Information Security. Big Data: Big Promises for Information Security. Retrieved October 2, 2022, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arn umb er=7035946
- Alan Johnson. View of The Ethics of Big Data in Higher Education. View of the Ethics of Big Data in Higher Education. Retrieved October 2, 2022, from https://informationethics
- Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, *17*(1), 1–23. https://doi.org/10.1186/S41239-020-00223-0/TABLES/11
- UpGrad. (2022). *Top 5 Interesting Big Data Applications in Education [2022] | upGrad blog*. UpGrad. https://www.upgrad.com/blog/big-data- applications-in-education/