2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA) | 978-1-7281-9437-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/CITISIA50690.2020.9371793

# Improving the quality of education system using Data Science Technologies : Survey

Pattinige Ravindra R Fernando
School of Computing and Mathematics
Charles Sturt University, Study Centre
Melbourne, Australia
itzmeraveen@gmail.com

*Abstract*— **Education is the most important and silent weapon in a country for both individual and country's economy. However lower level of adoption in the education system, poor decision making with less accuracy levels, adoption to new curriculums or subjects, teaching and learning styles are the main issues in education systems. These factors also have further long-term consequences for a country such as unemployment rates rises high, lack of suitable workforce for the demanding fields, individual dissatisfaction while being unemployed as well as in the community and socially. Unemployment rates are risen in Australia from past few years and this as a factor will be an ongoing issue if the government does not take any further actions to overcome these issues will definitely be direct hit to their economy in relation to work force in the present and future. Therefore the right technology should be implemented in order to obtain extract insights, obtain accurate decisions and high level adoption in education sector, as an example technologies such as data warehousing, big data, data mining, business intelligence and data analytics are in the peak of other industries such as aviation, retail, banking and other financial institutions. The main objective of this project is to facilitate a guide or a review for having data science technologies implemented in education sector in order to accomplish better education, as well as emphasis potential advantages of data technologies if it has been implemented in and around education systems.**

**Keywords- Data Science, Data warehousing in education; data warehouse potentiality in education sector; big data in education; data mining in education; business intelligence**

## I. INTRODUCTION

Education is the most essential element of a country's progress in terms of the financial, social and individual wellbeing. Simply if the government invest more on the education in return they are likely to get more human capital equipped with literacy and critical thinking [1]. Therefore, education must be given the number one priority by a country's government more specifically ministry of education, higher education institutions, colleges, other educational institutions both private and public etc.

Education sector shows lesser percentage of adoption to new syllabuses, curriculums, courses and other educational matters such as students' performance, student classification and many more. There is lack of accurate decision making and efficiency in the most of education systems all around the world. Therefore, use of data science and other technologies indeed to be implemented in and around the education systems as a solution.

Data is considered and treated as the most valuable intangible asset in any sector nowadays. Implementations of data science and other technologies such as big data, data mining and data analytics, business intelligence and machine learning has begun a new era of turning data into more advantageous information to improve the accuracy of decision making process and adoption to face many challenges. Simply the decisions are made on valuable information with proper insight.

Traditional Data warehousing Technique no longer capable of facilitating unstructured data analytics or mining like big data. To uplift the quality of education Big Data, Data Mining, business Intelligence would be very helpful for both students and educational institutions to maintain the quality in many ways such as student predictions, performance analysis, effective and efficient decision making, and patter recognition etc. Data science and data technologies are popular in most of the organizations such as banking, retail, aviation, insurance, travel etc.

Although, those sectors are using and benefited with such technologies, but the education sector has low popularity to adopt these technologies [2] because of implementation of data warehousing, big data infrastructure is costly, considerable cost for software hardware, technical staff, skilled staff to be acquired and cost of consultation for these projects..

### *A. 1.1 Education Data Warehousing*

Educational data warehouse or in short form EDW is implemented on data that has been created by the educational institutions. Either it could be implemented for individually for institutions or if the government identifies the benefits that could be implemented with huge scope. Therefore the data that has been created through systems or by users and other external sources will be gathered into one repository with relational database approach [3] and built OLAP method for querying purposes rather than using OLTP. OLTP is used in mostly legacy applications as it is suitable for handling insert, update or delete functions. OLAP is mostly used in data warehousing environments as it enables to use of multidimensional data and analyze. There is a specific mechanism or process called when developing an EDW and to maintain the data flow, it is called the ETL process. Data in data warehouse will be basically

978-1-7281-9437-0/20/$31.00 ©2020 IEEE

filled with this process which in long term Extract, Transform and load as shown in the figure 1. These phases of the process have tasks formerly data is been available in the data warehouse.

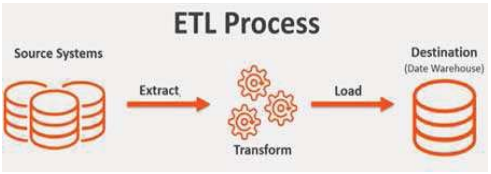


Figure 1 ETL Process example

### *B. 1.2 Education Data Mining*

Educational data mining EDM could also be used as another data technologies which will help to find hidden patterns of the education to find alternative solutions arises areas. There are special tools and software for Data Mining such as WEKA, Python, Orange, Rapid Miner etc. in data mining the data warehousing data and other cleansed data form the sources will be applied then it will be loaded in to Data mining tool to evaluate the patterns for knowledge purposes and decision making.

### *C. 1.3 Business Intelligence (BI)*

Business Intelligence is having high demand in other sectors although BI implementations in Education environment are not that easy task since this has to adopt and synchronize with technology stack, higher management support and relevant other factors. Business Intelligence implementation benefits mostly higher level of management to take better decisions. BI implementation is a collection technologies, tools and analysis that produces visualized reports for quick and easy understandings. Although higher education Business Intelligence oversee various advantageous outcomes, it is recommended perform a readiness check with couple of main factors such as Organizational, Technology and Social [9].

## II. LITERATURE REVIEW

This latest study about educational data mining proves that yet EDM is in emerging area and researches classifies the different types of EDM as well as suggested by the researches that it could basically tracks students in Australia from grade 1 to higher school and then to university achievement as well which would help to find patterns of large amounts of data. EDM implementation would be a benefit to universities, statistical departments and other government agencies in order to reveal insight and information to make efficient and effective decisions. The study also revealed that gender plays an important role when it comes to education [4]. The implementation of logical database for data warehousing would manipulate large amounts of data with millions of records of people would store and computerized. This would allow data mining techniques to make better accurate decisions [5]. The researched has identified that it is important to predict the future and improve the productivity of the higher education by bringing these data analytical tools together [6].

The study by these researches has proposed a framework to predict students' academic performance. To initiate the process first year bachelor students in computer science data was gathered from 2006 to 2014. In order to predict performance, Decision Tree, Naïve Bayes, and Rule Based classification techniques used and they have determined that the Rule Based model is the most suitable model which has provide highest accuracy value of 71.3% [7]. Study presented a review of latest big data technologies could associated with data mining and analytics in the education sector. Students generate large amounts of data by online that will not be able to use because they are not capable of processing those huge volumes of unstructured data. Business Intelligence tools has a significant capability to process raw data to be viewed as information which could support decision making process [8].

The big data applications now a days could result in innovative idea of teaching based on big data applications to enhance learning processes. Thematic analysis has deployed to create a model for higher education institutions. [10].The main issues with higher education sector are that the academic success is linked with the retention of students globally. The study has pointed out that the traditional data processing of analysing structured and data warehousing methods will not be able to accomplish goals and advantages achieved by big data implementation. The big data technologies would have many advantages. Therefor to overcome these issues big data analytics has being recommended where it would be advantageous in many ways, such as it will help students for better placement processes, accurate enrolment forecasts and more importantly early warring of the students who are more likely to be at a risk by failing or dropping out [11]. Data warehousing technologies are no longer be capable of handling the large amount of data that is being collected by universities and suggest that the data warehousing techniques must be extended to big data solutions to handle data and process information for decision making strategies. Therefore, big data analytics tools like Hadoop to data staging. Big data technologies implemented by educations system that will be helpful to make decisions effectively and efficiently [12].

## I. RESULT

Education has lower level of adoption and has lack of efficient, accurate decision-making capabilities. At this stage number of unemployment rates are increasing all around the world including Australia this has caused mainly because of poor education system and many other factors. However, the education factor is only considered in this research and focus of gaining more control over education with the latest data technologies available to mitigate the inefficiency and inaccuracy levels of decision making as well as for higher level of adoption. Therefore this research scope is determined as the data warehousing and other data technologies to use in order to enhance the quality of education sector in any country or by educational institution.

*A. Analysis/Findings*

In order to control the scope, articles published from 2011 to 2019 are being reviewed. However, the data warehousing concepts are published up to mid-2010. Data Mining, Big data, BI readiness are topics had more popularity later years of the articles published. Still the scope or having these data technologies implemented was the topic although haven't had any sort of confirmation of vastly implemented these technologies in the education sector.

According to [4] which was latest research study that has published about data mining in education and having track of students from grade 1 to higher school and then to university achievements will definitely be able to find patterns, trend analysis and will be help to find answers for unresolved issues in the education sector which is a better idea.

Furthermore the impact of data warehousing and other technologies implementations are mostly considered by the researches as a topic, although there is a lesser amount of real world implementations and most of the articles enforces and emphasize the need of education data warehousing, educational data mining and analytics, big data in education etc.

**Figure 2**. Below shows about the reviewed research paper's main data technologies which has been discussed. This also reflects one important factor that these technologies are potential implementing in education sector.

**Figure 3.** Below shows the year that these research papers has been published. Which clarifies that DWH concepts were discussed from 2011 and gradually evolving other data science technologies such as in 2012 educational data mining (EDM), 2013 big data, 2016 business Intelligence (BI) and finally modern DWH in 2018. . Moreover it also reflects another important factor that researchers of these articles had an interest using these technologies in education sector exactly the same time frame where other sectors such as banking, aviation and retail using data science technologies.

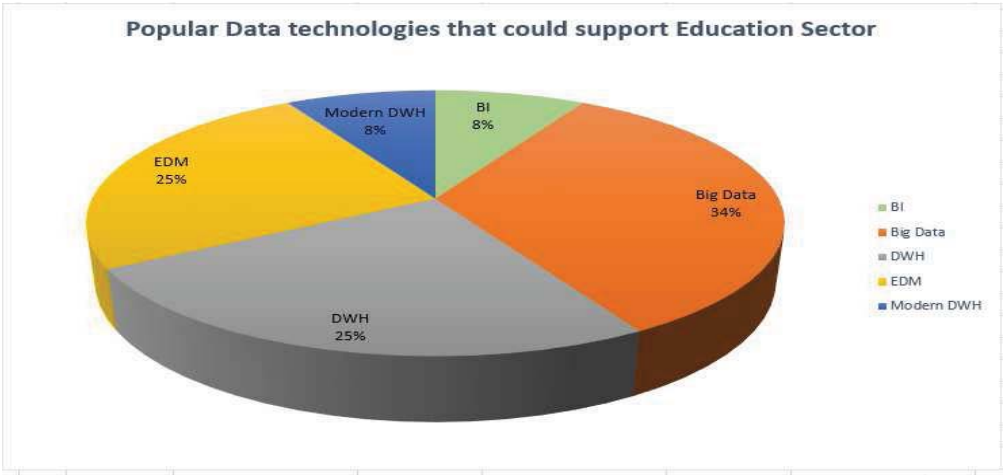


Figure 2 Popular Data technologies that could support Education Sector

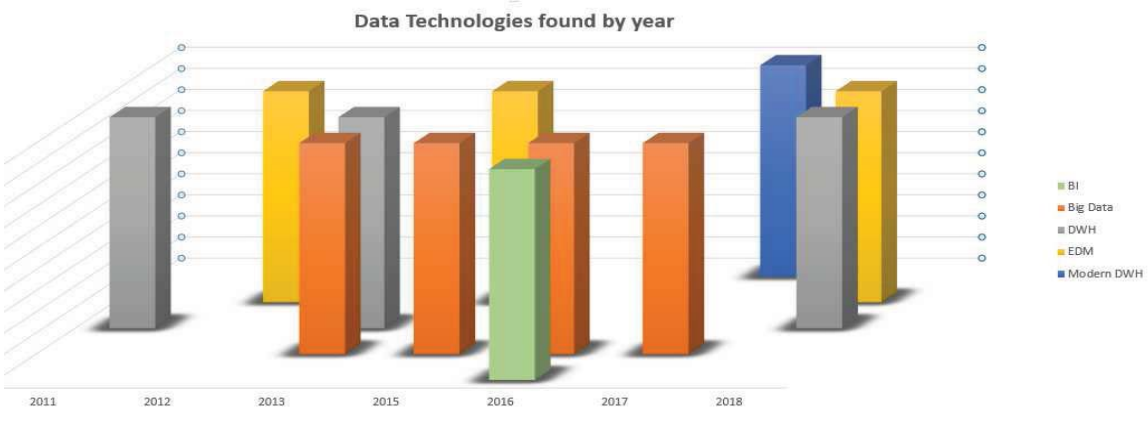


Figure 3 Data technology articles found by the year

TABLE I. ANALYSIS OF DIFFERENT DATA TECHNOLOGIES

| Techniques | Techniques | Environment | Description | Findings |
|---|---|---|---|---|
| T1 [5] | Data warehousing | Logical architecture of data warehouse design in University and higher education system | Large volumes of data could be handled in a data warehousing system to e-governance | Data warehousing implementation with data mining tools will aid to take strategic decisions |
| T2 [3] | Data Mining | Applications of Data mining in Higher Education | Use of Educational data mining in the context of predicting student performance and better understand students | This article yet in 2012 oversee the advantages using Educational data mining with regards to support students and also take decisions |
| T3 [1] | Data Warehousing | Data mining in Higher education systems with the use of student's exam performance data to group students | The exam performance data has been inserted in to a MySql table to run through the data mining model | K- Means clustering is used to predict student's results which will be helpful for instructors and students to take necessary actions before failing. |
| T4 [6] | Big Data Analytics | Big Data Analytics in Higher Education | Different data generating resources will gone through the big data phases which are acquisition, Extraction, Integration, Analyzing and Interpretation. | Teaching process as well as learning process could be improved by the implementation of Big data environment and most importantly it cold predicts the future of students before they fail and help. |
| T5 [7] | Data Mining | IHS has large amount of information about students. Data Mining is the best way to disclose patterns and to discover valuable information. | Data Mining with the aid of machine learning, visualization and statistical techniques for Institutions of higher learning (IHS). | Accuracy level will be depending on the larger the data set and its completeness. |
| T6 [8] | Big Data Analytics | Smartphone usage of the students and online learning environments are being vastly spread over the years and now there have been created so much of data by students themselves as well. | as of years passed the data that is being created on behalf of students and other individuals in the higher education scope is very high and the emerge of the bid data concepts have been more popularized as data has been created through online with diverse formats and in large volumes | Since the volumes of data has risen up after being created online data in the educational systems the big data concepts has been much popular along with data mining techniques. |
| T7 [9] | Business Intelligence | Higher education institution BI project implementation | It is challenging to implement a Business Intelligence environment in Higher education institutions although it returns greater benefits | Large number of data is created and retrieved from both external and internal sources and it is not that easy to implement a BI solution for HEI, helpful for making decision making for higher management. |
| T8 [10] | Data Analytics<br>Data Mining<br>Big data | Presence of big data platforms will enable innovative teaching and learning. | Massive amount of data is being created through smart phones and apps by students, teachers and tutors every moment. These data would help to make decisions as well as innovating teaching and learning. | Effective learning experience will be able to enable through big data platforms |
| T9 [11] | Data Analytics<br>Data Mining<br>Data Visualization | Big data in higher education is one of the technologies that is to be used in the education institutions as there are vast number of online information sources and other resources with structured and unstructured data formats available. | Data with high volume, velocity, variety and veracity nowadays created using big data platforms to support to take decisions in meaningful, accurate and efficient approach. | Big Data architecture of learning analytics in education has been proposed. |
| T10 [12] | Traditional vs Modern Data warehousing | Modern Data warehousing concepts with the aid of big data technology for higher education. | Since traditional data warehousing concept and techniques no longer facilitates handling large amount of both unstructured and semi structured data, the modern way of data warehousing with big data technologies would facilitate the need in the higher education data analytics | Existing traditional data warehousing concepts has been replaced by the need of big data into modern data warehousing concept. |
| T11 [2] | Data warehousing | Data Warehousing implementation within educational system. | The publication gives more detailed information about SDLC in DWH for higher education project. | Data storage infrastructure has not been paid enough attention to provide more quality data for data analytics and BI solutions although articles discuss about data mining and data analytics. |
| T12 [4] | Data Mining, Clustering | Phases of EDM Data Analysis Procedure | Australian educational system has applied used data mining and techniques to benefit its education system. | Educational Data mining and machine learning and statistics information are used to find out more insights and valuable information about education diversity of gender. |

**Table 1.**

The above table provides a summarized insight of the reviewed articles. The techniques column states about main data technologies that has been discussed in the research papers. As an insightful summary it appears that the data warehousing, data mining and clustering, data analytics, business intelligence, big data are more discussed in the higher education environments other than education systems such as primary and secondary. As of the technologies implemented it implies that data warehousing implementation is an essential part of other data technologies to implement in and around education system appropriately.

TABLE II. METRICS CONSIDERED BY THE DATA TECHNOLOGIES THAT SUPPORTS EDUCATION SECTOR

| Techniques | Structured Data | Unstructured Data | Supports decision making | Supports Analytics and predictions | Supports KPI and dashboards | Pattern recognition | Require Specific Tools | Require Specific Hardware |
|---|---|---|---|---|---|---|---|---|
| T1[5] | ✓ | X | ✓ | ✓ | X | X | ✓ | ✓ |
| T2[3] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T3[1] | ✓ | X | ✓ | ✓ | X | X | ✓ | ✓ |
| T4[6] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T5[7] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T6[8] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T7[9] | ✓ | X | ✓ | X | ✓ | X | ✓ | ✓ |
| T8[10] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T9[11] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |
| T10[12] | ✓ | X | ✓ | X | X | X | ✓ | ✓ |
| T11[2] | ✓ | X | ✓ | X | X | X | ✓ | ✓ |
| T12[4] | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ |

**Table 2.**

Includes information about the key elements of the different research papers reviewed. Most importantly by this analysis reflects the factors such as supporting decision making, require specific tools and hardware to implement data science technologies in the education sector and the fact we have to consider is that this will be costly considering acquiring specific tools and hardware.

## II. CONCLUSION

Data science technologies are in the peak of their demand facilitating many other sectors and it is essential to implement data science environments in and around education systems in Australia to improve the quality of education because of various factors. There are significant advantages of data science technology implementations such as accurate decision making, adopt to new curriculums or subjects, strategic planning and most importantly performance monitoring of the students, teachers etc. As a guide, developing a data warehouse is the most essential part to use other data science technologies where we can recognize as the core of this concept, therefore it will be expensive and whilst indeed of various resources such hardware, software and skilled staff. So the data warehouse stores data in OLAP methodology to optimize the querying and analysis process of millions of data. Then on top of that Data mining, clustering, machine learning and business intelligence tools will appropriately implemented to enhance the quality of education in order to address the issues discussed in the paper.

## REFERENCES


[1] Bhise. R.B, "'Importance of Data Mining in Higher Education System,'" *IOSR J. Humanit. Soc. Sci.*, vol. 6, no. 6, pp. 18–21, 2013, doi: 10.9790/0837-0661821.

[2] O. Moscoso-Zea, J. Paredes-Gualtor, and S. Lujan-Mora, "A Holistic View of Data Warehousing in Education," *IEEE Access*, vol. 6, pp. 64659–64673, 2018, doi: 10.1109/ACCESS.2018.2876753.

[3] M. Goyal and R. Vohra, "Applications of data mining in higher education," *Int. J. Comput. Sci.*, vol. 9, no. 2, p. 113, 2012.

[4] B. M. M. Alom and M. Courtney, "Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia," *Int. J. Inf. Technol. Comput. Sci.*, vol. 10, no. 2, pp. 1–9, 2018, doi: 10.5815/ijitcs.2018.02.01.

[5] P. Bhanti, U. Kaushal, and A. Pandey, "E-Governance in Higher Education: Concept and Role of Data Warehousing Techniques," *Int. J. Comput. Appl.*, vol. 18, no. 1, pp. 15–19, 2011, doi: 10.5120/2249-2879.

[6] B. Tulasi, “Significance of Big Data and Analytics in Higher Education,” *Int. J. Comput. Appl.*, vol. 68, no. 14, pp. 21–23, 2013, doi: 10.5120/11648-7142.

[7] F. Ahmad, N. H. Ismail, and A. A. Aziz, “The prediction of students’ academic performance using classification data mining techniques,” *Appl. Math. Sci.*, vol. 9, no. 129, pp. 6415–6426, 2015, doi: 10.12988/ams.2015.53289.

[8] K. Sin and L. Muthu, “"Application of Big Data in Education Data Mining and Learning Analytics – a Literature Review "," *ICTACT J. Soft Comput.*, vol. 05, no. 04, pp. 1035–1049, 2015, doi: 10.21917/ijsc.2015.0145.

[9] N. A. Hasan *et al.*, “Business intelligence readiness factors for higher education institution,” *J. Theor. Appl. Inf. Technol.*, vol. 89, no. 1, pp. 156–163, 2016.

[10] M. Huda *et al.*, “Innovative teaching in higher education: The big data approach,” *Turkish Online J. Educ. Technol.*, vol. 2016, no. NovemberSpecialIssue, pp. 1210–1216, 2016.

[11] F. Matsebula and E. Mnkandla, “A big data architecture for learning analytics in higher education,” in *2017 IEEE AFRICON*, Sep. 2017, pp. 951–956, doi: 10.1109/AFRCON.2017.8095610.

[12] L. W. Santoso and Yulia, “Data Warehouse with Big Data Technology for Higher Education,” *Procedia Comput. Sci.*, vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.