DATA ANALYTICS TO IMPROVE EDUCATION – A SURVEY

Dr. Sangeetha Muthuraman
Assisstant Professor,
Saradha Gangadharan College,
Velrampet,
Pondicherry

Abstract: Data analytics is an important research area as it is the most trending one which is used to solve a number of prediction and subsequent decision-making problems in different domains like manufacturing, banking, retail, healthcare, education and many more. Analytics in education domain is not popularly used, however is an important area of research by which there are a number of benefits that could be reaped. Especially in case of online education activities, enormous amountof useful data could be captured which may be used to improve the quality of education. With this idea,this paper identifies the major areas where analytics could be applied and presents its sub-areas as well.

Keywords: Data analytics in education, Classification in education analytics.

Introduction:

Although data analytics have been used in a number of domains, it has not been used much in the education domain. There are a number of benefits that could be reaped if it is used in an effective manner [1,2, 17]. A few important benefits include, firstly, personalisation of learning experience based on student's capability thereby increasing subject interest to students. Secondly, it could also be used in designing new syllabus for students based on the current job demands etc. These timely decisions would help improve students' knowledge and their performance according to the market need.Therefore, research of analytics in education is extremely essential. Having this in mind a survey has been made to identify the areas of interest where analytics can play a role in the field of education.

The rest of the paper is organised as follows. Section 2 gives an introduction on data analytics and its classification. Section 3 presents an introduction on machine learning and a classification of its algorithms. Section 4 briefly describes the areas in the education domain where data analytics could be applied. Section 5 presents the conclusion.

Data Analytics:

It is a scientific process in which systematic analysis of data is made, which can be used to find relationships among data, predict future outcomes from the analysed data and make optimised decisions based on predictions. Data Analytics can be classified into three categories. They are,

- Descriptive analytics,
- Predictive analytics and
- Prescriptive analytics

Descriptive analytics [2, 17]: This category could be used to mine information from the past or current data by asking a question what happened? This gives the students, teachers and other stakeholders clarity on what has happened in the pastand what is the current scenario, or presents the past history of data until when the request was made. This history may consist of qualitative as well as quantitative data.

Predictive analytics [2, 17]: This categoryforecasts what will happen tomorrow using different statistical models. Predictive analytics is an important branch of analytics in which learning from the available data and the various conditions imposed on the data helps in gaining experience about the problem and uses this experience to predict future situations [7]. Predictive analytics rests on machine learning to learn from the available data and make predictions.

Prescriptive analytics [2, 17]: This category uses the results of both predictive and descriptive analytics to decide what needs to be done for attaining better quality? Or how to produce optimised results? Example, when considering the teaching and learning process, it analyses the cause-effect relationships and the predictions made to improve the teaching and learning principles, rules, models etc. and provides a better teaching and learning process.

Therefore, the heart of the work lies mainly on the predictions made during the predictive analytics. Since this category heavily relies on machine learning algorithms for its prediction, the next section presents an overview of machine learning.

3. Machine Learning:

Machine learning is a field in artificial intelligence which develops algorithms that learns to carry out tasks based on pragmatic data [7]. The machine learning algorithms employ a trial and error process [8] in order to learn how to carry out tasks from the given data. These machine learning algorithms could be classified in two types. They are the supervised and unsupervised learning algorithms.

In case of supervised learning, a set of labelled training data is given and an agent does a lot of trial runs and observes the input-output pairs in this labelled data and models a function to map the outputs from the input [7, 8]. Further, supervised learning algorithms are classified into classification and regression algorithms [3]. The difference between a classification and a regression problem is that the output variable in case of a classification problem is a category variable such as diseased or not diseased, yes or no etc. but in case of a regression problem the output is a real value such as rupees, weight etc.Example: Naïve Bayes, Support vector machines etc.are used for classification problems, while linear regression, is used for regression problems, random forests for both classification and regression problems [3].

In case of unsupervised learning, a set of unlabelled training data is given and an agent does a lot of trial runs and observes some patterns in this data from which the predictions can be made [7, 8]. Further unsupervised learning algorithms are classified into clustering and association rule learning algorithms. A clustering algorithm is one which clusters the given data set into n-clusters, based on conditions relevant to the problem, like grouping clients

based on buying patterns. Example of clustering algorithm include k-means algorithm [3]. An association rule learning algorithm is one that learns new association rules on the given dataset, like one buying pattern could be associated with another. Example includes the apriori algorithm [3].

There is another king of problem which is called as semi-supervised learning problems. In case of semi-supervised learning, only a small set of training data is labelled while the majority of the data are unlabelled. In this case both the supervised as well as unsupervised learning algorithms could be used.

Once again, these machine learning algorithms can be classified into parametric and non-parametric algorithms [4]. In case of parametric algorithms,a form of mapping function which is appropriate to the problem given,is assumed and the co-efficient for the function are learned from the training data. Examples include linear regression, logistic regression etc. In case of non-parametric algorithms, there is no strong assumption made regarding the mapping function, hence they can learn any functional form freely from the training data. Examples include neural networks, decision trees, support vector machines etc.

Having given an overview on data analytics and machine learning, the next section presents the survey on application of analytics in the education domain.

4. Data analytics in Education Domain:

Education analytics is nothing but collecting and analysing all data related to education in a systematic manner in order to make future predictions and optimised decisions to improve the quality of education.As far as the education process is concerned analytics can be applied in six major areas. They aredepicted in the fig.1 below

Fig. 1 Areas of Education Analytics

4.1 Teaching analytics [1]: collects and analyses data from which information are predicted and hence informed decisions could be made which would improve the quality of teaching.

a. Applications of teaching data analytics include analysing and predicting

i. Teachers' skill and performance – This could be measured using a number of factors such as
- Lesson preparedness
- Organisation of presentation
- Clarity of presentation
- Interesting presentation
- Use of examples and illustrations
- Course workload
- Course difficulty etc.

ii. Teaching activities [9] – This could be measured using a number of factors such as
- mode and environment of teaching
- collaborating with students etc.

iii. Teachers' behaviour detection [9] - This could be measured using a number of factors such as
- Attitude towards students and questions
- Fairness of testing and grading
- Adjustment of rate and level to students
- Accessibility to students outside class
- Helpfulness to homework assignments etc.

4.2 Learning analytics [1]: collects and analyses data about the student's learning abilities.

a. Applications of learning data analytics include analysing and predicting

i. Students' skill and performance [9, 10, 11, 12] – This pertains to the curricular and co-curricular activities assessments. This can be done by assessing their concept maps, concept tests, assignments, group works, creating and using rubrics [15] etc.

ii. Students' attrition risk detection [9]- This helps in predicting the students who may drop out from the degree or course and increasing his/her interest by paying more attention to the student thereby enabling retention of the student.

iii. Students' behaviour detection [9, 13, 14]- This could be measured using a number of factors such as
- Behaviour towards his peers
- Behaviour towards his teacher
- Students interest towards the course etc.

iv. Student collaborations [9] – This is assessed when he/she works in a team. This could be measured using a number of factors such as
- Students' attitude
- Students' friendliness to work in a team
- Students' preparedness for a collaborative task
- Collaboration score etc.

v. Students'Learning Personalisation: Any class may constitute students with different levels of basic knowledge. In addition, some students may be fast learners, some moderate and the rest slow learners. So, in order to cater to such variety of students by promoting their learning in an equal manner is a challenge. Additional importance should be given to slow and moderate learners so that they can cope equally with the fast learners. This can be done by deciding how the learning content should be delivered for slow, moderate and fast learners, without demotivating the entire class, thereby increasing the subject interest equally to all of them. Addressing this is a big challenge [6].One solution is to monitor and measure the learning abilities of the student [6]. This kind of analytics which shows how students learn, enables educators to adapt their teaching style according to the student.

4.3 Administrative analytics:collects and analyses data on routine administrative operations to assess how well it is performed.

a. Applications of administrative data analytics include analysing and predicting

i. Students' demographic data [9]
ii. Students'personal data
iii. Teachers'personal and demographic data [9]
iv. Administrative data
v. Administrative staffs' personal data
vi. Administrative staffs' skill and performance
vii. Administrative staffs' behaviour
viii. Administrative staffs' demographic data
ix. Intelligent feedback [9] – Having in place a system that could give immediate feedback to enquiries, which could be an automated one, would improve students' or any stakeholder's interaction and interest.
x. Students' fees etc.

4.4 Academic analytics:collects and analyses data related to academic activities such as degree programs available, courses offered [7, 9], course evaluation, research etc. This analysis could be made at the institutional level to improve institutions performance. Some process like course evaluation are done to improve on the placement opportunities of the students thereby benefiting the stakeholders as well as improving the success rate of the institution.

a. Applications of academic data analytics include analysing and predicting

i) Degree performance and recommendation
ii) Course requirements and recommendation
iii) New course design [6,9]: One concept which is important in the education field is to modify the curriculum constantly according to the needs of the businesses or industries. This can be done by analysing the big data collected to understand the current employment trends in the market.
iv) Students' pathway – Based on his/her interest, skills, performance and behaviour, what has to be done once this course or degree is completed could be predicted.
v) Faculty development
vi) Publication of Faculties [16]
vii) Award nomination strategies [16]– Gold medallists etc.

4.5 Examination analytics:collects and analyses data related to pre-exam preparations, conducting exams and post-exam tasks.

a. Applications of examination data analytics include analysing and predicting

i) Mode (online or offline) and environment (home, college, university) of examination
ii) Type of examination (Quiz, MCQs, Written exam etc.)
iii) Question paper quality analysis
iv) Students evaluation
v) Result analysis etc.

4.6 Institutional analytics:collects and analyses data about the various institutions, their performance etc.

a. Applications of institutional data analytics include analysing and predicting

i. Institutions performance – can be assessedbased on the quality of education they provide [9], based on the students' placementsetc.
ii. Institutions infrastructure
iii. Institutions' alumni
iv. Institutions research
v. Funding opportunities or Grants received
vi. Annual program assessment
vii. Institutional progress [16]
viii. Creating research centre [16]
ix. Conferences conducted
x. Faculty development programs conducted [16]
xi. Awards obtained [16]
xii. Institutional Ranking etc.

Therefore, on a broad base the categorisation of education analytics has been made with its sub-categories to some level.

5. Conclusion: Data analytics is a growing research area. Analytics has been applied to many fields such as business, engineering, healthcare etc. But it is not so popularly used in the education domain, even after the popularity of online mode of education. Hence, this was the motivation to identify the areas where data analytics could be applied to improve the overall quality of education. Therefore, this paper classifies the major areas under the education domain where analytics can play a role stating a few sub-areas. These areas could be studied in detail in the future, and the implementation of how analytics could be applied to these areas, which machine learning technique may be more suitable etc. can be demonstrated in detail.

References:

1. Ifeanyi G. Ndukwe, Ben K. DanielID and Russell J. Butson, "Data Science Approach for Simulating Educational Data: Towards the Development of Teaching Outcome Model (TOM)", Big Data Cognitive Computing, 2018, 2, 24; doi:10.3390/bdcc2030024
2. IBM Software, "Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics", static.ibmserviceengage.com
3. jason brownie, "Master Machine learning algorithms", https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
4. K. Sree Divya, Peyakunta Bhargavi and Jyothi Singaraju, "Machine Learning Algorithms in Big data Analytics", INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING", 6(1):63-70, January 2018 DOI: 10.26438/ijcse/v6i1.6370
5. Anindita Desarkar and Ajanta Das, "Big-Data Analytics, Machine Learning Algorithms and Scalable/Parallel/Distributed Algorithms", Internet of Things and Big Data Technologies for Next Generation Healthcare, Springer Studies in Big Data, Part of the Studies in Big Data book series (SBD, volume 23), 03/01/2017, Pages 159-197, DOI: 10.1007/978-3-319-49736-5_8
6. Anonymous, "THREE WAYS EDUCATORS ARE USING BIG DATA ANALYTICS TO IMPROVE THE LEARNING PROCESS",Written on 29 May 2018, https://master.edhec.edu/news/three-ways-educators-are-using-big-data-analytics-improve-learning-process
7. Pekka Malo, "Introduction to Predictive Analytics – MyCourses", *mycourses.aalto.fi › content › PredictiveAnalytics-Part1-2016-handout.pdf*
8. The data science booklet, https://365datascience.com/wp-content/uploads/2018/09/The-365-Data-Science-Booklet.pdf
9. Katrina Sin and Loganathan Muthu, "APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS – A LITERATURE REVIEW", ICTACT JOURNAL ON SOFT COMPUTING: SPECIAL ISSUE ON SOFT COMPUTING MODELS FOR BIG DATA, JULY 2015, VOLUME: 05, ISSUE: 04, ISSN: 2229-6956 (ONLINE) DOI: 10.21917/ijsc.2015.0145

10. Vince Kellen, Cutter Consortium, Adam Recktenwald and Stephen Burr, "Applying Big Data in Higher Education: A Case Study", *Cutter Consortium*, Vol. 13, No. 8, pp. 1-39, 2013.
11. Paulo Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks", *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pp. 110-116, 2011.
12. Behzad Beheshti and Michel Desmarais, "Predictive performance of prevailing approaches to skills assessment techniques: Insights from real vs. synthetic data sets", *Proceedings of the 7th International Conference on Educational Data Mining*, pp. 409-410, 2014.
13. Joseph Grafsgaard, Joseph Wiggins, Kristy Elizabeth Boyer, Eric Wiebe and James Lester, "Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue", *Proceedings of the 7th International Conference on Educational Data Mining*, 2014.
14. Seong Jae Lee, Yun-En Liu and Zoran Popovic, "Learning Individual Behavior in an Educational Game: A Data Driven Approach", *Proceedings of the 7th International Conference on Educational Data Mining*, 2014.
15. Carnegie Mellon University, "Eberly Center – Teaching Excellence & Education Innovation, How to assess Students' Learning and Performance", https://www.cmu.edu/teaching/assessment/assesslearning/index.html
16. About Academic Analytics, https://academicanalytics.com/wp-content/uploads/2018/08/About-Academic-Analytics-2.pdf
17. Hugh J. Watson, All About Analytics, International Journal of Business Intelligence Research, IGI Global, 4(1), 13-28, January-March 2013.