The role of data science in online education

Songlak Sakulwichitsintu[1]

Abstract

This article presents a review of the major literature in data science and its implications over the past 12 years. It concurs with the view that choosing to study data science requires a research focus on big data, data analytics, learning analytics, and machine learning. Starting from the premise that the importance of data science is a given, the focus here is on the following categories: there are three types of data; big data includes five properties; data analytics consists of four types of data analytics; learning analytics; machine learning; online education; and data science in online education. Additionally, there are perspectives on ‘data scientists’, and the benefits of data science for educational institutions. In the online education environment of Massive Open Online Courses (MOOCs) and blended learning, big data and learning analytics are implemented to analyse students’ learning and track their learning success efficiently. The literature itself points to the crucial importance of the quality and management of data. Educational Data Science (EDS) can be used to solve problems encountered in students’ online learning.

Keywords: Data Science, Big Data, Data Analytics, Learning Analytics, Online Education

---

[1] Information Science Program, School of Liberal Arts, Sukhothai Thammathirat Open University, Nonthaburi, Thailand, E-mail: songlak.sak@stou.ac.th, songlaks.stou@gmail.com

(Received: 1 July 2021; Revised: 13 September 2021; Accepted: 8 August 2021)

Introduction

Data science is a field that uses calculation methods to obtain valuable data. The process begins by collecting sets of raw data. In data science, work ordinarily involves the following three types of data: (Anderson, Wiley, Anderson, & Swanstrom, 2015; Baldassarre, 2016).

• Structured data: Data is stored, processed and managed in a traditional relational database management system.

• Unstructured data: Data is generated by human activities in general and reveals inconsistencies with the structured database format.

• Semi-structured data: Data does not match in a structured database system but is structured by tags that are beneficial for making forms of orders and hierarchies in the data.

Although data contains both interesting and uninteresting data, it is often the case that the interesting data will be meaningful and will be analysed and stored. In terms of the data life cycle, the focus is on the data in digital form. It starts with creating data: first through cleaning and structuring of data, then using or analysing it through easy-to-understand concrete visualizations, and the exchange or dissemination of data. Finally, used information is kept while unused information is destroyed. The data life cycle consists of important elements usually in this order: acquire, clean, use/reuse, publish, and preserve/destroy. It resembles a biological structure: beginning with the birth of data, then on through active life until expiration occurs in the final stage. (Berman et al., 2018)

'Data Science' is defined as "a set of fundamental principles that support and guide the principled extraction of information and knowledge from data". The effective data scientist is expected to assess business problems and estimate the extent of value (Provost & Fawcett, 2013). Data science requires the precision of math and statistics skills, which are essential in understanding data and its importance. These skills are also beneficial in data science, because they are used for predictive forecasting, decision modelling, and hypothesis validating.

In other words, data science arises from the need to gather structured and unstructured data. It fundamentally employs statistical analysis methods, machine and deep learning, prediction, and problem solving. To be sure, Artificial Intelligence, machine learning, deep learning, and data science are related, but not exactly the same. Artificial Intelligence aims to create machines to operate and look like humans. However, these devices cannot completely replace humans.

Machine Learning seeks to create tools for extracting knowledge out of the data. It develops prototypes that are designed from independent data. This includes training features incorporating data that has been prepared. Deep Learning is the creation of multiple levels of neuron networks to enable in-depth learning. It facilitates more and faster analysis as needed, including the use of mathematical calculations. Data Science integrates arrays of data, visualization, insightful collection, and decision making relating to these data. Typically the steps of data science include data collection, validation, analysis, visualization, and decision making.

Data Science differs from mathematics or statistics in the sense that data science is an application of activities to meet specific needs or problems of users. Before we can solve a problem, we need to define the problem clearly. Optimum problem identification consists of openness, inquisitiveness, originality, and voluntariness (Saltz & Stanton, 2017). Data science is a combination of academics, skills, techniques, and tools that help organizations to explain trends and make assessments that lead to important decisions. It is in a constant state of evolution due to new technologies regularly being developed and applied. The achievements of data science and the skills of data scientist skills are essential to practicality of data analysis. Successful applications can be used for the benefit of individuals and organizations (Parks, 2017).

The new advances are derived from experience and observable research outcomes. An empirical approach in a technology-based environment, along with flexible and complex data, has yielded a new type of research called Data Science. From the beginning, this form of research methodology has been known as quantitative research. Then, a second research methodology was formulated called qualitative research. Subsequently a third research methodology was developed called mixed-method research. Lastly, a fourth research methodology with data-related form was created. It is known as data science research (Daniel, 2019).

New technologies were developed to support the use of big data for education in the form of changes in learning and teaching methods. These included the use of mobile technology for online teaching and learning. This led to student records being stored digitally. A feature of the management of the learning process is to solve student dropout issues and help students to be successful in their studies. This requires sufficient high-quality data in order to solve problems and enhance academic objectives. This would include accurate prediction of future events for effective self-learning, the benefits of collaborative learning, and students helping one another, such as spending time together to study creatively (KlaA nja-Milicevic, Ivanovic, & Budimac, 2017). Furthermore, machine learning techniques using student data, for example, forecast a dropout prediction rate for e-learning courses with accuracy, sensitivity, and precision (Lykourentzou, Giannoukos, Nikolopoulos, Mpardis, & Loumos, 2009).

Big Data

Big data is typically used to describe the data in which its capacity exceeds the processing capacity of general database systems, and because the data is too big and will change too quickly. Moreover, it may not be suitable for the structural needs of the traditional database structure. In addition, the amount of data will reach the terabyte level or the petabyte level. Data engineering also offers solutions that can be designed to meet the requirements and used based on the existing data. The five properties (five Vs) are used to profile big data characteristics including volume (size of data), variety (different types of data), velocity (speed of data generation), veracity (accuracy and reliability of data), and value (extracting useful information from data) (Grover & Kar, 2017; Gupta, Kar, Baabdullah, & Al-Khowaiter, 2018; Schleder et al., 2019). In the case of a value chain model of a typical big data ecosystem, the value chain functions work in collaboration with big data by using various analysis platforms until getting the desired value.

Big data essentially is a wide variety of collection and management techniques focusing on large data. Data science, in contrast, has developed a model that captures, visualizes, and analyses the basic patterns for developing business utilization (George, Osinga, Lavie, & Scott, 2016). The development and rapid evolution of big data has led to the adoption of these concepts in coordination with both the public and private sectors. This would include the influence of big data on academic development and growth. Big data uses analytical techniques of both structured and unstructured data. The process of extracting big data into five stages includes two main sub-processes of data management and analytics. Data management uses relevant technologies to support three of the stages: acquisition and recording; extraction, cleaning and annotation; and integration, aggregation and representation. Analytics is involved with the techniques of the remaining two stages: modelling and analysis; and interpretation of the intelligence from big data (Gandomi & Haider, 2015).

Data Analytics

Converting your raw data into actionable insights is the first step in the process of using the information you have collected that is truly beneficial to you. Data scientists focus on business by using data analytics to create insights from existing raw data. Data analytics is one approach to solve problems associated with big data by extending the analysis of bigger data sets from various data sources (Aasheim, Williams, Rutner, & Gardiner, 2015).

There are four types of data analytics that are most often encountered:

1) Descriptive analytics: This type is used to answer the question "What happened?" Descriptive analytics uses references from past and present data. Business analysts, as well as business-critical data scientists, will use modern business data for descriptive analytics.

2) Diagnostic analytics: This type is used to answer the question "Why did this particular thing happen?" or "What exactly went wrong?" This type of analytics is beneficial for deducing outcome probabilities. It suggests that the success or failure of the sub-elements derives from the data-driven reflection.

3) Predictive analytics: Although it utilizes past and present data, predictive analytics is quite capable of outperforming descriptive analytics. Predictive analytics is associated with the creation of complex models and analysis for predicting future occurrences or trends. In business-related scenarios, these analyses are conducted by business-critical data scientists.

4) Prescriptive analytics: This type needs to be adapted to work processes, structures, and systems to be more suitable for the processing of notices relying on predictive analytics. It is necessary to inform users what they should do. While both business analysts and business-critical data scientists can create prescriptive analytics, they have different methods and resources.

A data scientist is defined as “a high-ranking professional with the training and curiosity to make discoveries in the world of big data.” Many universities are planning to offer a degree in data science, which will include a big data analysis program. In addition, some companies seek to acquire their own data scientists to meet the needs of their employees and customers. The training departments of companies are therefore organizing training programs and providing certificates of data science and big data analytics. For instance, Kaplan has encouraged data scientists to provide productive learning strategies. It has been suggested that data scientists should have more practical skills, such as applying more traditional quantitative analysis to problems (Davenport & Patil, 2012).

Data science is of course centred on data, especially big data. Raw data is considered a part of data analysis. There is a growing number, especially of unstructured data such as text, images, and videos that are frequently networked among members. When connected, computers can exchange information with other units in the process of making decisions which may not need any input by humans. In such cases, users may see the need to be careful about ethics and accuracy, which can affect business operations or related laws. In some circumstances, databases may not be able to find information that would answer certain questions that users might have. The questions, however, sometimes may be formulated in incorrect patterns. An interesting point is that some things are not what is expected. The pattern that is expected to occur in the future is called 'prediction'. It can be said that there is a correlation between prediction and past data-related actions which can lead to significantly more confidence (Dhar, 2012).

Learning Analytics

One definition of learning analytics is "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs." (Ferguson, 2012). One feature of learning analytics is the collection of all data of the learners' learning experience. It covers studies undertaken before the registration, design of learning process, and assessing the learning needs of the learners. The development of a learning analytic needs to consider 1) developing innovative techniques, tools and staff, 2) access to all data, 3) the entire process of analysis, and 4) expert or any related features (KlaA nja-Milicevic, Ivanovic, & Budimac, 2017)

There are 18 weeks of teaching in the course, which consists of six learning activities. The teacher encourages students to keep focused on mastering the course and working on the homework exercises. Data sets of students' learning activities are categorized as online behaviour. After factoring in the learning data sets of blended learning course of the learners during the semester through learning analytics, it was found that those data can be used to predict students' learning performance at the end of the semester (Lu et al., 2018).

Learning analytics for online learning have the capacity to be an instrument for evaluating the quality of teaching based on the data of students' individual feedback. Moreover, big data are used for improving students' distance learning and supporting teacher and educational staff. For instance, big data contribute to enhancing students' learning outcomes, making student-centred programs, improving students' learning experience, reducing student dropout rates, and enhancing the learning guidance (Baldassarre, 2016). Educational big data utilizing the logs of the learning process in learning management systems has resulted in the improvement of students' learning, and teachers' ability to learn design in available learning environments (Hwang, Spikol, & Li, 2018). Learning analytics is the use of big data for academic purposes for better understanding the learners and their online learning environment. It can assist teachers in formulating teaching activities to suit the needs of learners, and better allocate the available resources to further improve the quality and value of students' learning experience (Scott & Nichols, 2017; Siemens, 2013).

Machine Learning

Machine learning procedures are mainly tools designed for generalizing from a given data set. They can solve multidisciplinary problems: such as the essential problem of machine learning applications involves understanding the classification rule from labelled data. Training data is considered as representative of future data (Blum, Hopcroft, & Kannan, 2016). Machine learning is one of the most important topics that drives Big Data. It is the ability to learn from data and to present a visual perspective of data in order to understand and predict the future. It is built on a statistical foundation, on which statistical data may be analysed in order to predict the direction that things might be taking. (L'heureux, Grolinger, Elyamany, & Capretz, 2017).

A machine learning framework was used to predict student dropout rates in Massive Open Online Courses. They were exclusively determined from clickstream data and forum data (Kloft, Stiehler, Zheng, & Pinkwart, 2014). A new theoretical framework was also developed with interleaving Bayesian models and the deep learning model to predict the dropout (Gal & Ghahramani, 2016).

Online Education (MOOCs & Blended Learning)

Massive Open Online Courses (MOOCs) are college or university courses offered in online platforms. The teaching is transmitted via video, and books are converted from texts into a digital file. There are also quizzes and exercises that can be taught through the website. In addition, students can communicate either with their peers or teachers through discussion boards. Scoring is also provided by the design of an automated assessment with quite strict rules. All of the elements in MOOCs rely heavily on the transformation of learning and teaching data into visualization, which include three steps: the separation of the source data, the creating or combining variables, and translation. Each step requires an expert in each area. Some hypotheses and heuristics have been established to transform and figure data in order to produce compelling and instructive visualization. For the MOOC platform, users can choose which course they want to visualize their functionality. It can be compared with which activities have a high percentage of usage and which have a low percentage of usage. In addition, users can download the learning files for review. They also can see which courses are new courses that have just been opened for admission (O'Reilly & Veeramachaneni, 2014). The visualization of Big Data shows learners' behaviours and learners' progress is subject to change. In addition, it enables teachers to prepare and utilize resources for teaching and learning more efficiently, as well as to develop learners' study habits and create satisfaction for learners (Xu & Ruan, 2018).

Blended learning is the combination of traditional learning, learning in the classroom, and online learning at all levels of education, especially in higher education. It is increasingly implemented by the integration of technology, academic standards, and organizational innovation. This is done to enhance learning and learning outcomes of students in more effective ways (Castro, 2019). The environment of blended learning involves both face-to-face teaching and online technology in many different formats. Effective use of learning analytics on students can provide learning information and track learners' comprehension success. It immediately shows which students are at risk of under-performing. In addition, students' persistence can be checked on a regular basis, along with the development of project planning and strategies for supporting learners (Picciano, 2014). In blended learning, information must be taken from both the digital and the physical sectors to give a complete picture of the whole teaching and learning process. The use of learning analytics facilitates a better understanding of the blended learning experience.

Areas covered include situations of suitability, correctness, relevance, and workability. The teacher can look at the monitoring report of each learning activity to evaluate students' progress in collaboration or participation in each workgroup, and keep up to date with resources to be used by individuals or groups (Rodríguez-Triana, Prieto, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2018). Furthermore, the visualization of learning analytics encourages reasonable and timely learning design. Learning analytics also reviews pedagogical considerations and relevant teachers' needs (Kaliisa, Kluge, & Mørch, 2020).

Data Science in Online Education

The data science with face-to-face learning is supported by the virtual learning environment (VLE) as an option to blended learning. It was found that students had different levels of skill and learning experiences (Musabirov, Pozdniakov, & Tenisheva, 2019). In American universities, there are undergraduate degrees in data analytics and data science with some similarities and some differences. Problems associated with big data can be solved by data analytics focusing on larger data sources and the consideration of statistics for machine learning and data mining. Data science is able to support the extraction of information and knowledge from data, not only with the skills possessed by data analytics but with analytical applications as well. (Aasheim et al., 2015).

Two extensive perspectives on the role of a 'data scientist consist of a "practical perspective" where data scientists are shown as business analysts that gain insights from scores of multi-dimensional sets of various data; and a "technical perspective" where data scientists are viewed as specialists in advanced computational instruments, data mining procedures, statistical analysis, and machine learning (Asamoah, Doran, & Schiller, 2015).

An increasing number of data science online courses are now available, especially at open universities in a Massive Open Online Course (MOOC) format. There is also more data science training in each university (Cao, 2017). Big data and cloud computing technology can ensure that all necessary information is available and give rise to innovative teaching techniques of English in educational institutions that can create more usable perspectives. Some examples of Big Data's applications include openness to reading materials written in English and the realization that big data technology can deliver speedy services, cost less, and the capacity to provide more services than a library. In addition, experts in foreign countries can have face-to-face communication with each other online. These services use real-time communication systems and rely on big data and cloud computing technology. Furthermore, teachers and students have access to the results and recent developments in research related to fields of interest via the latest innovations adopted by the Internet (Huang & Jin, 2018).

Nowadays, online learning in the form of MOOC is quickly gaining in popularity. However, there are some educational issues (e.g., dropout rates) that should be taken into account when helping students achieve more success in learning. A solution has been created for these problems by using Educational Data Science (EDS) to deal with the large amount of data online learning requires. It deals with matters related to the students' interaction with educational platforms and their engagement with courses offered online. In addition, the big data all learners interact with are a very valuable source of information for the application of EDS in MOOCs (Romero & Ventura, 2017). The primary objective of EDS is the application of data science to educational data, in the process of which academic functions are improved (P. Piety, Behrens, & Pea, 2013). Researchers view EDS as comprising four groups, each dedicated to their particular viewpoints of research work. This EDS framework has been developed for use in all educational contexts that use digital information to inform learners and manage learning. Firstly, Learning Analytics (LA)/Educational Data Mining (EDM) is used to determine activities at the micro level, or level of "learning", to describe which teaching methods and relevant academic fields are most likely to influence the promotion of learning specific content to specific learners. Secondly, Learner Analytics/Personalization (LA/P) and Educational Recommender Systems (ERS) works at the macro level, or "learner" level, to discover how differences between learners influence their resolve to learn, and how to increase overall study success from the use of adaptive techniques, modifications to suit each person, and giving advice. Thirdly, Academic/Institutional Analytics (AA) has been known to examine research institutes with an aim to focus on the institution rather than the learning process, or details related to various learning issues that take place. Lastly, Systemic/Instructional Improvement (S/II) is known for Data Driven Decision Making that makes use of data from test-based systems and identifies longitudinal data systems, mainly to assess teaching methods (P. J. Piety, Hickey, & Bishop, 2014).

The primary objective in pedagogy and learning development is to focus on students' growth along with the teacher's in-depth understanding about the learning dynamics, motivation, and learning-reacted activities of the pupils. For this reason, teachers need to build a rapport with their students, keep apprised of their concerns, and analyse the content that will be taught to suit the learners' goals. Teachers should learn to understand students' perspectives, look at the problems of the students, and analyse students' ways of thinking. In addition, teachers should design activities with psychology in mind, and take into account the personal characteristics of the students. These are the areas that delineate the teaching objectives most relevant to the students' needs. These are the fundamentals that lead to teaching innovations through research on Big Data's environment and artificial intelligence (Gao, 2020).

It has been demonstrated that the process of doing business in conjunction with Big Data can lead to innovative teaching in institutes of higher education. Students can generate information from a variety of sources, including learning management systems, social networks and mobile phones. Concurrently, educational institutions may obtain the information on students from various information systems, including academic data, financial data and library usage. These data sources will be analysed and stored as students' profiles based on their abilities and behaviour. By analysing the Big Data, it is possible to predict how students will progress in their studies, which subjects they like to study and what books they will borrow. These information sources may be used as the basis for sending alert warning messages to students to remind them to keep participating in required study projects. This all designed to meet the needs of the students themselves (Huda et al., 2016).

According to another study of Big Data's specialized features regarding learners' behaviours, the design of the MOOC course divides learners into five different groups in order to achieve different learning goals and attend to varying students' needs. This is done by analysing the patterns of learning access and performance for each student group; and then by monitoring key insights into the learning behaviour of each group of students. Some MOOC students do not place a high priority on their grades, but instead pay more attention to the value of the content. Commonly, a certain percentage of learners are unable to complete the exam and withdraw. Moreover, a small portion of students in any class perform poorly in their assigned activities. Therefore, it cannot be concluded that people who withdraw, or do not continue to study, are necessarily incapable of successfully learning basic MOOC contents (Douglas, Bermel, Alam, & Madhavan, 2016).

Conclusion

The major benefits of data science for educational institutions include improving resource allocation, improving the management of expected student accomplishments, reducing dropout rates, increasing business process efficiency, and financial and budget forecasting. Moreover, researchers consider the growth of Big Data in education correlates with the value learning brings to discovering additional critical concerns about using educational data with data ownership, ethics, privacy, and access (Daniel, 2019). In combining features of Data Science with Big Data to help improve teaching and learning, it is now possible to identify new phenomena experienced by today’s students. With this information, innovative teaching and learning methods can be adopted as part of an effort to stimulate interest among learners to examine new perspectives, including the introduction of technology with appropriate learning environments, all of which provides learners with a greater learning experience. In the online environment of MOOCs and blended learning, big data and learning analytics are implemented to analyse the pace of students’ learning and track their learning progress efficiently. The main activities consist of identifying students who are in unusual study situations, regular and continuous examination of learners' absorption and retention progress, planning learning projects for students, and creating strategies for supporting learners in various ways (Lu et al., 2018; Picciano, 2014).

It is widely acknowledged that developing tools to assist in data analytics or learning analytics can help teachers better understand interaction oriented data that is useful to online learners. The visualization of information clearly and effectively is able to show the learning behaviours and characteristics of the learners. This encompasses collaborative learning: the interaction between learners, and between learners and teachers in order to enable teachers to adjust learning activities or create appropriate teaching channels. All this encourages students to commit to serious study, and studying continuously until they are able to graduate as intended from the beginning.

A consensus reached by a comprehensive examination of related literature, finds that with online teaching and learning today Big Data technology is applied to meet the needs of most students, and supports students' learning in both offline and online interactions. Teaching materials are available in digital form including self-testing. Big Data technology can also analyse students' learning, enabling teachers' to supply up-to-date instruction, match the teaching materials with students' specified learning outcomes, and exchange modern content through online channels. Based on my experience of many years as a teacher and scholar in the field of Information Science and Information Technology, I strongly believe that students will be inspired to continue online learning that benefits greatly by interactions with their peers and teachers. New relationships may arise from the learning activities that require students to work together. Moreover, many channels for questioning are now available either by conversation in real time or making inquiries by sending messages via email or social media. The information obtained from learning analysis and the monitoring of learners' behaviour will be an important element in improving teaching content and methods. Now appropriate interactions developed for learning activities and online teaching plans can be adjusted to keep pace current trends. With the proper implementation of these innovations, the number of students dropping out along the way can likely be reduced significantly.

References

Aasheim, C. L., Williams, S., Rutner, P., & Gardiner, A. (2015). Data analytics vs. data science: A study of similarities and differences in undergraduate programs based on course descriptions. **Journal of Information Systems Education**, 26(2), 103-115.

Anderson, C., Wiley, Anderson, C., & Swanstrom, R. (2015). **Data Science for Dummies.** Somerset, UNITED STATES: John Wiley & Sons, Incorporated.

Asamoah, D., Doran, D., & Schiller, S. (2015). **Teaching the foundations of data science: an interdisciplinary approach.** arXiv preprint arXiv:1512.04456.

Baldassarre, M. (2016). Think big: learning contexts, algorithms and data science. **Research on Education and Media**, (2), 69. doi:10.1515/rem-2016-0020

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Stodden, V. (2018). Realizing the potential of data science. **Communications of the ACM**, 61(4), 67-72.

Blum, A., Hopcroft, J., & Kannan, R. (2016). **Foundations of data science**. Vorabversion eines Lehrbuchs.

Cao, L. (2017). Data science: a comprehensive overview. **ACM Computing Surveys (CSUR)**, 50(3), 1-42.

Cao, L. (2019). Data Science: Profession and Education. **IEEE Intelligent Systems, Intelligent Systems, IEEE, IEEE Intell. Syst.**, 34(5), 35-44. doi:10.1109/MIS.2019.2936705

Castro, R. (2019). Blended learning in higher education: Trends and capabilities. **Education and Information Technologies**, 24(4), 2523-2546.

Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. **British Journal of Educational Technology**, 50(1), 101-113.

Davenport, T. H., & Patil, D. (2012). Data scientist. **Harvard business review**, 90(5), 70-76.

Dhar, V. (2012). Data science and prediction. **Communications of the ACM.**, 56(12), 64-73.

Douglas, K. A., Bermel, P., Alam, M. M., & Madhavan, K. (2016). Big data characterization of learner behaviour in a highly technical MOOC engineering course. **Journal of Learning Analytics**, 3(3), 170-192.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. **International Journal of Technology Enhanced Learning**, 4(5-6), 304-317.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. Paper presented at **the international conference on machine learning.**

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. **International journal of information management**, 35(2), 137-144.

Gao, S. (2020). Innovative Teaching of Integration of Artificial Intelligence and University Mathematics in Big Data Environment. **MS&E**, 750(1), 012137.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. **Academy of Management Journal**, 59 (5), 1493-1507.

Grover, P., & Kar, A. K. (2017). Big data analytics: a review on theoretical contributions and tools used in literature. **Global Journal of Flexible Systems Management**, 18(3), 203-229.

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: a review for the future. **International journal of information management**, 42, 78-89.

Huang, Y., & Jin, X. (2018). Innovative college English teaching modes based on Big Data. **Educational Sciences: Theory & Practice**, 18(6).

Huda, M., Anshari, M., Almunawar, M. N., Shahrill, M., Tan, A., Jaidin, J. H., Masri, M. (2016). **Innovative teaching in higher education: the big data approach**. TOJET.

Hwang, G.-J., Spikol, D., & Li, K.-C. (2018). Guest Editorial: Trends and Research Issues of Learning Analytics and Educational Big Data. **Journal of Educational Technology & Society**, 21(2), 134.

Kaliisa, R., Kluge, A., & Mørch, A. I. (2020). Combining checkpoint and process learning analytics to support learning design decisions in blended learning environments. **Journal of Learning Analytics**, 7(3), 33-47.

KlaA nja-Milicevic, A., Ivanovic, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. **Computer Applications in Engineering Education**, (6), 1066. doi:10.1002/cae.21844

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. Paper presented at the **Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs.**

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. **IEEE Access**, 5, 7776-7797.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. **Journal of Educational Technology & Society**, 21(2), 220-232.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. **Computers & Education**, 53(3), 950-965.

Musabirov, I. i. h. r., Pozdniakov, S., & Tenisheva, K. (2019). Predictors of Academic Achievement in Blended Learning: The Case of Data Science Minor. **International Journal of Emerging Technologies in Learning,** 14(5), 64-74. doi:10.3991/ijet.v14i05.9512

O'Reilly, U.-M., & Veeramachaneni, K. (2014). Technology for mining the big data of moocs. **Research & Practice in Assessment**, 9, 29-37.

Parks, D. M. D. (2017). Defining Data Science and Data Scientist: University of South Florida.

Picciano, A. G. (2014). Big data and learning analytics in blended learning environments: Benefits and concerns. **IJIMAI**, 2(7), 35-43.

Piety, P., Behrens, J., & Pea, R. (2013). Educational data sciences and the need for interpretive skills. American Educational Research Association, 27.

Piety, P. J., Hickey, D. T., & Bishop, M. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. Paper presented at the **Proceedings of the fourth international conference on learning analytics and knowledge**.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. **Big data**, 1(1), 51-59.

Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018). The teacher in the loop: Customizing multimodal learning analytics for blended learning. Paper presented at the **Proceedings of the 8th international conference on learning analytics and knowledge.**

Romero, C. c. u. e., & Ventura, S. (2017). Educational data science in massive open online courses. **WIREs: Data Mining & Knowledge Discovery**, 7(1), n/a-N.PAG. doi:10.1002/widm.1187

Saltz, J. S., & Stanton, J. M. (2017). **An introduction to data science**. SAGE Publications.

Schleder, G. R., Padilha, A. C., Acosta, C. M., Costa, M., & Fazzio, A. (2019). From DFT to machine learning: recent approaches to materials science–a review. **Journal of Physics: Materials**, 2(3), 032001.

Scott, J., & Nichols, T. P. (2017). Learning analytics as assemblage: Criticality and contingency in online education**. Research in Education,** 98(1), 83-105.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. **American Behavioral Scientist,** 57(10), 1380-1400.

Xu, N., & Ruan, B. (2018). An application of big data learning analysis based on MOOC platform. Paper presented at the **2018 9th International Conference on Information Technology in Medicine and Education (ITME).**