Towards intelligent E-learning systems

Mengchi Liu[1] · **Dongmei Yu[2]**

Abstract

The prevalence of e-learning systems has made educational resources more accessible, interactive and effective to learners without the geographic and temporal boundaries. However, as the number of users increases and the volume of data grows, current e-learning systems face some technical and pedagogical challenges. This paper provides a comprehensive review on the efforts of applying new information and communication technologies to improve e-learning services. We first systematically investigate current e-learning systems in terms of their classification, architecture, functions, challenges, and current trends. We then present a general architecture for big data based e-learning systems to meet the ever-growing demand for e-learning. We also describe how to use data generated in big data based e-learning systems to support more flexible and customized course delivery and personalized learning.

Keywords E-learning system · Big data · Smart learning · Personalized earning

1 Introduction

The fast development of e-learning systems has radically transformed the way in which learning resources are imparted to students. They make educational resources more accessible, interactive and effective to learners without the geographic and temporal boundaries. E-learning (Alonso et al., 2005) has been defined as the use of information and communication technologies to improve the quality of learning by

✉ Dongmei Yu
dongmeiyu@shao.ac.cn

Mengchi Liu
liumengchi@scnu.edu.cn

1   Guangzhou Key Laboratory of Big Data and Intelligent Education School of Computer Science South China Normal University Guangzhou, Guangdong, 510631, China

2   Shanghai Astronomical Observatory Chinese Academy of Sciences, Shanghai, 200030, China

enabling access to resources and services, as well as remote exchange and collaboration. Some synonymous terms including open learning, distance learning, web-based education and online learning have been alternatively used over the past decades. In general, it is considered as an educational process that enables transfer of knowledge and skills flexibly to a large number of recipients at various times and locations. The combination of education and technologies provides a new way for people to learn in the era of information and communicationtechnology.

However, it has become challenging to provide excellent e-learning services as modern e-learning applications are increasingly data-intensive due to the following reasons:

1. the number of learners and courseworks increases dramatically as e-learning applications get more and more popular, especially during the covid pandemic;
2. different roles generate a huge amount of interactive information when posting or exchanging messages;
3. the diversity of resources make each type of information isolated;
4. a variety of personal information and sensitive data need related access control and security policies;
5. the gigantic amounts of data need to be stored and managed properly.

As a result, e-learning systems need to evolve to provide smart services. In this context, intelligent technologies have been gradually used to collect, preprocess, analyze, store, and visualize huge amount of data from various learning sources. They are utilized to eliminate noise and extract valuable information to improve the effectiveness of e-learning. Also, they enable learning resources to be tailored for each individual learner according to the contents and learners' interactive behaviors (Gamalel-Din, 2010; Kumar et al., 2016; Hu et al., 2020). In other words, precise customization and personalization of knowledge or services should be provided to each individual learner accordingly. Consequently, smart e-learning systems make it possible for educational organizations to offer improved teaching and learning deliveries, thus deal with the challenges in current e-learning services.

This paper mainly reviews the studies that apply various technologies to e-learning systems and hence provide personalized and precise teaching and learning services. We first systematically investigate the status of current e-learning systems in terms of their features, classifications, architecture, challenges and trends. Then we present big data based e-learning systems for more flexible course delivery and personalized learning, and also show how data can be processed to facilitate learning and teaching. Finally we discuss the beneficial effects of integrating big data technology with e-learning systems.

2 E-learning system classifications

E-learning systems facilitate the planning, management, and delivery of content for e-Learning. Based on the target users and the cost, we classify current e-learning systems into two kinds: Massive Open Online Course (MOOC) platforms and Learning Management Systems (LMSs). Firstly, MOOC platforms are open to a large number

of individuals who are intended to learn. Even though some courses are produced by certain universities, they are not limited to student in post-secondary institutions. They can be accessed by people regardless of their location, culture, nationality, and any other criteria. On the other hand, LMSs are usually implemented for post-secondary institutions. Thus, they are not by default open to the general public, only a certain group of people can have access to it. Secondly, most MOOCs are free of cost or cost little so that individuals can afford, while the cost for LMS is higher based on the number of users and usually borne by post-secondary institutions rather than individuals.

2.1 Massive open online courses (MOOCs) platforms

MOOCs are online courses open to unlimited participants that are offered by many universities and institutions on the web. The term was firstly coined in 2008 by Canadian researchers Dave Cormier and Brian Alexander (Goldie, 2016). In fall of 2011, Stanford University offered the first MOOC course, which was originally registered by more than 160,000 learners from all over the world, and eventually about 20,000 of them completed it. Then, three significant MOOCs platforms Udacity[1], edX[2], and Coursera[3] were developed in 2012 and used to offer MOOCs for free. Since then, the number of available MOOCs and MOOC learners increased dramatically, with more online platforms available. By the end of 2020, about 950 universities around the world launched 16,300 MOOCs with 180 million MOOC learners (Shah, 2020).

Generally, there are two kinds of MOOCs based on different learning theories: cMOOCs (the connectivist MOOCs) and xMOOCs (extended MOOCs) (Alonso et al., 2005). cMOOCs mainly emphasize connection and promote interaction by digital tools like blogs, learning communities and social media platform. Learners can also create and generate knowledge by themselves. xMOOCs are based on traditional university courses by focusing more on the content and deliver knowledge in small units so that the number of students can be increased significantly. Currently, the most popular and influential MOOCs providers are Coursera, edX and Udacity.

In addition, various national online platforms have emerged in a number of countries (Shah, 2021), including FutureLearn in Great Britain[4], XuetangX in China[5], France Université Numérique (FUN) in France[6], OpenHPI in German[7], EduOpen in Italy[8], SWAYAM in India[9], gacco in Japan[10], ThaiMOOC in Thailand[11], the National

---

[1] https://www.udecity.com

[2] https://www.edx.org

[3] https://www.coursera.org

[4] https://www.futurelearn.com

[5] https://www.xuetangx.com

[6] https://www.fun-mooc.fr

[7] https://open.hpi.de

[8] https://learn.eduopen.org

[9] https://swayam.gov.in

[10] https://gacco.org

[11] https://thaimooc.org

Open Education Platform (NOEP) in Russia[12], etc. In summary, these MOOCs are open, participatory and distributed (Baturay, 2015). They have the potential to disrupt the traditional education due to their easy accessibility and free or low-cost content delivery, especially considering educational credentials including micro-credentials, specializations or degrees from accredited institutions (Pickard et al., 2018).

2.2 Learning management systems (LMSs)

LMSs are e-learning systems for hosting, assigning, managing, reporting and evaluating e-learning courses. Many postsecondary education institutions adopt LMSs as critical educational tools to support course management and to foster interaction among students, teachers and content resources. They are also used to identify training and learning gaps, implementing a wide range of pedagogical methods to promote education process. LMSs are generally classified as commercial and non-commercial systems. Commercial LMSs like WebCT, Blackboard, D2L Brightspace have been frequently and very successfully used in the past decades. They are basically easy to deploy and use, and technical support services are provided without additional costs. However, some non-commercial LMSs such as Moodle[13], Canvas[14], Open edX[15] and Sakai [16] also become popular recently. The open source feature of non-commercial LMSs makes them attractive since they are easy to obtain, as many are free, especially those that provide a basic level of service. They also provide more flexible and scalable architecture to meet users' needs. Generally, the most successful LMSs in North America are the Big Four (Hill, 2019): Blackboard, Canvas, D2L Brightspace and Moodle.

3 Architecture of current E-learning systems

Many conventional frameworks have been used to create and improve e-learning system effectiveness. One is based on service-oriented architectures (SOA) that allow to easily extend the capabilities and functionalities of the system by dynamically adding services. For example, Fajar et al. (2018) present a SOA system architecture and reference for an e-learning system, which consists of six components: data layer, resource layer, application layer, business process layer, presentation layer and governance layer. The data of each layer are treated as the service in the SOA system, which makes it more reusable, flexible and accessible to extended tools. This method allows business and wider society to improve e-learning and offer affordable education. Similarly, (González et al., 2009) extend existing e-learning systems

---

[12]https://openedu.ru/university/hse

[13]https://moodle.org

[14]https://www.instructure.com

[15]https://open.edx.org

[16]https://www.sakailms.org

to external mobile scenarios based on SOA as well. The architecture ensures the independence of e-learning systems, mobile applications and external applications, and provides a reliable data exchange and interoperability between them. Furthermore, (Kappe & Scerbakov, 2017) present an innovative object-oriented architecture for the implementation of e-learning systems on a single software platform to meet the requirements of various e-learning scenarios. Abstract data objects (ADOs) that encapsulate private memory together with some methods are widely used as the main components of an functional objects like courses, announcements, curriculum and so on. The implementation shows that this architecture is highly modular since documents and objects can be created independently but also re-used through a flexible nesting or containment mechanism.

Recently, the availability of high speed networks, low-cost computers and storage devices has resulted in the significant advances in the cloud computing technology, which is the on-demand usage of a network of remote servers hosted on the internet to store, manage, and process data, rather than on one or more local servers. (Riahi, 2015) reviewed recent cloud-based systems and proposed an e-learning cloud architecture, which includes hardware resource layer, software resource layer, resources management layer, service layer and business application layer. They also conclude several advantages of cloud based e-learning like low cost, improved performance and compatibility, information security and benefits for both students and students. Other researches (El Mhouti et al., 2018; Masud & Huang, 2012; Riahi, 2015; Sun et al., 2015; Chao et al., 2015; Hendradi et al., 2020; Rani et al., 2015) also focus on combining e-learning systems with cloud computing. There are two advantages in doing so. Firstly, it is easy to create and maintain, and the investment cost can be reduced significantly using the pay-as-you-go method. Also, it allows to scale the services according to the need. (Sun et al., 2015) introduce a cloud-based virtual learning environment called MLaaS, which aims to provide adaptive micro learning contents and customized learning route for every single learner. Education data mining scheme is used to discover features of learning resources and understand learners' behaviors. In addition, (Chao et al., 2015) propose a cloud-based ecosystem called CLEM for teachers and learners. Their implementation shows that the cloud-based platform gathers heterogeneous and distributed devices in a common pool that makes computational resources more accessible and sharable. Furthermore, (Jeong et al., 2013) introduce a private-cloud-based e-learning system with six components: a private cloud platform, an XML based common file format, an authoring tool, a content viewer, an inference engine and a security system. By using these components, it can deliver and share various types of educational resources effectively. However, according to some literatures (El Mhouti et al., 2018; Laisheng & Zhengxia, 2011), the challenges of cloud-based e-learning system are mainly related to cloud privacy, security and confidence. At the same time, these concerns also provide opportunities for e-learning promotion and development in cloud computing environment.

Based on previous researches, we conclude a general framework for current e-learning systems in Fig. 1, which basically consists of three logical layers contributing to better teaching and learning effectiveness: presentation layer, e-learning system layer and data layer.
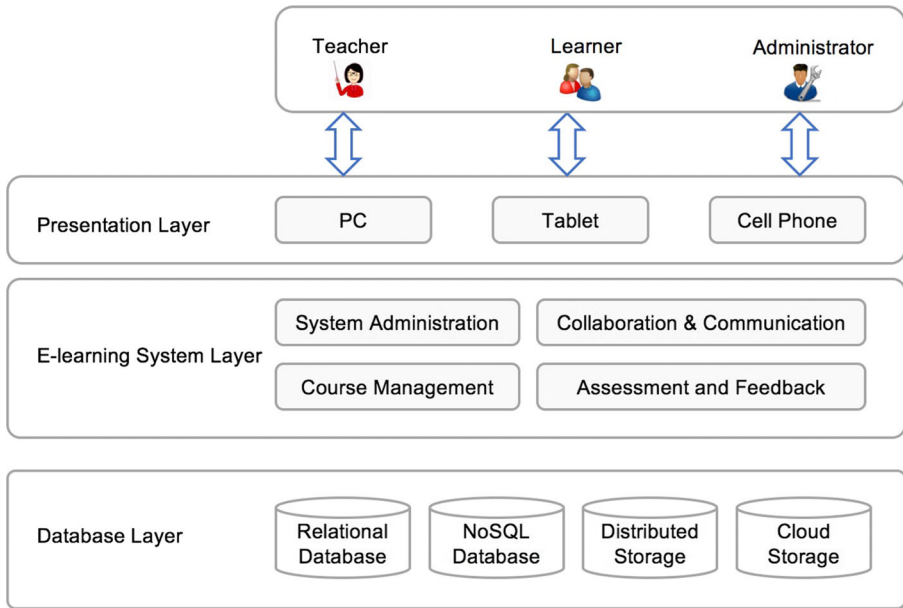
Fig. 1 A General Framework for E-learning Systems

3.1 The presentation layer

The presentation layer focuses on the human computer interaction by providing accessible user interface and learning resources to end users. It aims at improving the usability, accessibility, credibility and the user experience of the learning ecosystems (García-Holgado & García-Peñalvo, 2018). Firstly, it provides a unified interface for all the services or functionalities provided by the lower layers and hide system complexity from users. Users can utilize this interface to construct and control the contents of e-learning systems. The feedback from the e-learning system layer is delivered through this interface. Secondly, due to the prevalence of various mobile devices (e.g. mobile phones, laptops, tablets and other portable devices), the presentation layer should ensure that e-learning systems support mobile learning paradigm (Schuck et al., 2017). In other words, e-learning systems are adaptable to distinct screen sizes, which allows learners to gain any information flexibly. Also, mobile learning has been proved to be able to improve student participation and engagement during learning process, while learners have high levels of motivation and satisfaction (Cheng et al., 2015). It also has a positive influence on learners' academic performance (Han & Shin, 2016). Therefore, it is necessary to use proper front-end techniques such as HTML, XHTML, CSS, and JavaScripts to support mobile learning, in which presented pages can be rendered properly on a browser to meet the compatibility requirements of devices.

3.2 The E-learning system layer

The e-learning system layer aims at synthesizing educational resources by way of various functions such as course enrolment and management, user profile and activities, teaching or learning assessment and feedback, user communication or collaboration and so forth. It can also be an integration of related components which support instructional model or learning model (Lu et al., 2015). Users are able to choose the components to satisfy the different needs for teaching and learning. For most MOOCs and LMSs, this layer plays an important role between the presentation layer and the database layer. Learning and teaching information including users profile, learning resources, teaching and learning activities is collected and passed through e-learning system layer. It is also a teaching and learning platform that enable each learner to access specific education resources flexibly.

3.3 The database layer

The database layer hosts data generated by using e-learning systems. It is the critical place where education data is collected, stored and used. Due to the individual differences, collecting the massive data and retaining the diversity and dynamic features is very important. Additionally, all the collected data need to be stored until their use. Alternatively, some processed results may be put to use immediately, while most of them will serve some purposes later on. The main benefit is that it enables the collected or processed data to be accessed and retrieved easily.

Usually, existing solutions for e-leaning storage mainly rely on relational database, such as Mysql (Wangmo & Ivanova, 2017) and Oracle (Datta & Bhattacharyya, 2018). Moodle's database is typically MySQL or Postgres, and can also be Microsoft SQL Server or Oracle. Sakai and Blackboard can be deployed in a SQL or Oracle environment as well.

Also, NoSQL databases are increasingly used for large sets of distributed data due to flexible and scale-out architecture. They work as a complementary technology for the relational databases system and are suitable for distributed applications with the demand of high data scalability and availability (Davoudian et al., 2018). For example, MongoDB is choosed by Open edX for storing large files which are text files, PDFs, audio/video clips, etc.

Additionally, distributed storage technology is increasingly used to replace traditional local storage. Some run on top of file systems while others work as standalone systems. For example, (Zhang et al., 2020) use distributed storage technologies for experimental education systems. Specifically, the interplanetary file system (IPFS), an external storage server and external cloud storage are combined for storage management. Among them, IPFS determines the overall performance of the storage module and contributes to system reliability and flexibility. Additionally, a file table is defined to manage each learning content such as documents, video and problem books in a distributed database. (Kawato et al., 2020) create an e-learning system by implementing Apache Cassandra, which is an open-source distributed database

system to handle large volumes of data. By combining distributed hash tables (DHT), which hold information of the connected computer nodes, it is able to share various education resources spanning multiple servers. (Otoo-Arthur & van Zyl, 2020) present a framework on a distributed and parallel computing environment to provide new value to teaching and learning process.

Moreover, cloud storage as a large scale distributed storage paradigm is also used to education system, in which learning resources is stored on remote storage systems. Compared with traditional storage ways, it has many advantages in terms of scalability, flexibility, safety, ease of use and cost saving. (Sun et al., 2015) deploy Mobile MOOC learning on the Amazon EC2, and Amazon S3 is considered as the MOOC learner and course data storage because of its robustness and mature disaster recovery mechanisms. (Jeong et al., 2013) propose a content-oriented smart education system based on a small-scale, private cloud. A common file format based on XML are defined as a means of representing data and meta - data. The Document Type Definition (DTD) and the eXtensible Style sheet Language (XSL) are used to described the schema and styles for the XML document structure seperately, which enables the same content can be viewed on multiple devices. Furthermore, (Rani et al., 2015) deployed e-learning system on remote cloud host, where all required learning resources are stored. They also build a simple MySOL on the cloud host for authentication of the system. By doing so, an expanded and secure environment is built to raise e-learning system.

4 Functions of E-learning systems

The functions of e-learning systems depend on its potential usage such as system scale requirement, organizational objectives, online training strategy and desired pedagogical outcomes. (Cavus & Zabadi, 2014) summarize that standard LMSs should have various tools for e-learning systems. They compare six popular open source LMSs in terms of video services, discussion forum, file exchange, email, realtime chat and so forth, and discover that communication tools provided by Moodle and ATutor are efficient, but it is not easy to obtain information on Claroline and Sakai due to their complex webpages. Similarly, (Chung et al., 2013) suggest that LMSs should have five components: transmitting course content, creating a discussion, evaluating students, evaluating courses and instructors, and creating computer-based instruction. However, most of the existing e-learning systems do not contain all the features in a single system. So, we highlight the general function components (Fig. 2) that most e-learning systems have to support teaching and learning process.

4.1 System administration

This module includes a full range of functions for the management and configuration of system parameters and attributes in terms of users, courses grades, appearance reports and so forth. It covers components such as user authentication, user
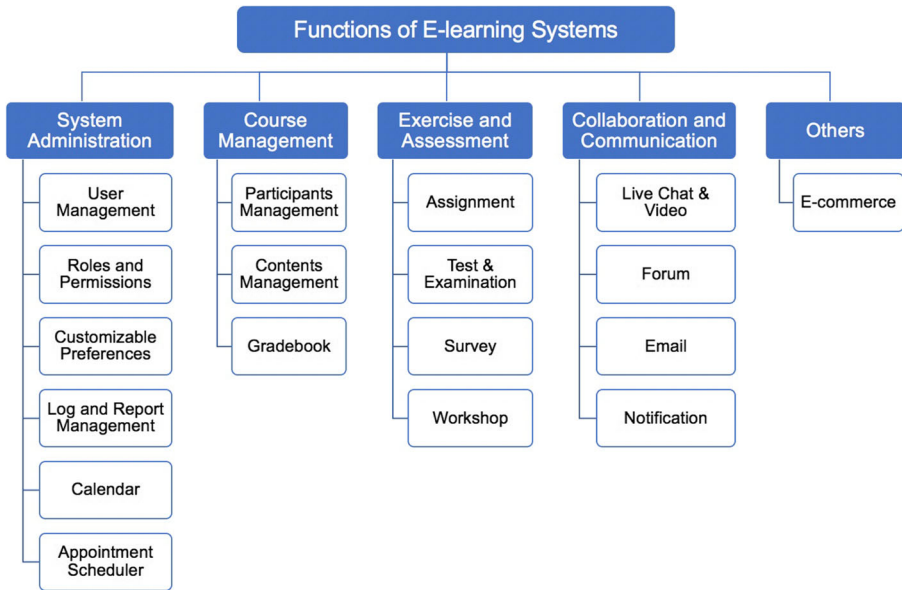
Fig. 2 Functions of E-learning Systems

management, roles and permission management, customizable preferences, log and report management, calendar and appointment scheduler.

- `User management`: For e-learning users, email or mobile phone-based self-registration authentication method is commonly adopted to fulfill user authentication. Also, a category hierarchy is usually built to organize users from different organizations. Common operations such as adding, deleting, modifying and querying related to user management are supported. Normally, there are several categories of e-learning users including administrators, instructors, teaching assistants (TAs) and learners. The administrators set up and configure the system. The instructors prepare the lessons and access the learners' progress. TAs assist instructors with instructional responsibilities. Learners are anyone interested in learning and being educated in the courses.

- `Roles and permissions`: E-learning systems usually support several standard user roles and has the potential to create an unlimited number of additional roles. Therefore, it is necessary to control users' access rights only to the information they need and to prevent them from accessing information that does not pertain to them. For example, role based access control (RBAC) is used to control client access and consents in DidaTec LMSs platform (Laura et al., 2018). Another example is that the Access Control List (ACL) is used to maintain the user information and their permissions, while group key is utilized to secure course materials and to ensure that only approved participants have access to it (Kanimozhi et al., 2019).

- `Customizable preferences`: Personalized setting for user profiles and system preferences such as privacy, design and layout of the websites is allowed to enhance users' experience.
- `Log and report management`: Event log analysis is displayed through graphical user interface to assist teaching and learning. Also, analysis reports are available and exported to help administrators and teachers to make decisions based on the statistical results.
- `Calendar`: It displays a consolidated view of all the course-related events by day, week, or month in the e-learning system. It allows users to view the available learning programs or courses with specific due dates. Also, the calendar usually automatically synchronizes with other teaching or learning activities such as syllabus, assignments, tests, and grades. In the case that users create, change, or delete the date of an activity in the LMS, the change will show up in calendar and vice versa. Finally, systems such as Moodle and Coursera also allow users to export calendar, so they may be imported into other calendar programs, as a backup or to create a copy. Conversely, other agendas can also be imported to calendars in LMSs to facilitate time management in some university (Mei, 2016).
- `Appointment scheduler`: It helps teachers schedule appointments with their students. Teachers specify time slots and locations for online or offline activities and students choose for their attendance.

4.2 Course management

It is a basic but most important component of an e-learning system (Cavus & Zabadi, 2014), which mainly refers to create, organize and deliver various coursework. Most LMSs allow users to add course material from various sources in different formats such as text, graphics, audio, video and so forth. Platforms like Moodle, Open edX allow to use the SCORM (Shareable Content Object Reference Model) standard for its online courses. The benefit is that it provides a standardized course model that supports the reusability of learning objects. For example, multiple individual lessons can be stringed together into a complete course. Participants are also encouraged to have more interactivity within e-learning systems. With the proper authoring tool, they can create their own courses and eliminate the need to outsource course development. Similarly, (Gamalel-Din, 2010) tailor course materials by drawing multimedia Learning Objects(LO) from LO repositories, which are composed of small granular multimedia objects. This idea helps teachers to find the best available assets and LOs for course design. Students are also able to get tailored learning strategy based on their abilities and previous knowledge. Basically, the specific functions of course management are as follows.

- `Participants management`: It enables an administrator or teacher to easily enrol, view, filter, edit and delete participants for each course, and also group participants or invite learners. It is also a centralized place where teachers are able to trace student's attendance, increase student enrollment and avoid high drop-out rates during the course. By comparing user activity and identifying attendance

trends, regular attendance of all students is recorded and ready for further analysis. Since student attendance is strongly linked to learning outcome, it is also necessary for teacher to give a warning to those with poor attendance during online learning.

- `Contents management`: Contents are organized in descriptive categories so that users can easily find their desired resources. Both static contents and interactive resources are delivered according to students needs. Some contents might be made either for a restricted audience or for a wider population, either as a free offering or as paid courses. Generally, a LMS allows course creators to freely structure their e-learning offerings in a manner that best fits their purposes and requirements. Also, instructors can trace the progress of each course and adjust their pedagogical strategy accordingly.
- `Gradebook`: It is a central location where teachers can manage grades for courses and track student activities relative to gradable items. It plays an critical role in performance monitoring and feedback seeking associated with self-paced learning practices.

4.3 Exercise and Assessment

This module utilizes some testing and evaluation capabilities to monitor, track and evaluate the effectiveness of the e-learning process. Most e-learning systems support learning assessments periodically and some of them even support the teachers to identify gaps or intervene when necessary. Generally, a broad range of e-learning assessment methods are considered in terms of learners' progress and performance. Some offer built-in auto-graded evaluation tools (Baturay, 2015), such as quizzes, tests, assignment, group exercises, examinations and surveys so that both instructors and students can track the learning performance in gradebook easily. Some even have diagnostic assessments to evaluate the level of knowledge of learners and assign suitable level to them. Furthermore, peer assessment (Lynda et al., 2017) is widely used in MOOC platforms which involves learners in grading and giving feedback from the work of their peers. It is also recognized as one important feature that affects the effectiveness of e-learning systems. For example, Coursera has regarded peer review as a scalable and sustainable way to guide students in assessing each other's job as well as providing feedback. Lastly, evaluation reports used to assess e-learning are generated to query and display data in graphs and charts, allowing users to easily spot teaching or learning trends or issues. Additionally, this report should reflect the user performance on both individual and group level from multiple perspectives and monitor if the learners achieve their required objectives. Generally, the following activities are normally used to perform assessment and feedback.

- `Assignment`: It allows learners to submit their work online and teachers to grade and give response. Teachers are allowed to select excellent assignments to share with all students enrolled in the courses.
- `Test and examination`: It is necessary to assess course quality and learning outcomes. Teachers are allowed to create quizzes that are made up of a wide range of questions derived from a question bank. This enables a question

to be re-used in different quizzes and facilitate the teaching process. Examinations are also conducted online to assess student performance. Furthermore, a remote proctor or students' webcam can be used to monitor the student's activities and the surroundings during the examination, which is an effective solution to maintain academic integrity for e-learning examination.

- `Survey`: It is used to help teachers to gather information from students and reflect on their own teaching. It is also used to identify certain trends that may be happening among course participants.
- `Workshop`: It allows the learners to perform peer assessment activity according to teacher's guidance.

Collaboration and communication Effective collaboration and communication help the transfer, sharing and co-construction of knowledge as well as the sharing of experiences in teacher-student and peer-to-peer relationships (Chiu & Hew, 2018). There are various communication tools in existing e-learning systems that encourage participants to support each other in the learning process.

- `Live chat & video`: Live chat is an instant messaging application that allows users to discuss in real time while they participate in the teaching and learning process. It usually supports features such as real time chat monitor, chat history, file sharing and so forth. A platform-independent and web-based instant messaging can be embedded to support convenient communication among users. The chat tool (Bagarukayo et al., 2014; Carmona et al., 2008) can be integrated as a synchronous, live communication way to aid interaction and collaboration among e-learning users. It allows students and teachers to interact in real time, such as conducting group discussions or study sessions effectively. Live video tools such as Microsoft Teams, Skype, Zoom and Tencent Meeting are increasing used as communication tools for e-learning especially during COVID-19 pandemic. According to (Alameri et al., 2020), 80.7% of participants agree that Moodle, Microsoft teams and Zoom platforms enhance the communication between teachers and students in higher education. Learners are easy to concentrate on classes by constructing visual presentations. Most students believe Moodle, Microsoft teams and Zoom platforms are critical for them to handle learning process and they will be an indispensable part for online learning.
- `Forum`: It provides space where students and teachers can discuss a specific topic or a group of topics to exchange their ideas. Three types of forums can be built: public forum, course forum and class forum. A three-level forum can contribute significantly to successful collaboration and community building in an online environment. Forums are usually integrated into the e-learning systems (Kakasevski et al., 2008; Baturay, 2015) . It helps learners to exchange their ideas and knowledge effectively so that they are not constrained in a passive role but can instead help each other and engage in active ways. Also, forums allow instructors to post course-related questions that can be accessed and discussed by learners. Then the extensions of questions and ideas for interaction are available regardless of whether the instructor is available. Furthermore, forums can also be split into several subforums to provide specific discussions. For example, Coursera

provides a default partition of subforums, which includes study groups, general discussions, lectures, assignments, logistics and feedback (Rossi & Gnawali, 2014). The instructors are also allowed to customize the subforums flexibly.

- `Email`: The traditional email has been widely used by Moodle, Blackboard and Open edX and other LMSs (Kakasevski et al., 2008; Bagarukayo et al., 2014). It supports instructors to send email to individual learner or a group of students in the course without launching a separate email program.
- `Notification`: It alerts users about events or activities update in the system.

4.4 Others

- `E-commerce`: E-commerce exists in some e-learning systems especially MOOC platforms. It provides sophisticated business transaction functionalities such as payment processing, shopping cart, and customer analytics capabilities. In MOOC platforms, to complete a course or learning module, users need to provide their user profiles and make a one-time payment or agree to a monthly subscription. To earn a degree or a certificate, users also need to pay tuition fees accordingly. E-commerce integrated LMSs usually allow learners to carry out all their transactions starting from registration to making the payment through a single portal, which in turn helps improve user experience.

5 Challenges of E-learning systems

E-learning systems have profoundly changed the traditional methods of teaching and learning by offering enhanced access to information and interactive resources at all levels of education. They are a supplementary offer to traditional education and to some extent have the ability to substitute it. Despite the advantages it offers, there are still some pedagogical and technical problems that need to be addressed. (Moubayed et al., 2018) analysize several challenges from different aspects, which includes transmission/delivery, personalization, enabling technologies, collaborative/cooperative learning facilitation, and evaluation & assessment. (Islam et al., 2015) also discuss some challenges existing in the success of e-learning, which are mainly related to technology, learning style, training and management. During the Covid-19 pandemic, e-learning faces more challenges as a massive adoption of online education. (Hamdan et al., 2020) analyze several challenges and obstacles including the lace of access to ICT tools, the adequate training for teachers using technological devices, the limited budget for digital devices and poor e-learning environment. (Oryakhail et al., 2021) investigate barriers that hinder the implementation of e-learning in Afghanistan Higher Education. Their research shows specific challenges related to students, lectures, infrastructure and university management.

One major concern for e-learning systems is to use new pedagogy and cognitive approaches to achieve efficient transmission and delivery of e-learning system resources. Since e-learning is quite different from face-to-face education, the courses have to be adapted more attractive or interactive for students, which could be a

challenge for teachers who have been used to traditional teaching. E-learning systems require a different approach to pedagogy instead of simply uploading large amounts of resources onto the e-learning systems. (Bari et al., 2018) state that there is no adequate design strategies adapted to the e-learning process and the evaluation of its success implementation. (Andersson, 2008; Moubayed et al., 2018) also discover that some hands-on courses conducted through face-to-face teaching can be difficult to carry out on e-learning systems so that the students cannot fully grasp the content as they learn from traditional classroom-based training. For example, practical lessons or laboratory work are difficult to be conducted on e-learning system (Karjo et al., 2021). Another major concern of e-learning is human resistance, which refers to lack of motivation for both students and teachers. For students, the lack of learning motivation and persistence has been research widely. Since e-learning is self-regulated learning, unmotivated learners may get behind without adequate supervision and guidance. Some statistics show that there is a high dropout rate on MOOCs platforms, which means the majority of students who signed up the course in the beginning could not finish in the end. For example, The Open University found out that only 6.5% of those enrolled students complete the course (Jordan, 2014). The number of enrolments decreases over time and is strongly linked with the duration of the course. As a result, e-learning ends up with a high dropout rates and low effectiveness. To overcome this hurdle, it's important to stimulate the deep motivations that drive the learners to study or induce them to drop out based on data analysis methods. Also, some kind of useful interventions like self-regulated learning can be delivered to potentially prevent learners' dropout behavior (Min & Nasir, 2020). Based on (Hapsari et al., 2021), the heavy workloads and more time requirement for teachers have been a challenge affecting the adoption of e-learning. E-learning acceptability is important to the success of e-learning (Hapsari et al., 2021). If teachers have confidence in e-learning and are willing to master both technical and conceptual issues, it will be easy to achieve e-learning success.

Also, with the rapid development of technologies, e-learning systems grow dramatically in terms of the services offered and the available contents generated. Therefore, ensuring that the e-learning system has the means to adapt to the evolving scalability and robustness needs is particularly crucial. According to (Hapsari et al., 2021; Oryakhail et al., 2021; Hamdan et al., 2020; Karjo et al., 2021), lack of reliable internet connection has been a major barrier among e-learning challenges. Also, the lack of infrastructure capability is a common problem faced by both teachers and students in developing countries. If e-learning infrastructure fail to handle requests from thousands of users simultaneously, the system timeout or latency will definitely lead to the interruption of e-learning. Thus, we need to consider how to optimize various hardware or software resources to meet the storage and network requirements as well (El Mhouti et al., 2018). Technologies like cloud computing (Riahi, 2015; El Mhouti et al., 2018; Masud & Huang, 2012; Sun et al., 2015; Chao et al., 2015) have been introduced to provide efficient scalable architecture for e-learning systems.

Moreover, discovering useful information that can be utilized to help teachers determine proper pedagogical strategies and achieve better learning outcomes is also

difficult in an e-learning environment (Islam et al., 2015). However, using big data based statistical and mathematical procedure to identify and extract valuable knowledge from large data source is a feasible solution to solve the problem related to "information overload" (Brajkovic et al., 2018).

Furthermore, compared with traditional classroom, it is quite difficult for teachers to track or monitor student progress in e-learning system. AI provides a solution to this problem (Klašnja-Milićević & Ivanović, 2021). It allows teachers to monitor or assess student progress timely. If there is a problem with student performance, AI can be used to alert teachers and assist students based on their strengths and weaknesses.

Lastly, several social challenges faced by e-learning cannot be ignored. Firstly, the cost for e-learning is an issue. (Hamdan et al., 2020) find financial cost for students from low-income families might prevent them from online education. Students need more financial support to purchase computers or stable internet connection services. Secondly, cyber security and privacy is another social challenge facing e-learning. For example, live video applications like Zoom and Microsoft Teams have end-to-end encryption for videos or calls, which ensures the content is encrypted before it's sent and decrypted only by the intended recipient. However, for most e-learning systems, cyber security and privacy is an optional function which might place the systems and information at risk. Thus, it's critical to choose the reliable e-learning system and tighten up the security and privacy of online education.

6 Current trends

Modern e-learning has evolved as a multi-disciplinary process including pedagogy, psychology, various aspects of computer science and many other fields of engineering. Both pedagogy and technology factors have been considered in these trends. Concepts like blended learning, adaptive learning have been introduced to change the traditional in-class education into competency-based education. Also, the wide use of e-learning systems has resulted in huge amount of data generated. By applying data-intensive approaches to educational resources, we can get better understanding of learners, educational settings, and the education results and then improve the teaching and learning process.

6.1 Blended learning

Blended learning (or hybrid learning) is evolved from the original computer-based learning environment. It combines the benefits of classroom learning with the advantages of e-learning to ensure an effective learning environment. In other words, learning activities take place inside and outside the classroom. Especially during the COVID-19 pandemic, some or even all classroom teaching are replaced with online teaching (Müller & Mildenberger, 2021; Prahmana et al., 2021) and evaluating the effectiveness of blended learning have been studied widely. Normally, there are many education and technology elements that can be incorporated in learning and teaching processes based on different learning purpose. According to (Valiathan, 2002)

and (Prahmana et al., 2021), there are three blended learning models: skill-driven, attitude-driven and competency-driven.

- `Skilled-driven model`: It aims at providing students specific knowledge and skills, while teachers give feedback and guidance.
- `Attitude-driven model`: It enables learners to gain new attitude and behaviors, and interaction and collaboration between learners and teachers plays an important role.
- `Competency-driven model`: It aims at transferring learners tacit knowledge by observing and interacting with experts on the job.

Unlike traditional education where the classroom focus is on the teacher, blended learning allows the use of digital texts and tools, and the students become the protagonist of their own learning process, constructing their own knowledge together with the teachers. This mix between classroom learning and e-learning facilitates the students to carry out a more direct and flexible learning style that matches students' diverse needs.

6.2 Adaptive learning

Traditionally, a standard e-learning system does not consider individual differences of learners and treat all learners equally. However, adaptive learning or personalized learning aims to tailor massive information to them based on their features, preferences, background and learning behaviors (Aroyo et al., 2006; Gomede et al., 2021; Mavroudi et al., 2018). In order to do so, adaptive learning basically utilizes a data-driven method to identify the students' needs faster, and therefore enable the delivery of personalized learning at scale. It also needs differentiated teaching strategies and smart feedback to build learner skills (Sonwalkar, 2013). For instance, algorithms are used to evaluate students' current learning conditions using online tests, thus adapted modules will be provided to identify their learning gap and improve learning outcomes. According to (Onah & Sinclair, 2015) , the most common adaptive method for course development have the following aspects: the adaptive hypermedia information retrieval system; adaptive annotation system; adaptive recommendation system; adaptive web navigation; adaptive feedback. (Oxman et al., 2014) conclude that at least three components are needed for an adaptive system.

- a content model to structure the provided contents,
- a learner model to understand learners abilities, and
- an instructional model to match the content with learners in a dynamic and personalized way.

Generally, adaptive learning can support adjustments in faculty role, allow creative teaching methods, and facilitate learning process in multiple ways.

6.3 Educational data mining (EDM) and learning analytics (LA)

EDM refers to developing technical methods to explore different kinds of educational datasets for the purpose of better understanding of learners and educational

settings (Mohamad & Tasir, 2013). LA is a closely related endeavor to EDM, and mainly emphasizes on the process of collecting, measuring, analyzing and reporting data about learners and their contexts, in order to understand and optimize learning and the environments in which it takes place (Knobbout & van der Stappen, 2018). However, there are several differences between them according to (Siemens & Baker, 2012). EDM focuses much more on automated adaptation and discovery, whereas LA is mainly in support of teachers and learners judgment. Technically, EDM focus on using typical data mining methods to assist learning process analysis including commonly used techniques like classification, clustering and association rule mining (Mavroudi et al., 2018). In addition to these methods, LA may also take advantage of methods like statistical analysis, Social Network Analysis (SNA) and visualization tools to enable users to gain an overview of the learning results. They establish an ecosystem to reshape the existing models of education and provide new solutions to facilitate the teaching and learning process.

In response to these trends, the focus has been shifted from traditional e-learning toward smart e-learning by integrating big data technology within e-learning paradigm (Kumar et al., 2016) or embedding students' cognitive model and theories into intelligent learning environments (Gamalel-Din, 2010).

7 Big data based E-learning systems

Generally, technologies like data science and artificial intelligence (AI) are driving blended learning and adaptive learning effectively. However, combining these technologies with e-learning is still in the early stages. Where and how to place these technologies in e-learning need to be exploited to achieve smart blended learning and adaptive learning. Many researches have been conducted during the past few years. We present a general architecture of big data based e-learning in Fig. 3. It combines essential technologies contributing to collect, aggregate, preprocess, analyze, store and manage big data in e-learning systems. These logical components
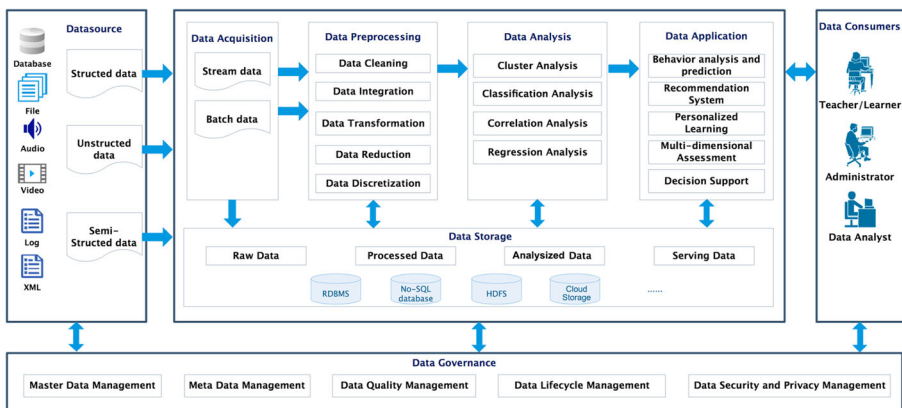

Fig. 3 The Architecture of Big Data Based E-learning Systems

perform specific functions to enable readers to understand the lifecycle of transforming the different e-learning data into valuable teaching and learning guidance through state-of-the-art technologies (Davoudian & Liu, 2020).

Firstly, data in e-learning systems come from different sources with various formats and different granularity. Generally, there are four types of data sources: learning resources, user information, user behavior/activities, and collaboration information. Specifically, learning resource can be audios, videos, slides, texts, presentations or any other kind of documents or files that are used to create course content and pedagogical resources. Another data source comes from users' demographic information including their gender, age, education, occupation, experiences, and prior knowledge etc. The third type of data source includes users' behavior and interactions with the system (e.g., log in/out activities, visited contents, quizzes, tests, assignments). The last type of data source comes from the users' cooperation via emails, instant communication tools, forums and so on.

Taking into account the heterogeneous and hierarchical nature of data sources, it becomes essential to determine data structures and formats that reflect an event (Romero & Ventura, 2013; Otoo-Arthur & van Zyl, 2020). There are three kinds of data: structured, semi-structured or unstructured. Structured data are organized through predefined structures and are stored as records in tables with predefined columns in relational database systems such as MySQL, Oracle, SQL Server and PostgreSQL. Semistructured data have internal structures but are not organized through predefined structures. They are represented in CSV, XML, JSON and other markup languages and are stored in NoSQL stores such as Redis, MongoDB, Cassandra and HBase (Davoudian et al., 2018). Unstructured data have no predefined structures and are stored as text files, emails, videos, audios, web pages, etc. For e-learning data analysis, these different types of data need to be considered together.

7.1 Data acquisition

It is the very first step for e-learning data processing. It collects data generated from distributed information sources with various frequencies, sizes, and formats (Wang et al., 2018). It is important that the collected data align with the research questions to be solved in the system. The commonly used open protocols for data acquisition are Advanced Message Queuing Protocol (AMQP) and Java Message Service(JMS) (Lyko et al., 2016). The former is an open source standard for asynchronous messaging between applications considering security, reliability, and performance, whereas the latter allows programs to access system's messages easily. Regarding common techniques for data acquisition, several tools can be used to collect and aggregate multiple datasets effectively based on different data sources. For instance, Apache Flume can transport large amounts of streaming data such as log files into a centralized store like HDFS at a higher speed, or turn it into a producer of Kafka (Landset et al., 2015). Additionally, Apache Sqoop is designed to transfer batch data between Apache Hadoop and structured database, and dump structured data into HDFS (Geng et al., 2019; Dahdouh et al., 2018).

7.2 Data preprocessing

It involves converting original data into an understandable format by using multiple data mining methods or techniques. The collected data are often incomplete, inconsistent, noisy, superfluous and containing errors, which can contribute to incorrect or misleading data analysis and consequently make data analysis slow and inaccurate. So it is not directly applicable to start a data mining process. Alternatively, for further processing and evaluation, the raw data need be converted into correct and helpful information. Generally, data preprocessing includes the following techniques.

- Data cleaning. It is utilized to discovery missing values, highlight errors, identify outliers, smooth the noisy data, get rid of duplicates, convert improperly formatted or address the inconsistencies in the data to improve data accuracy and quality.
- Data integration. It is used to combine data with different representations from multiple sources and also detect and resolve data conflicts.
- Data transformation. It converts the original data from one format to another by using normalization, aggregation and generalization and includes steps such as interpretation, pre-translation data quality check, data translation and post-translation data quality check.
- Data reduction. It is used to reduce the amount of data required for analysis, through methods such as dimensionality reduction, compression, deduplication, numerosity reduction and so on.
- Data discretization. It involves the process of partitioning continuous attributes to several discretized values to prepare datasets well for some mining algorithms such as decision tree that assume discrete values.

Data analysis

It is utilized to process various types of data and perform proper analyses for teaching and learning assistance and improvements. Generally, the common analysis methods can be carried out in realtime, offline and hybrid way. Realtime analysis involves a continual input, analyze and output of data that need to be processed within a short period of time. Parallel processing and memory-based processing are two general methods for realtime analysis (Wang et al., 2018). Tools like Apache Storm, Spark, Flink, S4 (da Silva & et al, 2016), Kafka, SAP Hana (Chandio et al., 2015) are used to deal with realtime data. Offline analysis is utilized to analyze historical data for the purpose of identifing patterns in the environment without high requirement on response time. Most traditional e-learning systems utilized offline analysis method to handle and analyse the learning data in batch. The typical examples include components of Hadoop framework. Specifically, Spark and Hadoop MapReduce are commonly used tools to deal with batch processing. Additonally, a hybrid computation model can be used to deal with massive data by combing the advantages of both batch and realtime analysis.

Several representative data analysis techniques are introduced below to analyze huge data and mining valuable information concealed in the raw data sets.

- Cluster analysis. It involves grouping same or similar records into clusters. For example, Coursera utilizes t-distributed stochastic neighbour embedding (t-SNE) algorithm to group courses into categories. This method makes categorization scheme simpler with less redundant. Clusters can also be developed in educational information at several distinct grain sizes. Student activities can be combined to explore behavior patterns (Baker & Inventado, 2014). Similarly, teachers or students can be grouped together to discovery similarities or differences among them.
- Classification. It is used to classify a value into a specific category in which all the items have very similar or the same characteristics. Classification has been commonly utilized in e-learning to predict the performance of learners (Ahmed & Elaraby, 2014; Baker & Inventado, 2014; Rasheed & Wahid, 2021). For example, the possibility of a student to pass the exam (Rustia et al., 2018) or the students who tend to dropout (Pal, 2012) can be predicted based on students and behavior characteristics.
- Correlation analysis. Its goal is to find correlations between variables or attributes. This may be an attempt to determine which variables are most closely related to a single variable of specific concern (Sachin & Vijay, 2012). For example, correlation analysis tools can be used to identify cheat behaviors based on knowledge test and exercise tasks databases (Teodorov et al., 2011).
- Regression analysis. It is a predictive modelling techinique, which usually uses statistical method to predict one variable by examining the relationships among a series of variables. Usually, a variable is identified as the predicted variable and a set of other variables as the predictors (Angeli et al., 2017). For example, regression analysis is applied to predict which metrics help explain bad examination results (Feng et al., 2005) in the smart e-learning system.

7.3 Data storage

Its main objective is to guarantee the efficient storage of raw data, processed data, analyzied data, and serve data with various types throughout the lifecycle of big data architecture. Since the data are produced at ever-increasing velocity, they must be gathered and stored at low price. Distributed file systems (DFS) are commonly adopted to keep low storage cost and ensure data availability and reliability for data analytics. The typical file systems include GFS, HDFS, TFS and FastTFS by Taobao, Microsoft Cosmos, and Facebook Haystack (Chandio et al., 2015). Other alternatives for distributed data storage are also available that either operate on top of the file storage system or run as standalone devices. Traditional relational databases like Oracle and Postgres are widely used to store structured data. Techniques including replication, caching, horizontal or vertical scaling can be used to deal with the huge volume of data. Non-relational databases known as NoSQL (Not only SQL) stores, are appropriate for intelligent applications as they support multiple data structures (Landset et al., 2015; Davoudian et al., 2018). The types of NoSQL databases usually include

key-value storage like Redis, document storage like MongoDB, column-oriented storage like HBase and Graph-based storage like Titan, Neo4J and OrientDB.

7.4 Data governance

It refers to a collection of practices and processes about security, integrity, usability and availability of the data employed in an application. It builds rules and guidelines for data quality management across the organization, and allocates responsibilities to each role it defines (Wende & Otto, 2007). It also implements access control and other data security measures, capture the meta data of datasets to support security efforts and facilitates end-user data consumption. Generally, data governance usually includes master data management, meta data management, data quality management, data lifecycle management, data security and privacy management.

- Master data management. Master data is data most critical to an organization's operations and analytics. Master data management is a technology-enabled discipline which aims to control master data assets and ensure their consistency and accuracy (Allen & Cervo, 2015). It can help to improve data quality and facilitate computing in multiple system architectures, platforms and applications.
- Meta data management. Meta data is the information that describes the semantics of other data. It specifically identifies the attributes, properties and tags that describe and classify information. Meta data management is a critical component for data governance practice since data are scattered in various formats and coming from many sources in the big data environment.
- Data quality management. It plans, uses, organizes and disposes data in a quality-oriented manner for the purpose of improving decision making and business services (Weber et al., 2009). It is one of the most significant fields since unqualified data ultimately have enormous adverse effects on data analysis. For example, if we intend to accurately predict student behavior, high quality data are required to develop prediction models.
- Data lifecycle management. It refers to decisions that define the definition, production, retention and retirement of data (Khatri & Brown, 2010). In other words, it comprehensively monitors and tracks the life process of data presentation, which paves the way for policies on sensitive data protection and security control.
- Data security and privacy management. For the learning process, there are very few researches specifically on data security and privacy, even though both play a significant role in e-learning. Usually, we need measures to implement users authentication and privacy protection, exam protection for data integrity, courseware copyright protection, floor control security for synchronized communication activities (Eibl, 2009) and so forth. For example, student information that will not be used for analysis purpose should be excluded in data acquisition stage to maintain the confidentiality of individual information. Also, personal privacy data need to be protected and encrypted during the lifecycle of data analysis process. Therefore, considering the security and privacy factors while developing e-learning applications can ensure a reliable, efficient learning environment for teaching and learning activities.

8 Data application

It refers to the application domain that EDM/LA technologies can be adopted in e-learning systems. (Fauvel et al., 2018) reviewed innovative AI techniques that affects MOOCs education. They focused on the researches about student learning behaviors, engagement, and learning performance by constructing intelligent and personalized learning tracks. They categorized the research into three areas including learner modeling, learning experience improvement and learner assessment. Similar, (Bakhshinategh et al., 2018) reviewed current applications that have been used for EDM and classified them into multiple groups and subgroups. Generally, the major goals of EDM are as follows (Vora & Iyer, 2018): (1) providing feedback to support instructors; (2) detecting student behavior; (3) predicting student's performance; (4) recommendations for students; (5) constructing courseware; (6) planning and scheduling. According to these researches, we review the literatures which focus on the specific learning/teaching purposes including behavior analysis and prediction, recommendation system, personalized learning and multidimensional assessment below.

8.1 Behavior analysis and prediction

With big data analytic techniques, instructors can monitor and analyze various online activities accurately such as how long the learners take to answer a question or submit an assignment, how much time they spend in a course, which questions they skip on the test, which part of knowledge they are interested most and so forth. They could also discover various factors that influence students performance and predict the future trends based on these explorations. Additionally, disruptive behaviors include low engagement, excessive lateness, high dropping out rate, cheating on assignments and tests, low learning effectiveness, derogatory comments in online discussion or email can also be found. Since instructors and learners do not have face-to-face interaction, disruptive student behaviors existing among traditional e-learning education environment cannot be disclosed immediately. Big data technology can detect and determine those unacceptable behaviors, while adhering to procedures for reporting disruptive incidents. Then, proper interventions aiming to stop those behaviors or motivate weaker students can be conducted accordingly to mitigate negative factors in e-learning environment. Generally, behavior analysis and prediction mainly focuses on aspects like student learning motivation, engagement, participation, dropout rate, performance success and so on. Several examples are listed in Table 1.

8.2 Recommendation system

Learners are overwhelmed with the large numbers of learning resources available online. It is becoming more and more difficult for them to select suitable learning materials in e-learning environments. Recommendation systems provide an effective approach to solve this issue by assisting learners to discover appropriate learning contents and improve learning outcomes (Fauvel et al., 2018). Basically, there are four strategies commonly used in recommendation system: collaborative filtering

Table 1 Behavior analysis and prediction

Topic	Methods	Reference
User satisfaction, student motivation and engagement, learning performance	Gamification techniques	(Lisitsyna et al., 2015; Urh et al., 2015; Muntean, 2011)
Correlation between students' involvement in asking questions and final performance	Data mining & text mining	(He, 2013)
Relationship between achievement objectives and help-seeking strategies	Clustering	(Vaessen et al., 2014)
Student behavioral structure, predict possibility to obtain a certificate	Latent Dirichlet Allocation	(Seaton & Chuang, 2015)
Patterns of similar user behaviors	Clustering	(Talavera & Gaudioso, 2004)
Students' learning pace and progress	Data clean and analyze	(Hadavand & Leek, 2018)
Predict & evaluate low-engagement	Classification	(Hussain et al., 2018)
Dropout prediction	Classification	(Liang et al., 2016; Feng et al., 2019; Xing & Du, 2019)
Detect learning style	Classification	(Rasheed & Wahid, 2021)

Table 2 Recommendation system

Topic	Methods	Reference
Dissimilarity among courses	Content-based	(Hou et al., 2018)
Relationships between courses activities	Association rules	(Dahdouh et al., 2019)
Recommend learning resources	Context-based method, sequential pattern mining, CF algorithms	(Tarus et al., 2018)
Distribute tailored learning contents	Association rules, content filtering, collaborative filtering	(Xiao et al., 2018)
Recommend learning resources	Cluster-based hybrid algorithm	(Bhaskaran et al., 2021)

(CF), matrix and tensor factorization, content-based (CB) techniques and association rule mining (Klašnja-Milićević et al., 2017; Ibrahim et al., 2020). Also, many researchers adopt a hybrid recommendation approach by combining the advantages of above-mentioned strategies to promote the quality of recommendations. Some recommendation system examples are shown in Table 2.

8.3 Personalized learning

It refers to a customized way that adapts to the learners' personalized requirements in e-learning systems. Personalized learning is critical since numerous learners with various background and needs will get involved in e-learning systems. The predefined procedure of learning resources to be followed by the students in a course cannot meet all learners' particular objectives. According to students learning pattern, individual preferences, and knowledge states (Jeong et al., 2013), personalized learning content can be provided to each individual learner accordingly. Learners are able to reduce the time spent on finding proper contents and get personalized service and meaningful learning experience. Additionally, it can assist them to access the individualized sequence of resources produced and adapted to what they need, rather than following the predefined learning route. Also, the identification of targeted content can satisfy teaching needs among large, heterogeneous and complicated resources. Some examples of personalized learning can be found in Table 3.

Table 3 Personalized learning

Topic	Methods	Reference
Produce tailored learning paths	AI planning techniques	(Garrido et al., 2016)
Recognize learning style and adjust learning content	ThoTh Lab platform	(Deng et al., 2018)
Personalized curriculum sequencing with matched difficulties	Genetic algorithm	(Huang et al., 2007)

Table 4 Multidimensional assessment

Topic	Methods	Reference
Autonomous assessment of assignments	Web-based system	(Le Ru et al., 2015)
Evaluate assignment	Formative assessments	(Klimova, 2015)
Peer assessment	Clustering	(Lynda et al., 2017)
Learning performance assessment	Classification	(Cheng et al., 2011)

8.4 Multidimensional assessment

Student learning evaluation and assessment is a key characteristic for e-learning systems. However, it is a challenging task to conduct self-sustainable or personalized evaluation to suit the learner population. Therefore, various dimensions need to be adopted to assess the efficiency of e-learning systems. Table 4 provides some examples for multidimensional assessment.

8.5 Decision making

Decision making is an essential part of e-learning systems. The participants and related stakeholders need to make appropriate decisions to adjust the teaching/learning method or the program process. According to (Galvis, 2018), many contextual factors have an influence on an institution's decisions. Users are able to establish decision processes that convert education data into actionable insight that can help improve learning performance (Picciano, 2012). Basically, they identify problems and spot opportunities for positive change. Consequently, they can draw precise conclusions and make a better decision by examining and analyzing data consecutively. Additionally, e-learning systems provide possibility to improve decision making capability based on extensive data analysis. Table 5 gives some examples for decision making.

Table 5 Decision making

Topic	Methods	Reference
Precision teaching interventions	Data-driven decision model	(Zhu et al., 2016)
Provide knowledge for decision-maker	Knowledge-based data prototype	(Abu-Naser et al., 2011)
Decisions for mobile learning	Task-interaction model	(Fulantelli et al., 2015)

9 Conclusion

E-learning systems have been increasingly used to provide efficient learning services, especially after the declaration of the global COVID-19 pandemic by the World Health Organization in mid-March 2020. A lot of post-secondary institutions have introduced e-learning systems alongside online courses.

In this paper, we provide comprehensive review on the efforts of applying new information and communication technologies to improve e-learning services. We first systematically investigate current e-learning systems in terms of their classification, architecture, functions, challenges, and current trends. We then present a general architecture for big data based e-learning systems to meet the ever-growing demand for e-learning. We also describe how to use data generated in big data based e-learning systems to support more flexible and customized course delivery and personalized learning.

Based on the general architecture presented, we have systematically implemented a novel big data based e-learning system called Weblearn[17] that has been used by several universities in China for e-learning since 2021.

We are now working on data preprocessing, data analysis, and data application as shown in Fig. 3 in order to provide customized course delivery and personalized learning.

References

Abu-Naser, S., Al-Masri, A., Sultan, Y. A., & Zaqout, I. (2011). A prototype decision support system for optimizing the effectiveness of elearning in educational institutions. *Int. J. Data Mining Knowledge Management Process (IJDKP)*, *1*, 1–13. https://doi.org/10.5121/ijdkp.2011.1401.

Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. *World Journal Computer Applied Technology*, *2*(2), 43–47. https://doi.org/10.13189/wjcat.2014.020203.

Alameri, J., Masadeh, R., Hamadallah, E., Bani, H., & Fakhouri, H. (2020). Students' perceptions of e-learning platforms (moodle, microsoft teams and zoom platforms) in the university of jordan education and its relation to self-study and academic achievement during covid-19 pandemic, vol. 11:13.

Allen, M., & Cervo, D. (2015). Multi-domain master data management: Advanced MDM and data governance in practice.

Alonso, F., López, G., Manrique, D., & Viñes, J. M. (2005). An instructional model for web-based e-learning education with a blended learning process approach. *British Journal of Educational Technology*, *36*(2), 217–235.

Andersson, A. (2008). Seven major challenges for e-learning in developing countries: Case study ebit, sri lanka. *International Journal of Education and Development using ICT*, *4*(3), 45–62.

---

[17]http://www.weblearn.cn/

Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P.A (2017). Data mining in educational technology classroom research: Can it make a contribution?. *Computers Education*, *113*, 226–242. https://doi.org/10.1016/j.compedu.2017.05.021.

Aroyo, L., Dolog, P., Houben, G.-J., Kravcik, M., Naeve, A., Nilsson, M., & Wild, F. (2006). Interoperability in personalized adaptive learning. *Journal of Educational Technology Society*, *9*(2), 4–18.

Bagarukayo, E., Mpangwire, V., & Kalema, B.M. (2014). Integrating a chatting tool into a learning management system. In *3Rd international conference on informatics engineering and information science (ICIEIS2014), pp. 271–275.*

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics, pp. 61–75.*

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education Information Technologies*, *23*(1), 537–553. https://doi.org/10.1007/s1063.

Bari, M., Djouab, R., & Hoa, C.P (2018). Elearning current situation and emerging challenges. *PEOPLE: International Journal of Social Sciences*, *4*, 2. https://doi.org/10.20319/pijss.2018.42.97109.

Baturay, M. H. (2015). An overview of the world of moocs. *Procedia - Social and Behavioral Sciences*, *174*, 427–433. https://doi.org/10.1016/j.sbspro.2015.01.685, International conference on new horizons in education, INTE 2014, pp. 25–27, June 2014, Paris.

Bhaskaran, S., Marappan, R., & Santhi, B. (2021). Design and analysis of a cluster-based intelligent hybrid recommendation system for e-learning applications. *Mathematics*, *9*(2), 197.

Brajkovic, E., Rakic, K., & Kraljevic, G. (2018). Application of data mining in e-leaming systems. In *2018 17th international symposium on INFOTEH-JAHORINA, INFOTEH 2018 - proceedings 2018-Janua(March), pp 1–5.* https://doi.org/10.1109/INFOTEH.2018.8345536.

Carmona, A. B. C., Muñoz, C. M., & Vargas, A.C. (2008). Virtual whiteboard and chat for a learning management system. In *E-learning, pp. 263–270.*

Cavus, N., & Zabadi, T. (2014). A comparison of open source learning management systems. In *Procedia - social and behavioral sciences, vol. 143, pp. 521–526.* https://doi.org/10.1016/j.sbspro.2014.07.430. 3rd cyprus international conference on educational research, CY-ICER 2014, 30 January–2014, 1 February, Lefkosa, North Cyprus.

Chandio, A. A., Tziritas, N., & Xu, C.-Z. (2015). Big-data processing techniques and their challenges in transport domain ZTE Communications, vol. 1(010).

Chao, K.-M., James, A. E., Nanos, A. G., Chen, J.-H., Stan, S.-D., Muntean, I., Figliolini, G., Rea, P., Bouzgarrou, C. B., Vitliemov, P., & et al (2015). Cloud e-learning for mechatronics: Clem. *Future Generation Computer Systems*, *48*, 46–59. https://doi.org/10.1016/j.future.2014.10.033.

Cheng, C.-H., Chen, T.-L., Wei, L.-Y., & Chen, J.-S (2011). A new e-learning achievement evaluation model based on rbf-nn and similarity filter. *Neural Computing and Applications*, *20*(5), 659–669. https://doi.org/10.1007/s00521-009-0280-0.

Cheng, P.-H., Yang, Y.-T. C., Chang, S.-H. G., & Kuo, F.-R.R (2015). 5e mobile inquiry learning approach for enhancing learning motivation and scientific inquiry ability of university students. *IEEE Transactions on Education*, *59*(2), 147–153. https://doi.org/10.1109/TE.2015.2467352.

Chiu, T. K., & Hew, T. K. (2018). Factors influencing peer learning and performance in mooc asynchronous online discussion forum. Australasian Journal of Educational Technology, vol. 34(4), https://doi.org/10.14742/ajet.3240.

Chung, C.-H., Pasquini, L. A., & Koh, C.E. (2013). Web-based learning management system considerations for higher education. *Learning and Performance Quarterly*, *1*(4), 24–37.

Seaton, D. T., & Chuang, I. (2015). Probabilistic use cases: Discovering behavioral patterns for predicting certification. In *Proceedings of the second (2015) ACM conference on learning scale. LS '15, pp. 141–148. ACM, https://doi.org/10.1145/2724660.2724662*.

da Silva, V. A., & et al (2016). Strategies for big data analytics through lambda architectures in volatile environments. *IFAC-PapersOnLine*, *49*(30), 114–119. https://doi.org/10.1016/j.ifacol.2016.11.138.

Dahdouh, K., Dakkak, A., Oughdir, L., & Messaoudi, F. (2018). Big data for online learning systems. *Education and Information Technologies*, *23*(6), 2783–2800. https://doi.org/10.1007/s10639-018-9741-3.

Dahdouh, K., Oughdir, L., Dakkak, A., & Ibriz, A. (2019). Building an e-learning recommender system using association rules techniques and r environment. *International Journal of Information Science and Technology*, *3*(2), 11–18.

Datta, S., & Bhattacharyya, S. (2018). Simple spirited scalable e-learning system. In *2018 IEEE 5th international congress on information science and technology (CiSt), pp. 368–373*. https://doi.org/10.1109/CIST.2018.8596631.

Davoudian, A., & Liu, M. (2020). Big data systems: A software engineering perspective. *ACM Computing Surveys*, *53*(5), 110–111039. https://doi.org/10.1145/3408314.

Davoudian, A., Chen, L., & Liu, M. (2018). A survey on nosql stores. *ACM Computing Surveys*, *51*(2), 40–14043. https://doi.org/10.1145/3158661.

Deng, Y., Lu, D., Chung, C.-J., Huang, D., & Zeng, Z. (2018). Personalized learning in a virtual hands-on lab platform for computer science education. In *2018 IEEE frontiers in education conference (FIE), pp. 1–8. IEEE*. https://doi.org/10.1109/FIE.2018.8659291.

Eibl, C. J. (2009). Privacy and confidentiality in e-learning systems. In *2009 fourth international conference on internet and web applications and services, pp. 638–642*. https://doi.org/10.1109/ICIW.2009.102.

El Mhouti, A., Erradi, M., & Nasseh, A. (2018). Using cloud computing services in e-learning process: Benefits and challenges. *Education and Information Technologies*, *23*(2), 893–909. https://doi.org/10.1007/s10639-017-9642-x.

Fajar, A. N., Nurcahyo, A., & Sriratnasari, S.R. (2018). Soa system architecture for interconected modern higher education in indonesia. *Procedia Computer Science*, *135*, 354–360. https://doi.org/10.1016/j.procs.2018.08.184. The 3rd international conference on computer science and computational Intelligence (ICCSCI 2018) : Empowering smart technology in digital era for a better life.

Fauvel, S., Yu, H., Miao, C., Cui, L., Song, H., Zhang, L., Li, X., & Leung, C. (2018). Artificial intelligence powered moocs: A brief survey. In *2018 IEEE international conference on agents (ICA), pp. 56–61*. https://doi.org/10.1109/AGENTS.2018.8460059.

Feng, M., Heffernan, N. T., & Koedinger, K.R. (2005). Looking for sources of error in predicting student's knowledge. In *Educational data mining: Papers from the 2005 AAAI workshop, pp. 54–61*.

Feng, W., Tang, J., & Liu, T.X. (2019). Understanding dropouts in moocs. In *AAAI 2019*.

Fulantelli, G., Taibi, D., & Arrigo, M. (2015). A framework to support educational decision making in mobile learning. *Computers in Human Behavior*, *47*, 50–59. https://doi.org/10.1016/j.chb.2014.05.045.

Galvis, Á. H. (2018). Supporting decision-making processes on blended learning in higher education: Literature and good practices review. *International Journal of Educational Technology in Higher Education*, *15*(1), 25. https://doi.org/10.1186/s41239-018-0106-1.

Gamalel-Din, S. A. (2010). Smart e-learning: A greater perspective; from the fourth to the fifth generation e-learning. *Egyptian Informatics Journal*, *11*(1), 39–48. https://doi.org/10.1016/j.eij.2010.06.006.

García-Holgado, A., & García-Peñalvo, F. J. (2018). Human interaction in learning ecosystems based on open source solutions. In P. Zaphiris, & A. Ioannou (Eds.) *Learning and collaboration technologies. Design, development and technological innovation, pp. 218–232. Springer*.

Garrido, A., Morales, L., & Serina, I. (2016). On the use of case-based planning for e-learning personalization. *Expert Systems with Applications*, *60*, 1–15. https://doi.org/10.1016/j.eswa.2016.04.030.

Geng, D., Zhang, C., Xia, C., Xia, X., Liu, Q., & Fu, X. (2019). Big data-based improved data acquisition and storage system for designing industrial data platform. *IEEE Access*, *7*, 44574–44582. https://doi.org/10.1109/ACCESS.2019.2909060.

Goldie, J. G. S. (2016). Connectivism: A knowledge learning theory for the digital age? *Medical teacher*, *38*(10), 1064–1069.

Gomede, E., Barros, R. M. D., & Mendes, L.D.S. (2021). Deep auto encoders to adaptive e-learning recommender system - sciencedirect. Computers and education: Artificial intelligence.

González, M. C., García-Peñalvo, F., Guerrero, M. C., & Forment, M.A. (2009). Back and forth: From the lms to the mobile device. a soa approach. In *Proceedings of the IADIS international conference mobile learning 2009 (Barcelona, Spain, 26-28 February, 2009), pp. 114–120. IADIS Press, Portugal*.

Hadavand, A., & Leek, J. (2018). crsra: A package for cleaning and analyzing coursera research export data. bioRxiv, pp. 275537, https://doi.org/10.1101/275537.

Hamdan, M., Jaidin, J. H., Fithriyah, M., & Anshari, M. (2020). E-learning in time of covid-19 pandemic: Challenges & experiences. In *2020 sixth international conference on e-learning (econf), pp. 12–16. IEEE*.

Han, I., & Shin, W. S. (2016). The use of a mobile learning management system and academic achievement of online students. *Computers & Education*, *102*, 79–89. https://doi.org/10.1016/j.compedu.2016.07.003.

Hapsari, I. N., Parhusip, A. R., Wahyu, S., Sutanto, I., Firmansyah, G., & Rosyid, A. (2021). E-learing service issues and challenges: An exploratory study. In *2021 1st international conference on computer science and artificial intelligence (ICCSAI), vol. 1, pp. 196–201. IEEE*.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, *29*(1), 90–102. https://doi.org/10.1016/j.chb.2012.07.020.

Hendradi, P., Abd Ghani, M. K., Mahfuzah, S., Yudatama, U., Prabowo, N. A., & Widyanto, R.A. (2020). Artificial intelligence influence in education 4.0 to architecture cloud based e-learning system. *International Journal of Artificial Intelligence Research*, *4*(1), 30–38.

Hill, P. (2019). State of higher Ed LMS market for US and Canada: 2018 Year-End Edition. https://eliterate.us/state-higher-ed-lms-market-us-canada-end-2018.

Hou, Y., Zhou, P., Xu, J., & Wu, D.O. (2018). Course recommendation of mooc with big data support: A contextual online learning approach. In *IEEE INFOCOM 2018-IEEE conference on computer communications workshops (INFOCOM WKSHPS), pp. 106–111*. https://doi.org/10.1109/INFOCOMW.2018.8406936.

Hu, H., Zhang, G., Gao, W., & Wang, M. (2020). Big data analytics for mooc video watching behavior based on spark. Neural Computing and Applications, vol. 32, https://doi.org/10.1007/s00521-018-03983-z.

Huang, M.-J., Huang, H.-S., & Chen, M.-Y. (2007). Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Expert Systems with Applications*, *33*(3), 551–564. https://doi.org/10.1016/j.eswa.2006.05.019.

Hussain, M., Zhu, W., Zhang, W., & Abidi, S.M.R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. Computational Intelligence and Neuroscience, vol. 2018, https://doi.org/10.1155/2018/6347186.

Ibrahim, T. S., Saleh, A. I., Elgaml, N., & Abdelsalam, M.M. (2020). A fog based recommendation system for promoting the performance of e-learning environments. *Computers & Electrical Engineering*, *87*, 106791.

Islam, N., Beer, M., & Slack, F. (2015). E-learning challenges faced by academics in higher education. *Journal of Education and Training Studies*, *3*(5), 102–112. https://doi.org/10.11114/jets.v3i5.947.

Jeong, J.-S., Kim, M., Yoo, K.-H., & et al (2013). A content oriented smart education system based on cloud computing. *International Journal of Multimedia and Ubiquitous Engineering*, *8*(6), 313–328. https://doi.org/10.14257/ijmue.2013.8.6.31.

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. The International Review of Research in Open and Distributed Learning, vol. 15(1), https://doi.org/10.19173/irrodl.v15i1.1651.

Kakasevski, G., Mihajlov, M., Arsenovski, S., & Chungurski, S. (2008). Evaluating usability in learning management system moodle. In *ITI 2008-30th international conference on information technology interfaces, pp. 613–618. IEEE, https://doi.org/10.1109/ITI.2008.4588480*.

Kanimozhi, S., Kannan, A., Suganya Devi, K., & Selvamani, K. (2019). Secure cloud-based e-learning system with access control and group key mechanism. Concurrency and computation: Practice and experience, 4841, https://doi.org/10.1002/cpe.4841.

Kappe, F., & Scerbakov, N. (2017). Object-oriented architecture of a modern learning management system. In *EDULEARN17 proceedings. 9th international conference on education and new learning technologies, pp. 4910–4916*. https://doi.org/10.21125/edulearn.2017.2101.

Karjo, C. H., Andreani, W., Herawati, A., Ying, Y., Yasyfin, A. P., & Marie, K. (2021). Teachers' challenges and needs in e-learning environment. In *2021 international seminar on application for technology of information and communication (iSemantic), pp. 405–409. IEEE*.

Kawato, T., Higashino, M., Takahashi, K., & Kawamura, T. (2020). Proposal of distributed e-learning system using idle resources. In *2020 5Th international conference on computer and communication systems (ICCCS)*.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148–152. https://doi.org/10.1145/1629175.1629210.

Klašnja-Milićević, A., & Ivanović, M. (2021). E-learning personalization systems and sustainable education MDPI.

Klašnja-Milićević, A., Vesin, B., Ivanović, M., Budimac, Z., & Jain, L.C. (2017). Recommender systems in e-learning environments. In *E-learning systems, pp. 51–75*.

Klimova, B. (2015). Assessment in smart learning environment – a case study approach. In V. L. Uskov, R. J. Howlett, & L. C. Jain (Eds.) *Smart Education and Smart e-Learning, pp. 15–24. Springer*.

Knobbout, J., & van der Stappen, E. (2018). Where is the learning in learning analytics? In V. Pammer-Schindler, M. Pérez-sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.) *Learning, lifelong technology-enhanced, pp. 88–100. Springer*.

Kumar, P., Malali, P., & Noronha, H. (2016). Big data integration for transition from e-learning to smart learning framework. In *2016 3rd MEC international conference on big data and smart city (ICBDSC), pp. 1–4. IEEE, https://doi.org/10.1109/ICBDSC.2016.7460379*.

Laisheng, X., & Zhengxia, W. (2011). Cloud computing: A new business paradigm for e-learning. In *2011 third international conference on measuring technology and mechatronics automation, vol. 1, pp. 716–719*. https://doi.org/10.1109/ICMTMA.2011.181.

Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the hadoop ecosystem. *Journal of Big Data*, *2*(1), 24. https://doi.org/10.1186/s40537-015-0032-1.

Laura, G., Bogdan, O., Aurelia, C., & Serban, M. (2018). University learning management system based on office 365. In *2018 17th international conference on information technology based higher education and training (ITHET),pp. 1–6. IEEE, Olhao, Portugal, https://doi.org/10.1109/ITHET.2018.8424799*.

Le Ru, Y., Aron, M., Gerval, J.-P., & Napoleon, T. (2015). Tests generation oriented web-based automatic assessment of programming assignments. In V. L. Uskov, R. J. Howlett, & L. C. Jain (Eds.) *Smart education and smart e-learning, pp. 117–127. Springer*.

Liang, J., Li, C., & Zheng, L. (2016). Machine learning application in moocs: Dropout prediction. In *2016 11th International conference on computer science & education (ICCSE), pp. 52–57, https://doi.org/10.1109/ICCSE.2016.7581554*.

Lisitsyna, L. S., Pershin, A. A., & Kazakov, M.A. (2015). Game mechanics used for achieving better results of massive online courses. In V. L. Uskov, R. J. Howlett, & L. C. Jain (Eds.) *Education, smart e-learning, smart, pp. 183-192. Springer*.

Lu, P., Cong, X., & Zhou, D. (2015). E-learning-oriented software architecture design and case study. *International Journal of Emerging Technologies in Learning (iJET)*, *10*(4), 59–65. https://doi.org/10.3991/ijet.v10i4.4698.

Lyko, K., Nitzschke, M., & Ngonga Ngomo, A.-C. (2016). Big data acquisition. In *New horizons for a data-driven economy, pp. 39–61*.

Lynda, H., Farida, B.-D., Tassadit, B., & Samia, L. (2017). Peer assessment in moocs based on learners' profiles clustering. In *2017 8th international conference on information technology (ICIT), pp. 532–536. IEEE, https://doi.org/10.1109/ICITECH.2017.8080054*.

Masud, M. A. H., & Huang, X. (2012). An e-learning system architecture based on cloud computing. *System*, *10*(11), 255–259.

Mavroudi, A., Giannakos, M., & Krogstie, J. (2018). Supporting adaptive learning pathways through the use of learning analytics: Developments, challenges and future opportunities. *Interactive Learning Environments*, *26*(2), 206–220. https://doi.org/10.1080/10494820.2017.1292531.

Mei, J. (2016). Learning management system calendar reminders and effects on time management and academic performance. *International Research and Review*, *6*(1), 29–45.

Min, H., & Nasir, M. (2020). Self-regulated learning in a massive open online course: a review of literature. *European Journal of Interactive Multimedia and Education*, *1*(2), 02007.

Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences*, *97*, 320–324. https://doi.org/10.1016/j.sbspro.2013.10.240.

Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H., & Shami, A. (2018). E-learning: Challenges and research opportunities using machine learning & data analytics. *IEEE Access*, *6*, 39117–39138. https://doi.org/10.1109/ACCESS.2018.2851790.

Muntean, C. I. (2011). Raising engagement in e-learning through gamification. In *Proc. 6th international conference on virtual learning ICVL, vol. 1, pp. 323–329. Romania*.

Müller, C., & Mildenberger, T. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educational Research Review*, *34*, 100394. https://doi.org/10.1016/j.edurev.2021.100394.

Onah, D. F., & Sinclair, J. (2015). Massive open online courses: an adaptive learning framework. In *9Th international technology, education and development conference, pp. 2–4*.

Oryakhail, M. S., Saay, S., & Nasery, H. (2021). Challenges in the implementation of e-learning in afghanistan higher education. In *2021 international conference advancement in data science, e-learning and information systems (ICADEIS), pp. 1–6. IEEE*.

Otoo-Arthur, D., & van Zyl, T. (2020). A scalable heterogeneous big data framework for e-learning systems. In *2020 international conference on artificial intelligence, big data, computing and data communication systems (icABCD), pp. 1–15. IEEE,*. https://doi.org/10.1109/icABCD49160.2020.9183863.

Oxman, S., Wong, W., & Innovations, D. (2014). White paper: Adaptive learning systems. Integrated Education Solutions, pp. 6–7.

Pal, S. (2012). Mining educational data using classification to decrease dropout rate of students. arXive:1206.3078.

Picciano, A. G. (2012). The evolution of big data and learning analytics in american higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20.

Pickard, L., Shah, D., & De Simone, J.J. (2018). Mapping microcredentials across mooc platforms. In *2018 learning with MOOCS (LWMOOCS), pp. 17–21*. https://doi.org/10.1109/LWMOOCS.2018.8534617.

Prahmana, R. C. I., Hartanto, D., Kusumaningtyas, D. A., & Ali, R.M. (2021). Muchlas: Community radio-based blended learning model: A promising learning model in remote area during pandemic era. *Heliyon*, *7*(7), 07511. https://doi.org/10.1016/j.heliyon.2021.e07511.

Rani, M., Nayak, R., & Vyas, O.P. (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. *Knowledge-Based Systems*, *90*, 33–48. https://doi.org/10.1016/j.knosys.2015.10.002.

Rasheed, F., & Wahid, A. (2021). Learning style detection in e-learning systems using machine learning techniques. Expert Systems with Applications (3):114774.

Riahi, G. (2015). E-learning systems based on cloud computing: A review. *Procedia Computer Science*, *62*, 352–359. https://doi.org/10.1016/j.procs.2015.08.415.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075.

Rossi, L. A., & Gnawali, O. (2014). Language independent analysis and classification of discussion threads in coursera mooc forums. In *Proceedings of the 2014 IEEE 15th international conference on information reuse and integration (IEEE IRI 2014), pp. 654–661. IEEE*. https://doi.org/10.1109/IRI.2014.7051952.

Rustia, R. A., Cruz, M. M. A., Burac, M. A. P., & Palaoag, T.D. (2018). Predicting student's board examination performance using classification algorithms. In *Proceedings of the 2018 7th international conference on software and computer applications, pp. 233–237. ACM*. https://doi.org/10.1145/3185089.3185101.

Sachin, R. B., & Vijay, M. S. (2012). A survey and future vision of data mining in educational field. In *2012 second international conference on advanced computing & communication technologies, pp. 96–100. IEEE*. https://doi.org/10.1109/ACCT.2012.14.

Schuck, S., Kearney, M., & Burden, K. (2017). Exploring mobile learning in the third space. *Technology, Pedagogy and Education*, *26*(2), 121–137. https://doi.org/10.1080/1475939X.2016.1230555.

Shah, D. (2020). By The Numbers: MOOCs in 2020. https://www.classcentral.com/report/mooc-stats-2020.

Shah, D. (2021). Massive List of MOOC providers around the world, https://www.classcentral.com/report/mooc-providers-list.

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2Nd international conference on learning analytics and knowledge. LAK '12, pp. 252–254. ACM, New York*. https://doi.org/10.1145/2330601.2330661.

Sonwalkar, N. (2013). The first adaptive mooc: A case study on pedagogy framework and scalable cloud architecture—part i. In *MOOCs Forum, vol. 1, pp. 22–29. Mary Ann Liebert*. https://doi.org/10.1089/mooc.2013.0007.

Sun, G., Cui, T., Yong, J., Shen, J., & Chen, S. (2015). Mlaas: a cloud-based system for delivering adaptive micro learning in mobile mooc learning. *IEEE Transactions on Services Computing*, *11*(2), 292–305. https://doi.org/10.1109/TSC.2015.2473854.

Talavera, L., & Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In *Workshop on artificial intelligence in CSCL. 16th European conference on artificial intelligence, pp. 17–23*.

Tarus, J. K., Niu, Z., & Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing*, *22*(8), 2449–2461. https://doi.org/10.1007/s00500-017-2720-6.

Teodorov, G., Kir, O., & Zheliazkova, I. (2011). Correlation analysis of educational data mining by means a postprocessor's tool.

Urh, M., Vukovic, G., Jereb, E., & Pintar, R. (2015). The model for introduction of gamification into e-learning in higher education. *Procedia - Social and Behavioral Sciences*, *197*, 388–397. https://doi.org/10.1016/j.sbspro.2015.07.154, 7th world conference on educational sciences.

Vaessen, B. E., Prins, F. J., & Jeuring, J. (2014). University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, *72*, 196–208. https://doi.org/10.1016/j.compedu.2013.11.001.

Valiathan, P. (2002). Blended learning models. *Learning circuits*, *3*(8), 50–59.

Vora, D. R., & Iyer, K. (2018). Edm–survey of performance factors and algorithms applied. *International Journal of Engineering & Technology*, *7*(2.6), 93–97. https://doi.org/10.14419/ijet.v7i2.6.10074.

Wang, Y., Kung, L., & Byrd, T.A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019.

Wangmo, C., & Ivanova, G. (2017). Analysis and design of e-learning system prototype for lifelong learning in bhutan. Applied Researches in Technics, Technologies and Education, vol. 5, https://doi.org/10.15547/artte.2017.04.007.

Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all—a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, *1*(1), 4. https://doi.org/10.1145/1515693.1515696.

Wende, K., & Otto, B. (2007). A contingency approach to data governance. In M. A. Robert, R. O'Hare, M. L. Markus, & B. Klein (Eds.) *Proceedings of 12th international conference on information quality, Cambridge, pp. 163–176*. https://www.alexandria.unisg.ch/213308/.

Xiao, J., Wang, M., Jiang, B., & Li, J. (2018). A personalized recommendation system with combinational algorithm for online learning. *Journal of Ambient Intelligence and Humanized Computing*, *9*(3), 667–677. https://doi.org/10.1007/s12652-017-0466-8.

Xing, W., & Du, D. (2019). Dropout prediction in moocs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, *57*(3), 547–570. https://doi.org/10.1177/0735633118757015.

Zhang, H., Zhang, J., Yin, X., Zhou, K., & Pan, Z. (2020). Cloud-to-end rendering and storage management for virtual reality in experimental education. *Virtual Reality & Intelligent Hardware*, *2*(4), 368–380.

Zhu, Z.-T., Yu, M.-H., & Riezebos, P. (2016). A research framework of smart education. *Smart Learning Environments*, *3*(1), 4. https://doi.org/10.1186/s40561-016-0026-2.