Journal of
Distance Learning and Open Learning
Print ISSN : 2314-8829 Online ISSN2314-8837

An insight and paradigm in Open and Distance Education in the Era of Industrial 4.0

Dewi Juliah Ratnaningsih

Jl Cabe Raya Pondok Cabe Pamulang, Ciputat

djuli@ecampus.ut.ac.id

Abstract:

In the digital and industry 4.0 era, data is very valuable precious. All sectors, both industry and education need data as a source of information and insight into improving quality. Technology and information systems are growing so that the amount of data that can be processed is very large and fast. Those phenomena are called big data. In the pandemic era, the use and utilization of big data in education are very significant and determine the learning process. To obtain meaningful information in decision-making, big data analytics is needed. In addition, another important aspect is data science. As a distance education institution, Universitas Terbuka has a high opportunity to apply and offer data science courses. The prospect of being a data scientist is very high and it occupies a prestigious position in the modern world. This article explains big data, big data analytics, and data science and its prospects at Universitas Terbuka. These three parts are interconnected and needed in the digital and industry 4.0 era.

Keywords: Big data, data science, open distance learning, online learning, online tutorial, Universitas Terbuka.

INTRODUCTION

We are currently entering the era of big data or big data. Big Data is an important aspect of productivity, innovation, and global competition that gets a lot of attention from both academics and practitioners (Manyika et al, 2011). With the growing use of technology and information, the phenomenon of big data in everyday life is something we cannot deny. According to the National Security Agency, the internet can process 1,826 petabytes (PB) of data per day. The Google search engine can process more than 40,000 searches every second or 3.5 billion searches per day. Facebook users can upload 300 million photos, 510,000 comments, and 293,000 status updates per day (Gantz & Reinsel, 2012). Assert that by 2020, more than 40 Zettabytes (or 40 trillion gigabytes) of data will be generated, emulated, and consumed every day.

We can no longer avoid Big Data, and this also applies to the world of education. Amid the pandemic that has hit the world since the end of 2019 until now, the distribution and size of data in various media have increased sharply. Apart from that, another impact is on the learning system. All learning is done online, so the use of the internet has become a necessity and the data processed is getting bigger, reaching 1,826 PB or more per day. All learning from elementary to tertiary level has shifted from face-to-face to online learning.

Different from the Open University in Indonesia that implements a Distance Learning system. From its inception, the Open University has implemented a Distance Learning system. In the Distance Learning system, the learning process can be carried out through various modes, both synchronous and asynchronous. The impact of pandemic has had quite a significant impact on the Open University, namely that it has to improve synchronous learning support services which were previously carried out face-to-face. A learning support service developed during the pandemic that is synchronous is a tutorial via the web. Asynchronous learning assistance services known as online tutorials have increased during the pandemic with the increase in the number of students.

Tutoring at the Open University uses the Learning Management System (LMS) platform. With the increasing number of students, the volume, speed, and various variables and data structures processed in the LMS are very large, so the role of big data is very necessary. Apart from that, concerning big data, data science is needed which can translate the data into a form or information that is easier to understand. Through scientific data, information, and insights, data products can be packaged well so that they produce maximum impact and can be used as recommendations in decision-making. Therefore, the role of data science in translating big data is very necessary. This article will explain the concept of big data and data science in the world of education which can be used as a paradigm and insight at the Open University. Through this article, it is hoped that can open the insight of readers and decision-makers at the top leadership level to be able to improve learning assistance services to Open University students precisely, quickly, and accurately. In turn, educational outcomes, namely graduates who have competencies equivalent to other universities, can be realized and public recognition of the Open University will increase.

BIG DATA, INDUSTRIAL ERA 4.0, AND DATA SCIENCE

Big Data

Data always plays an important role in decision-making. With advances in technology, the provision and dissemination of digital data are very rapid and available in almost all fields, including in the field of education. The availability of data through digital networks, social networks, mobile devices, and Wireless Sensor Networks (WSN) has caused the volume of data to increase drastically (Snijders et al., 2013; Ward & Barker, 2013). This is what led to the emergence of the term big data or Big Data (BD).

Big Data is defined as “a new generation of technologies and architectures designed to extract value for large, wide-variety data sets with high-speed data capture, discovery, and analysis (Lazer et al., 2014; Dede et al., 2016). Big Data describes phenomena involving complex and dynamic data growth. The definition of BD from several references varies. However, in essence, the meaning of BD refers to the structural and functional dimensions of data. The BD dimension includes 5 (five) main elements, namely: volume, speed, truth, variety, verification, and value (Poulovassilis, 2016; Manyika et al, 2011). Figure 1 presents the general characteristics of BD according to (Manyika et al, 2011). Below is an explanation of the five characteristics of BD.

Volume refers to the large amount of data generated every second and applies to the size and scale of a data set. The BD size is if the data size is in the Exabyte (EB) or Zettabytes (ZB) range (Chen & Zhang, 2014; Vajjhala & Strang, 2015). However, at this data size, there are still some challenges for datasets in the smaller size range. For example, Walmart collects 2.5 PB of more than one million customers every hour (Milicevic et al., 2017). In the world of education, for example, the data capacity in the LMS, or what is called records is becoming larger. This issue focuses on data processing capacity and also the system's ability to process it.

Variety refers to different forms of data in a data set including structured data, semi-structured data, and unstructured data. Structured data (e.g., stored in relational databases) is mostly well-organized and easy to sort, but unstructured data (e.g., text and multimedia content) is random and difficult to analyze. Semi-structured data (e.g., NoSQL databases) contain tags to separate data elements (Gandomi & Haider, 2015; Pokorny et al., 2015).

Difficulties that can occur with this variety of data structures are if we want to convert between different data types (for example, from unstructured data to structured data), and in mixed-type data. Therefore, in this case, a data mining algorithm is needed that is designed to consider well-formatted input data. This can be done to handle incomplete data and/or different data input formats (Tsai et al., 2015).

Velocity consists of speed in data processing. Speed here is represented in batch, real-time, and streaming form. Velocity emphasizes that the speed of processed data must meet the speed of the data produced and must be accurate (Chen & Zhang, 2014). Delays in data processing can result in incorrect analysis and fatal consequences. For example, in the medical field. If the data processing speed is inadequate, resulting in inaccurate sensor analysis resulting in incorrect information, it will result in injury or even death.

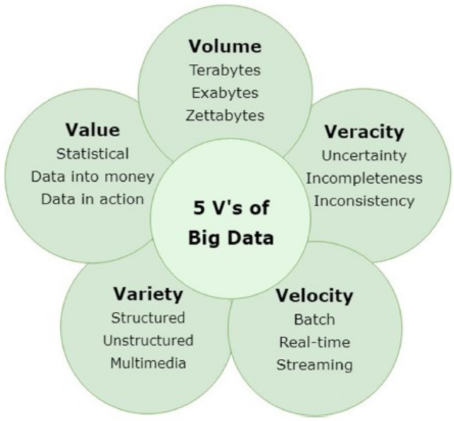

Figure 1.
General characteristics of big data.

Likewise in education, for example, if the processing of student registration data is inaccurate, it can cause errors in determining student academic or exam schedules. Another possibility is that when processing data, execution fails because the data is very large and the application is less capable so the data processing results fail to be sent on time. Therefore, velocity is very important in BD. Veracity represents the quality of data. For example, data that is uncertain, incompleteness, and inconsistency. Data quality greatly influences the losses or risks experienced by the organization. The more diverse the data, and data structures, the more difficult it becomes to establish accuracy and trustworthiness in big data analytics (BDA). For example, in the case of determining disease trends, any inconsistent and incomplete data from an existing data set will result in an invalid analytical process, resulting in an erroneous analysis.

Value represents the context and usefulness of data in decision-making. For example, Facebook, Google, and Amazon have leveraged the value of BD through analytics for their products. Likewise, Amazon analyzes a set of BDs to see the sales and purchasing patterns of their products. From the results of this analysis, they can provide recommendations on strategies for increasing product sales and how to capture consumer segmentation in purchasing their products. Likewise, Google can collect location data from Android users to improve location services on Google Maps. Facebook monitors user activity to target ads and recommend them to other users. Through the values that BD has, a company or organization can photograph and retrieve information that is useful for making better decisions (Court, 2015).

Industrial Era 4.0

Industrial terms are familiar in everyday life. In the Oxford dictionary, the industry is "an economic activity related to the processing of raw materials and the manufacture of goods in factories". According to history, the first industrial revolution began towards the end of the 18th century. Naturally, all industrial revolutions were driven by technological advances in manufacturing processes. The development of the industrial revolution from Industry 1.0 to Industry 4.0 is explained in Figure 2. The first industrial revolution (Industry 1.0) was characterized by industrial mechanization with limited production based on the use of oil and steam engines as energy sources. The second industrial revolution (Industry 2.0) was based on the organization of work and the use of electricity to promote mass production. The third industrial revolution (Industry 3.0) is based on the integration of electronic components and information technology in industry to automate production tasks. Today, as a direct consequence of Industry 3.0, we witness the development of intelligent industrial systems thanks to the permanent integration of advanced technologies (Artificial Intelligence, Cloud Computing, Internet of

Things (IoT), Big Data, and Robotics) supporting the emergence of the fourth industrial revolution (Industry 4.0). Industry 4.0 is characterized by the technological fusion of the boundaries between the physical, biological, and digital worlds to design the world as a cyber-physical system (Schwab, 2017).

In the industrial era 4.0, humans are required to have data literacy, technology, and human resource skills (Aoun, 2017). Data literacy is the ability to read, analyze, and use digital data information as recommendations for making better decisions. Technological literacy is the ability to understand mechanical and technological systems in the world of work. Meanwhile, human resource literacy is the ability to interact well, be flexible, and have good character.

According to (Lee et al., 2014), Industry 4.0 is characterized by increased digitalization of manufacturing driven by four factors. These four factors are (1) increasing data volume, computing power, and connectivity; (2) the emergence of analytics, capabilities, and business intelligence; (3) the emergence of new forms of interaction between humans and machines; and (4) improvements in transferring digital instructions to the physical world, such as robotics and 3D printing. The basic principle of Industry 4.0 is the integration of machines, workflows, and systems, by implementing intelligent networks along production chains and processes to control each other independently. Industry 4.0 is also driving a shift from mass production or service provision to customized products and services based on customer needs.

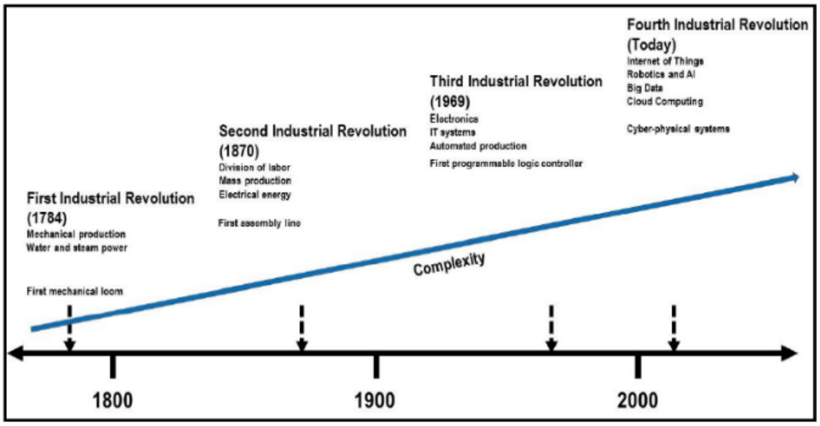

Figure 2.
Development of the industrial revolution.

The key technologies driving Industry 4.0 are wearable devices, augmented reality, simulation, autonomous vehicles and robots, additive manufacturing, distributed ledger systems (such as blockchain), big data analytics, mobile computing, and cloud computing. In addition to these technologies, there are social and economic factors that are also driving the fourth industrial revolution, such as telecommuting, emerging platforms, and more freelance and consultancy-style services, enabled by technology (Rautavaara, 2015; Lu, 2017). Industry 4.0 refers to the digitalization transformation of production or manufacturing-based industries driven by connected technology, namely cyber-physical systems, the internet of things, cloud computing, and cognitive computing (Gaj et al., 2013; Kagermann et al., 2013). The uniqueness of Industry 4.0 can be seen from four perspectives, namely: interconnection, information transparency, technical assistance, and decentralized decisions (Lopez et al., 2017).

Some of the skills needed in the industrial era 4.0 or the era of global competition are critical thinking, problem-solving, communication, collaboration, and creativity (Tobias et al., 2014; Spelt et al., 2009). These skills are often known as “21st century skills” which relate to learning and innovation, digital literacy, and career and life. Apart from professional knowledge and understanding in the world of technology, in the industrial era 4.0, people are needed who can make fast, precise, and accurate decisions.

Higher Education is seen as a forum for developing essential competencies to adopt Industry 4.0 (Buasuwan, 2018). Important aspects of Industry 4.0 competencies are information and communication technology, organizational learning, management, and an innovative environment (Bermúdez & Juárez, 2017; Tan et al., 2017). Innovative management is an important element of Industry 4.0 competency because it promotes creative thinking. In the context of the teaching and learning process in HEIs, innovative management can be seen as the ability to adopt new teaching and learning models (Bermúdez & Juárez, 2017). Such as changing the traditional teaching and learning process towards open-based education through Massive Open Online Courses (MOOCs). MOOCs originate from the ideal of openness in education, that knowledge must be shared freely and the desire to learn must be met without demographic, economic, and geographic constraints, which is in line with the decentralized concept of Industry 4.0.

Data Science

In the digital era and Industry 4.0, data is a very valuable item. Data has a very important role in providing various information needed by modern humans. Through data, a company can improve service quality, and sales strategy and attract customers, so that sales turnover increases (Marr B, 2017). However, data that is available in various forms and varied data structures must be able to be analyzed and translated into meaningful information and can be used as recommendations by policymakers (Davenport & Patil, 2012).

Data science (DS) or data science is a science that is currently on the rise and is built on the disciplines of mathematics, statistics, and computers (Cao, 2017; Weihs & Ickstadt, 2018). The understanding of DS based on several references varies. However, DS is essentially an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from a variety of data, both structured data, semi-structured data, and unstructured data (Dhar, 2013; Stanton, 2012; Provost & Fawcett, 2013). According (Song & Zhu, 2017) suggest that DS can be imagined as an intersection between data mining, big data analytics, artificial intelligence, statistical modeling, and other complex systems capable of transparently evaluating data quality and analytical process results.

Based on several definitions of DS, it can be understood that the essence of DS is that it is not only capable of analyzing numerical or interval/ratio data but also other unstructured data in the form of sound, images, text, and other forms of data. In DS, a combination of practices from computer science, mathematics, and statistics, as well as domain knowledge is necessary. The combination of these disciplines makes DS powerful for processing BD. Graphically, the relationship between the three can be described in Figure 3.

Leek (2013) and Waller & Fawcett (2013) provide an understanding of DS, namely science related to the development and use of tools and processes to extract and differentiate valuable knowledge from very complex data. Davenport & Patil (2012) states that DS is not everything related to complex modeling, nor is it the creation of impressive visualizations and the science of codification. However, DS are things related to making the maximum possible impact for the company. The impact can be in the form of insights, data products, or recommendations for company leaders. Furthermore, Davenport & Patil (2012) stated that DS is a unique combination of skills that can unlock data insights and tell fantastic stories through data. The product of DS is reinterpreting available data for different analytical purposes with diverse data collection techniques. All aspects of DS result in changes in scientific methods, research, and the way society makes decisions (Amato et al., 2017).

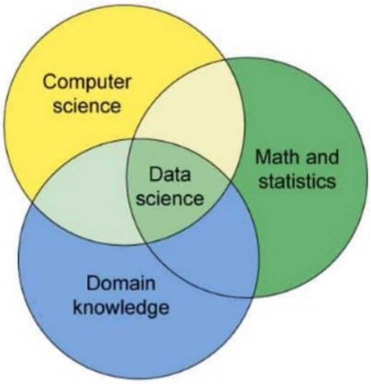

Figure 3.
Relationship between Computer Science, Mathematics, Statistics, and Domain Knowledge.

OPEN AND DISTANCE LEARNING

The Open and Distance Learning system (ODL) is a system that combines the concept of open learning with a distance learning system. In this case, there are two important terms, namely open learning and distance learning (DL) system. The concept of open education is a policy goal or ideal regarding the education system, namely emphasizing the importance of system flexibility to minimize constraints on place, time, and aspects caused by student characteristics such as economic conditions (Bates, 2005).

The term open in an open learning system means being freer from existing boundaries. The ODL system refers more to the system or method of delivering the learning process, the open learning system is related to changes in the structure of educational organizations, becoming an organization that is open in terms of place, time, learning materials, learning systems, and others. Open learning is a mindset and approach used to provide various choices in learning for students, as well as giving students as much control as possible to determine what they will discover and their learning strategies.

Belawati (2019) stated that not all distance learning is open. Some literature shows that the characteristics of open learning must at least contain an element of flexibility including aspects of age (no age limit), location (can be from anywhere), costs (cheap or even free), and length of study (no study limits). time), prerequisites (no need to have a previous education certificate), multi-entry, and multi-exit (can enter and stop at various alternative times/anytime). Furthermore, Belawati (2019) simply describes DL and ODL as slices depicted in Figure 4.

Bozkurt (2019) emphasized that the term ODL is more widely used to reflect a shift towards a more social and learner-centered view of learning that adopts a broader openness to social justice. The ODL is an important element in the education and training system of the future. Currently, the ODL plays an important role in determining the creation of a global knowledge-based society.

According UNESCO (2012), the role of ODL in diversifying education delivery systems, especially for technical and vocational education, encouraging cooperation and partnerships between companies, professional bodies, and distance teaching institutions. Support is also provided to ODL to meet the unique needs of people with disabilities, migrants, cultural and linguistic minorities, refugees, and population crises, which cannot be achieved efficiently by traditional shipping systems. In addition, ODL is deeply embedded in teacher education and training, as well as in teacher education practice.

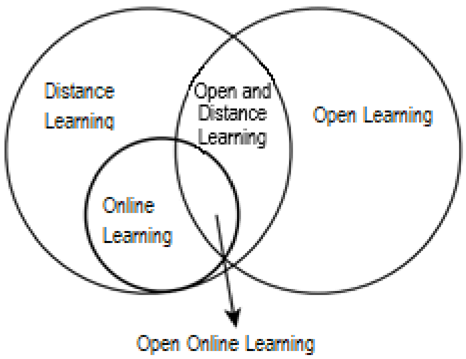

Figure 4.

Distance education, online education, and open and distance education.

One example of the "open" movement in education is reflected in OER (Open Education Resources) and MOOCs (Massive Open Online Courses), which are examples of new developments using various connected technologies. The main idea behind the emergence of the OER movement is that the world of knowledge is public intellectual property that is circulated openly using World Wide Web (www) technology. In other words, the OER movement is also driven by social responsibility, aiming to provide fair and universal access to knowledge through web platforms. Through the OER movement, it encourages the whole world to take the initiative to provide access to various educational resources, including lecture materials and other educational materials.

Two formal definitions are available for OER; one from UNESCO and one from the OECD. According (UNESCO, 2012) defines OER as digital teaching materials for learning and research processes. This OER is packaged in various digital or other media, is in the public domain has been released under an open license, is accessible at no cost, and may be distributed to others without restriction. Meanwhile, in (UNESCO, 2012) explains that OER is digital material offered freely and openly for use by educators, students, and independent students and reused for teaching, learning, and research. Based on this definition, OER places more emphasis on "openness", "digital" format, and "reusability/adaptation" or resources.

Apart from OER, another open movement is MOOC. MOOCs have been called “a major revolution in education” (Bates, 2014). MOOCs are a form of lecture that is offered online and free of charge using technology that can accommodate a very high (massive) number of students. Storme, et al., (2016) stated that MOOCs provide wider access to learning for people who do not have the opportunity to learn in face-to-face lectures. Through MOOCs, the original goals of distance education, social justice, and openness for all, as well as reflecting community-based and socially-based learning approaches will soon be realized. MOOCs are generally pure online learning, but now there are also blended MOOCs such as those organized by The Commonwealth of Learning (CoL) which combine off-line delivery of material via CD with online interaction. In terms of interaction methods, most MOOCs use asynchronous communication methods with a learning approach like in a classroom (class-type design model) with a specific LMS.

Bates (2014) identified several practices in organizing MOOCs, namely: (1) the use of social media (online courses are supported by various 'connected' tools and media), (2) the content of the material is determined by the participants or students who decide and contribute the content material, (3) communication is well distributed (communication is an automatic network with many sub-components), and (4) students decide for themselves whether what they have learned is appropriate for them, Anderson & Simpson (2012) suggest that mobile technology helps in the development of ODL. Mobile technology can improve learning anytime and anywhere without the limitations of space, time, and distance.

BIG DATA CONCEPTS AND BIG DATA ANALYTICS IN OPEN AND DISTANCE EDUCATION

Indonesia is a vast island country, comprising over 17,000 islands that are spread from Sabang to Merauke. The country's unique geographic characteristics pose significant challenges in providing equal access to education. Many areas are difficult to reach, particularly the frontier, remote, and disadvantaged areas. These areas often face obstacles such as infrastructure inequality, limited resources, and lack of access to technology which makes it challenging to provide quality education.

The geographical conditions have significant implications for the country's major problems. Firstly, limited telecommunications infrastructure and internet access make it difficult to access learning materials online, creating a gap in the quality of learning between urban areas, the frontier, remote, and disadvantaged areas. Secondly, the lack of relevant data on educational conditions in , the frontier, remote, and disadvantaged areas makes it difficult for the government and higher education institutions to plan and implement effective learning programs. Thirdly, the limited use of technology in distance learning in the frontier, remote, and disadvantaged areas affects the quality of learning as it causes a lack of direct interaction between teachers and students.

Several literature reviews have emphasized the significance of using Big Data in the educational context. Utilizing Big Data in distance learning allows higher education institutions to gather, analyze, and utilize data to understand patterns and trends in learning. The use of Big Data can personalize learning, provide learning materials that suit students' needs, and monitor and evaluate their progress individually. Therefore, incorporating Big Data and appropriate analysis can be a viable solution that can improve access, quality, and sustainability of learning in the frontier, remote, and disadvantaged areas.

The use of Big Data can be an effective solution in improving the accessibility, quality, and sustainability of distance learning in , the frontier, remote, and disadvantaged areas. By combining data from various sources, including demographic data, internet connectivity, and student learning performance data, distance learning strategies can be developed, taking into consideration the specific needs of each area. Big Data analysis can provide valuable insights to improve the quality of learning by identifying a curriculum that is relevant to local conditions and providing tailored learning content and support to students. The role and use of Big Data in distance learning can also contribute significantly to educational innovation by providing accurate and valid information to determine the novelty and superiority of learning innovations, particularly in the frontier, remote, and disadvantaged areas.

Universitas Terbuka (UT) has taken the initiative to make higher education accessible to all, with its jargon "Reaching the Unreachable" and motto: Making Higher Education Open to All. UT has 39 representative offices in Indonesia and 1 foreign service office unit that serves students in 39 countries. These offices are called Regional UTs, and their distribution is shown in Figure 5. UTs are also present in various countries, as shown in Figure 6.

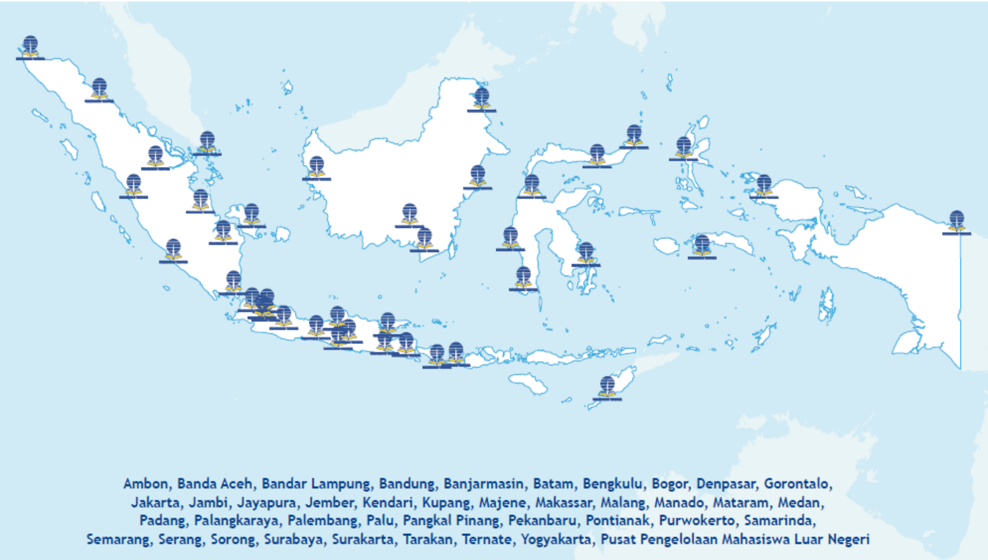

Figure 5.
Distribution of regional UT offices throughout Indonesia.

Figure 6.
Distribution of UT students in the world.

According to a report by Anonymous (2021), the University of Technology (UT) has managed to enroll 525,419 new students so far. These students are distributed among four faculties and a graduate school. The Faculty of Teaching and Education (FTTE) has the highest number of students with 135,027 people, followed by the Faculty of Law, Social and Political Sciences (FLSPS) with 188,822 people, the Faculty of Business Economics (FEB) with 163,905 people, the Faculty of Science and Technology (FST) with 33,386 people, and Postgraduate with 4,279 people. Figure 7 provides a graphical representation of the distribution of students per faculty. Additionally, the university offers approximately 1,363 courses to students at all levels (Anonymous, 2020). With a large number of students managed, courses offered, and various services provided to students, the university has to process a significant amount of data transactions. As a result, the data processing at UT is classified as big data (BD).

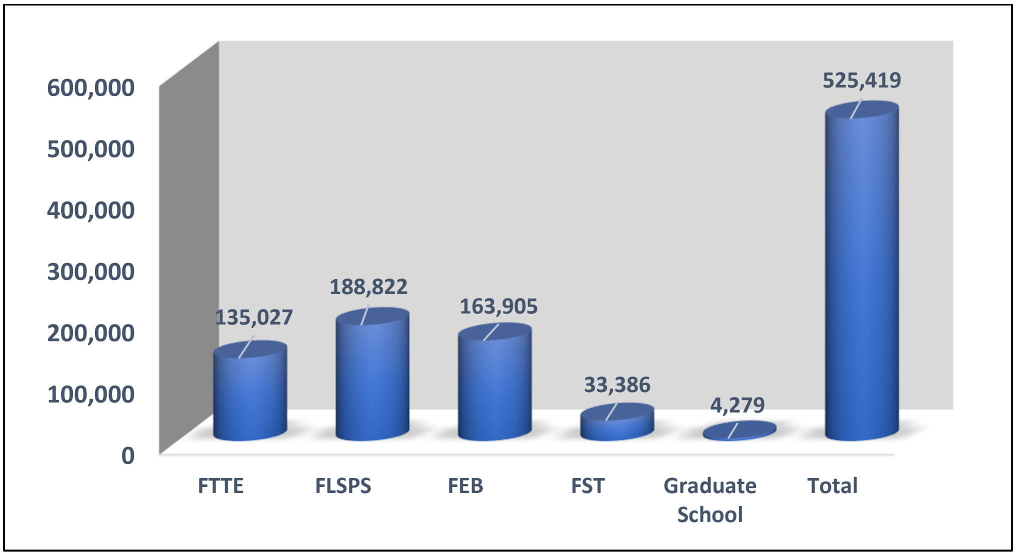

Figure 7.
Distribution of UT students by faculty.

One of the academic services provided by UT to students is tutorials and teaching materials. Some of the tutorial modes provided are synchronous and some are asynchronous. The synchronous tutorials given at UT are in the form of face-to-face tutorials, which during the pandemic were changed to webinar tutorials. Meanwhile, asynchronous tutorials are online tutorials using the LMS platform with Moodle software. In addition, there are also online courses that are provided on a massive and open basis to allow unlimited participation and can be accessed via the web. This type of online course is called Massive Open Online Course (MOOC). For teaching material services, UT provides printed and non-printed (digital) forms in the form of Resources Based View (RBV). Through this service, a certain amount of BD is generated. In this context, the data generated by students and tutors/lecturers during the learning process or when they register is mighty. For example, during the learning process in tutoring, activities are carried out such as inputting student profiles, registration, preferences, discussions, general comments, chat forums, assignments, feedback, assessment of discussions and assignments, interactions that occur between students and students or students and tutors will generate mighty amounts of data. Therefore, in this case, the resulting data category is BD.

The increasingly large volume of data with the increasing number of students encourages learning professionals in DL to look for methods to analyze abundant data sets in mighty volumes. BD conditions cannot be processed using traditional database management tools. Therefore, in this case, adequate BD technology is needed to analyze various information contained in BD in real time for making recommendations to stakeholders and decision-makers. However, in this context, diverse challenges arise. Anshari et al. (2016) said there are three main challenges in BD resulting from online learning, namely: data extraction from unstructured sources, integration of BD with various platforms used, and the need for the role of data scientists (data scientists) who have high skills in mathematics, statistics, and computer science. The role of scientific data is highly expected in interpreting analysis results to provide insights that can help users such as policymakers to make decisions to improve service quality faster and better.

In online learning, as a form of learning assistance service, (Dahdouh et al., 2019) explains the differences in the architectural design of ordinary (traditional) e-learning systems and those that use BD. The architectural design is presented in Figure 8 and Figure 9. From Figure 8 it can be seen that in a traditional e-learning system, there are 3 (three) different roles in the learning process. These role levels are: first, users who connect to the learning platform to learn; second, there is an online learning platform used by the web server; and third the database management system, namely the relationship between data, teaching resources, learner activities, and so on. Meanwhile, the architectural design of an e-learning system using BD is presented in Figure 9. From Figure 9 it can be seen that if the e-learning system uses BD then there are 4 (four) layers of e-learning processing infrastructure. First, the lowest layer is cloud infrastructure. At this layer, the infrastructure built includes virtual computing, storage, and network resources. Infrastructure process conditions like this are expected to provide the flexibility requested by users (Hwang, 2017).

The second layer is the upper BD layer: technologies for data storage, processing, analysis, optimization, visualization, predictive analytics, BD analytics, and BD applications. At this layer, there are several analyses for professional online learning in improving the efficiency and reliability of e-learning platforms. Meanwhile, the e-learning system is in layers consisting of teaching and learning platforms and educational technology (LMS, CMS, VLE, Virtual Classroom, and so on). Information from this layer synthesizes data in the form of content, user information student profiles, and learning registration. This data is critical to adapting educational content to meet the needs of each learner, and to offer a more appropriate learning environment. The final layer of this architecture is the user layer which consists of the administrator, instructor/tutor, and students.

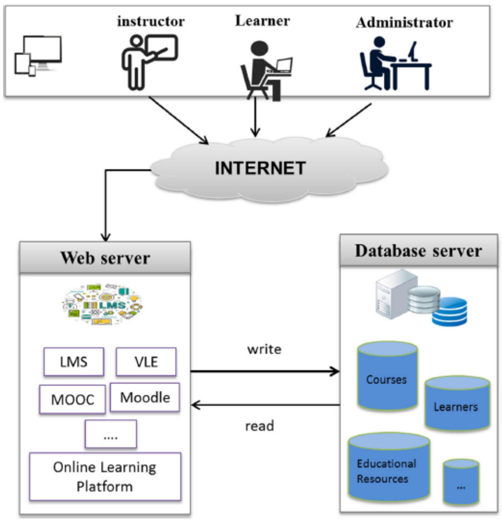

Figure 8.
Traditional e-learning architecture design without big data.

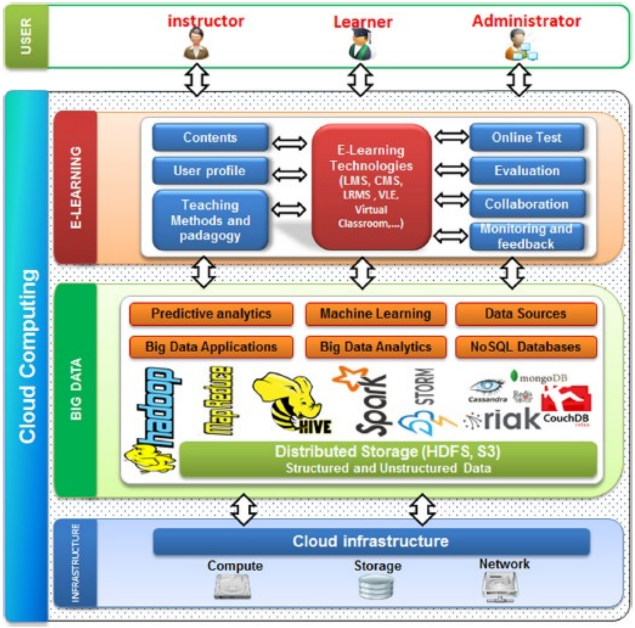

Figure 9.
Traditional e-learning architecture design with big data.

Big Data in education is a new phenomenon (Picciano, 2012), with much of the research discourse centering on the use of data to inform the quality of teaching and research (Eynon, 2013). For example, (Kalota, 2021) suggested the use of BD techniques in education which allows academic institutions to overcome learning problems faced by students and identify student learning strategies. Utilization of BD in education can be done to support the learning, teaching, and administration processes. Daniel (2015) proposed three uses of BD as explained in Figure 10.

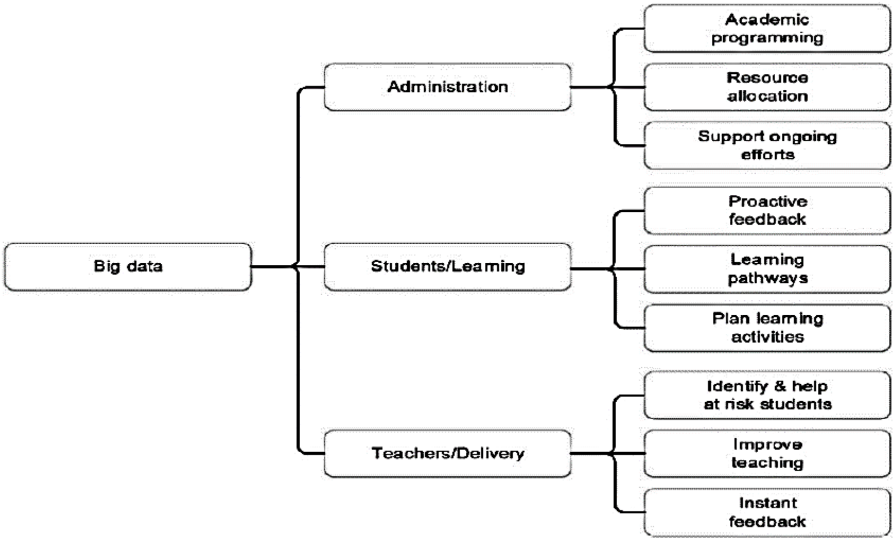

Figure 10.
Big data scenario in education (Daniel, 2015).

The BD scenario proposed by Daniel (2015) can be applied at UT. The tutoring platform at UT uses LMS. Through the LMS, various activities of students and tutors in tutoring can be collected and tracked. In LMS, BD plays an important role in understanding students' learning patterns more effectively. Apart from that, the use of BD and BD analytics can be used to track student learning experiences in tutoring. Student digital track records in the LMS can be analyzed to study student behavior and student success in learning. By analyzing BD collected from the LMS teachers/tutors can assess how well students understand the content and how much content must be modified so that students can understand the material well (Daniel, 2015).

By utilizing the BD concept, BD analytics (Big Data Analytics, BDA) is also needed, namely the entire process of data collection, data storage, data cleaning, and BD analysis. BDA aims to find different patterns/trends in data and other useful information to explore meaningful values so that they can be used for recommendations and decision-making (Fan et al., 2014; Labrinidis & Jagadish, 2012). In this case, an organization/institution needs to use efficient methods to process

BD so that it becomes information that is meaningful and easy to understand (Gandomi & Haider, 2015). The classification analysis method in BDA that can be applied in LMS is presented in Figure 11 (Sivarajah et al., 2016).

From Figure 11 it can be seen that there is an analysis stage in making BD as information that can be extracted into insight which can be used as a decision to make improvements. The three core analyses in BDA can be explained below.

a. Descriptive analysis. This analysis is related to what patterns/trends occur in the data. Descriptive analysis is the initial stage of data processing that creates a set of historical data. Data mining methods in this case help to reveal patterns of data conditions that occur to produce insight. The descriptive analysis provides future opportunities and trends by providing a picture of what might happen in the future (Joseph R.C & Johnson N.A, 2013).
b. Predictive analytics. This analytics relates to forecasting and statistical modeling to predict future events. This type of analysis studies the relationships between variables and then creates a statistical model that can predict the value of new events and future events (Waller & Fawcett, 2013).
c. Prescriptive analytics. This analysis is concerned with optimizing processes, structures, and systems through information generated from predictive analysis. The aim is to determine the appropriate actions to be taken to improve service quality by using optimal funds (Joseph R.C & Johnson N.A, 2013). Prescriptive analysis focuses more on determining the best solutions that need to be taken by the organization/company to achieve the desired targets or goals.

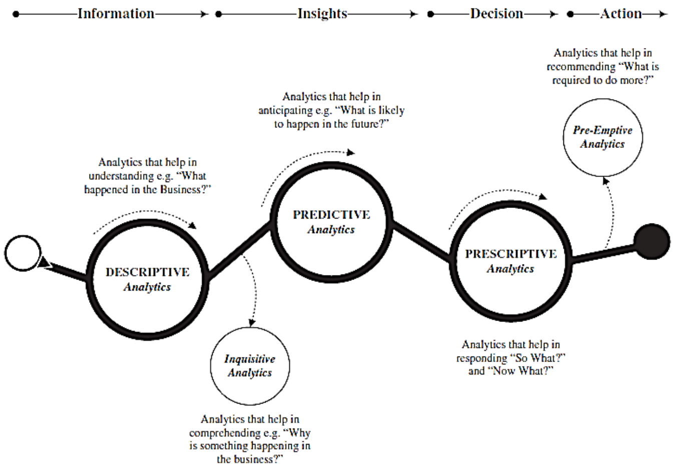

Figure 11.
Analysis Methods in Big Data Analytics.

DATA SCIENCE IN OPEN AND DISTANCE EDUCATION INSTITUTIONS

The condition of data available at UT from various forms of academic services is included in BD. The concept of BD and BDA management needs to be carried out well, precisely, and accurately to produce maximum impact in providing insights or recommendations. As explained in the previous section, DS applications in education, especially UT, can be used to obtain meaningful information to improve the quality of the Tri Dharma of Higher Education. This is in line with what was stated by (Provost & Fawcett, 2013; Milicevic et al., 2017) who say that DS applications in education can produce high-quality benefits.

In an educational institution, data scientists are needed to be able to extract various information as valuable knowledge for theoretical development. In this digital era, the need for DS is increasing. Many universities offer DS programs. Along with the development of the digital era and Industry 4.0, almost all sectors will need data scientists. Data science or DS can become a major scientific discipline that supports key aspects of industrial development. In the future, many industries will need data scientists who are skilled and talented in BD technology so that they are expected to be able to solve problems well, quickly, and accurately.

Davenport & Patil (2012) discussed that data scientists are high-level professionals in making discoveries in the world of BD because there is a huge curiosity about data as key information for business success. A data scientist can explore the information contained in the data communicate the findings obtained and make recommendations for business improvement. In this world, no decision is not supported by data. Data is the basis for innovation and an invaluable treasure for a company/organization/institution. Therefore, learning about data science (DS) at universities is very important.

The DS education across disciplines is very challenging for an educational institution, including UT. DS educational design requires a different approach. One approach that can be taken is cross-disciplinary. With the Independent Learning Campus curriculum, The DS education is very possible and has enormous opportunities. (Pournaras, 2017) suggests that the combination of skills in collaborative research projects that create multifaceted learning opportunities can be used in designing highly challenging DS learning. This is in line with one form of learning arena in the MBKM curriculum. (Pournaras, 2017) further argued that cross-disciplinary data science education meets the requirements of data scientists who are more versatile in the job market. The presence of data scientists in a company/agency/institution can reduce business costs for training and ultimately grow more democratic and participatory citizens who are ready to respond to the challenges that exist in the era of digital society.

One of the skills required in DS is understanding the DS cycle, namely the DS stages that can be used for data problem-solving. The DS cycle proposed by (Spelt et al., 2009) is presented in Figure 12. From Figure 12 it can be seen that there are 8 (eight) important stages in the DS cycle, namely: business understanding, data understanding, data preparation, model planning, model creation, evaluation, implementation, as well as review and monitoring. One of the most important skills in data science is DS project management, which is closely related to the DS cycle in Figure 9. To be successful in DS project management, one must have overall knowledge of BD technology, DS cycle, DS techniques, communication skills, and application. Specifically, one must have the ability to be able to identify business problems, ask specific questions about business problems, identify appropriate data sources, select the right DS tools and analytics platforms, evaluate results, and communicate with stakeholders.

Besides being able to be used to improve the quality of the DS learning curriculum which is aligned with the Independent Learning Campus curriculum, DS can also be used to improve the quality of three pillars of higher education. Through DS, a discovery or theory can be produced, namely by conducting experiments and analysis from the BD collection. Analysis carried out on BD can encourage the discovery of data patterns that can help educational researchers/practitioners in designing better theories and models.

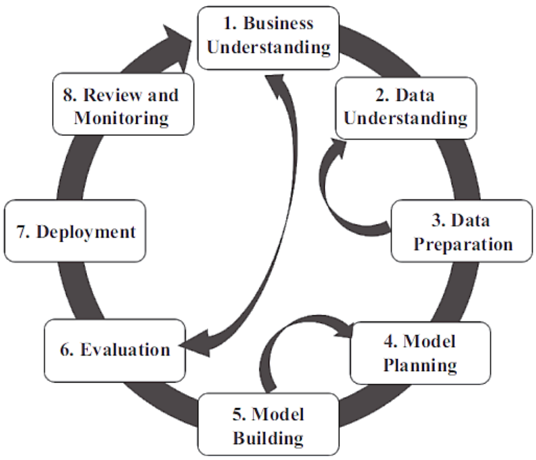

Figure 12.
Data science cycle in solving data problems (Song and Zhu, 2017).

Furthermore, researchers/practitioners can generate a deeper understanding of the complexity of social, economic, biological, technological, cultural, and natural phenomena. Data science or DS can be a determining factor for sustainable development in science, industry, and societal innovation(Traub, 2019). Therefore, DS is a very important opportunity in formulating policies, journalistic data, and marketing strategies in business. Apart from that, the use of application-based DS is very significant in developing infrastructure and services for a modern digital society.

CONCLUSION

The use of technology and information in the digital era and Industry 4.0 has resulted in the data being processed per day being very large and increasing sharply. Large data volumes and very fast processing result in large data conditions. The term big data is known as Big Data. Data is a very valuable treasure and is invaluable as an important source of information in decision-making. Various companies and businesses as well as other industrial sectors are developing rapidly by utilizing BD. Important criteria in BD include volume, variety, velocity, veracity, and value.

In the industrial era 4.0, humans are required to have data literacy, technology, and human resource skills. The six main skills that must be possessed to enter global competition in the digital era and industry 4.0: are critical thinking, creativity, a sense of empathy and compassion, communication skills, the ability to collaborate, and logic in computing. Additionally, key technologies driving Industry 4.0 are wearable devices, augmented reality, simulation, autonomous vehicles and robots, additive manufacturing, and distributed ledger systems (such as blockchain, big data analytics, mobile computing, and cloud computing).

As an institution that organizes Open Distance Learning, it is time for UT to utilize BD in the learning process. The UT learning process which is asynchronous in the form of tutorials using the LMS platform will produce BD. Big Data recorded in the Moodle database can be utilized and analyzed using statistical analysis to produce Big Data Analytics. Apart from that, a very important scientific discipline in the BD era is Data Science. Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from a variety of data, including structured data, semi-structured data, and unstructured data. In DS, a combination of practices from computer science, mathematics, and statistics, as well as domain knowledge is necessary. The combination of these disciplines makes DS powerful for processing BD.

The concepts of BD and DS can be applied at UT. This is very possible considering that the data processed in LMS UT is very large. Through BD and BDA UT can obtain valuable information to improve learning assistance services to students. Apart from that, the DS application in the world of education, including UT, can be used to gain insight and recommendations in improving the quality of the Tri Dharma of Higher Education. Data science or DS can become a major scientific discipline that supports key aspects of industrial development. The opportunity for UT to offer DS courses to produce data scientists is very large. In addition, DS can be a determining factor in sustainable development in science, industry, and societal innovation. The presence of data scientists in a company/agency/institution is very necessary to be able to provide input to decision-makers so that they can have a significant impact on improving quality.

REFERENCES

Anderson, B., Simpson, M. 2012. History and herit*age* in distance education. *Journal of Open, Flexible and Distance Learning, 16*(2), 1-10. Retrieved from http://files.eric.ed.gov/full-text/EJ1080085.pdf.

Anonim. 2020. Katalog Sistem Penyelenggaraan Universitas Terbuka 2020/2021. Tangerang Selatan: Universitas Terbuka.

Anonim. 2021. UT dalam Angka. https://www.ut.ac.id/ut-dalam-angka. Retrieved on Agust 10, 2021.

Anshari, M., Alas, Y., Sabtu, N. I. P. H., Hamid, M. H. S. A. 2016. Online learning: trends, issues and challenges in the big data era. *Journal of E-Learning and Knowledge Society*. 12(1):121–134.

Amato, G., Candela, L., Castelli, D., Esuli, A., Falchi, F., Gennaro, C., Giannotti, F., Monreale, A., Nanni,M., Pagano, P., Pappalardo, L., Pedreschi, D., Pratesi, F., Rabitti, F., Rinzivillo, Diwan, P. 2017. Is Education 4.0 an imperative for success of 4th Industrial Revolution? https://medium.com/@pdiwan/is-education-4-0-an-imperative-for-success-of4th-industrial-revolution-50c31451e8a4. Accessed on July 28, 2021.

Aoun, J. 2017. *Robot-Proof: Higher Education In The Age Of Artificial Intelligence.* Us: Mit Press.

Bates, T. 2005. *Technology, e-learning and distance education*. New York: Routledge.

Bates, T. 2014. *Comparing xMOOCs and cMOOCs: Philosophy and practice*. Retrieved 04 01, 2016, from Online learning and distance education resources: http://www.tony-bates.ca/2014/10/13/comparing-xmoocs-and-cmoocs-philosophy-and-practice/.

Belawati, T. 2019. Pembelajaran Online. Edisi Kedua. Kementrian Riset, Tenlogi, dan Pendidikan Tinggi. Tangerang Selatan: Universitas Terbuka.

Bozkurt, A. 2019. From Distance Education to Open and Distance Learning : A Holistic Evaluation of History ,” no. April, 2019, doi: 10.4018/978-1-5225-8431-5.ch016.

Bermúdez M.D. and Juárez, B. 2017. Competencies to adopt Industry 4.0 for operations management personnel at automotive parts suppliers in Nuevo Leon. Proceedings of the International Conference on Industrial Engineering and Operations Management 2017, pp. 736-747. 2017.

Buasuwan, P. J. A. E. and Studies, D. 2018. “Rethinking Thai higher education for Thailand 4.0,” Asian Education and Development Studies, vol. 7, no. 2, pp. 157-173.

Cao, L.: Data science: a comprehensive overview. ACM Comput. Surv. 2017. https://doi.org/10.1145/3076253.

Chen, C. P., Zhang, C. Y. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny)*., vol. 275, pp. 314–347, 2014, doi: 10.1016/j.ins.2014.01.015.

Court, D., 2015, ‘Getting big impact from big data’, McKinsey Quarterly 1, 52–60.

Dahdouh, K. Dakkak, A. Oughdir, L., Messaoudi, F. 2018. Big data for online learning systems. *Educ Inf Technol*. https://doi.org/10.1007/s10639-018-9741-3.

Daniel, B. K. 2015. Big Data and analytics in higher education: opportunities and challenges. *British Journal of Educational Technology*. 46: 904–920. doi:10.1111/bjet.12230.

Davenport, T.H., & Patil, D.J. 2012. Data scientist: The sexiest job of the 21st century. *Harvard Business Review*. 90(10): 70–76.

Dhar, V. 2013. Data science and prediction. *Communications of the ACM*, 56(12), 64–73.

Dede, C., Ho, A., Mitros, P. 2016. Big Data Analysis in Higher Education: Promises and Pitfalls Citation Terms of Use Share Your Story,” vol. 5, no. 5, 2016, [Online]. Available: http://nrs.harvard.edu/urn-3:HUL.InstRepos:34785368.

Eynon, R. 2013. The Rise of Big Data: what does it mean for education, technology, and media research? *Learning, Media, and Technology*. 38: 237–240. doi: 10.1080/17439884.2013.771783.

Fan, J., Han, F., & Liu, H. 2014. Challenges of big data analysis. *National Science Review*. 1(2): 293–314.

Gaj, P. et al., “Computer communication within industrial distributed environment—A survey,” IEEE Transactions on Industrial Informatics, vol. 9, no. 1, pp. 182-189, 2012.

Gandomi, A., Haider, M. 2015. Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

Gantz, J., Reinsel, D. 2012. The Digital Universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC – EMC Corporation. Online Available at http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf. Retrieved on July 20, 2021.

Hwang, K. 2017. Big-data analytics for cloud, IoT and cognitive learning [electronic resource] (1st ed.). Wiley.

Joseph, R. C., Johnson, N. A. 2013. Big data and transformational government. *IT Professional*. 15(6): 43–48.

Kagermann, et al., "Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry," Final report of the Industries 4.0 Working Group, 2013.

Kalota, F. 2021. Applications of Big Data in education. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*. 9: 1602–1607. doi: 10.1016/c2019-0-01296-4.

Labrinidis, A., & Jagadish, H. V. 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*. 5(12): 2032–2033.

Lee J, Bagheri B, Kao H. 2014. A. Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. *Int Conf Ind Informatics*. doi:10.13140/2.1.1464.1920.

Lopez, MM. et al. 2017. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT". Sensors, vol. 17, no. 9.

Lu Y. 2017. Industry 4.0: A survey on technologies, applications and open research issues. *J Ind Inf Integr*. 6: 1-10. doi:10.1016/j.jii.2017.04.005.

Lazer, D., Kennedy, R., King, G., Vespignani, A. 2014. The parable of Google Flu: traps in big data analysis. *Science*. 343, 1203–1205.

Leek, J. 2013. Six types of analyses every data science should know. Data scientist Insights blog. Retrieved July 12, 2021, from http://datascientistinsights.com/2013/01/29/six-types-of-analyses-everydata-scientistshould-know.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. 2011. Big data: The next frontier for innovation, competition, and produc-tivity. McKinsey Global Institute. Retrieved on July 21 from http://www.citeulike.org/group/18242/article/9341321Pang.

Marr, B. 2017. Really big data at Walmart: real-time insights from their 40+ Petabyte data cloud. https ://www.forbes.com/sites /berna rdmar r/2017/01/23/reall y-big-data-at-walma rt-real-time-insig hts-from-their -40-petab yte-datacloud /#2a0c1 6916c 10.

Milicevic, A.K., Ivanovis, M., Budimac, Z. 2017. Data science in education: Big data and learning analytics," *Comput. Appl. Eng. Educ.*, vol. 25, no. 6, pp. 1066–1078, 2017, doi: 10.1002/cae.21844.

Picciano, A. G. 2012. The evolution of big data and learning analytics in american higher education," *J. Asynchronous Learn. Netw.*, vol. 16, no. 3, pp. 9–20, 2012, doi: 10.24059/olj.v16i3.267.

Poulovassilis, A. 2016. Big Data and education. Technical Report BBKCS-16–01. Retrieved July 28, 2021, from http://www.dcs.bbk.ac.uk/research/techreps/2016/bbkcs-16-01.pdf.

Pokorný, J., Škoda, P., Zelinka, I., Bednárek, D., Zavoral, F., Kruliš, M., Šaloun, P. 2015. Big data movement: a challenge in data processing. *In: Big Data in complex systems. Cham: Springer*. 29–69.

Pournaras, E. 2017. Cross-disciplinary higher education of data science – beyond the computer science student. Data Science. 1: 101–117. doi: 10.3233/DS-170005.

Provost, F, Fawcett, T. 2013. Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59. doi: 10.1089/big.2013.1508.

Rautavaara E. (2015). Educating the Future Product Designers - Exploring the anatomy of a project-based capstone course.

Sivarajah, U., et al., 2016. Critical analysis of Big Data challenges and analytical methods, *Journal of Business Research*. http://dx.doi.org/10.1016/j.jbusres.2016.08.001. doi: 10.1016/j.jbusres.2016.08.001.

Schwab, K. 2017. The Fourth Industrial Revolution: what it means, how to respond (2017). https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-itmeans-and-how-to-respond/. Accessed on July 28, 2021.

Snijders, C., Matzat, U., Reips, U. D. 2013. Big Data': Big Gaps of Knowledge in the Field of Internet Science," *Int. J. Internet Sci.*, vol. 7, no. 1, pp. 1–5.

Song YI, Zhu Y. 2017. Big Data and Data Science: Opportunities and Challenges of iSchools. *Journal of Data and Information Science*. 2(3): 1-18. doi: 10.1515/jdis-2017-0011.

Spelt, EJ, et al. 2009. "Teaching and learning in interdisciplinary higher education: A systematic review," Educational Psychology Review, vol. 21, no. 4, p. 365. doi: 10.1007/s10648-009-9113-z.

Stanton, J. (2012). An introduction to data science.http://ischool.syr.edu/media/documents/2012/3/datasciencebook1_1.pdf. Accessed on Agust 1.

Storme, T., Vansieleghem, N. Devleminck, S., Masschelein, J., Simons, M. 2016. The emerging pedagogy of MOOCs, the educational design of technology and practices of study. J. Comput. Educ. 3(3):309–328. doi: 10.1007/s40692-016-0070-5.

Tan, J. P. L., Koh, E., Jonathan, C. R., Yang, S. 2017. Learner dashboards a double-edged sword? Students'sense-making of a collaborative critical reading and learning analytics environment for fostering 21st century literacies. *Journal of Learning Analytics*. 4, 117–140.

Tobias, S, et al., "Game-based learning. In Handbook of research on educational communications and technology," Springer, 2014.

Traub, J., Quiané-Ruiz, J., Kaoudi, Z., Markl, V.: Agora: Towards an open ecosystem for democratizing data science&artificial intelligence. CoRR abs/1909.03026 (2019). arxiv: 1909.03026.

Tsai, C.W., Lai, C.F., Chao, H.C., Vasilakos, A,V. 2015. Big data analytics: a survey. *J Big Data*. 15;2(1):21. doi: 10.1186/s40537-015-0030-3.

UNESCO .2012. *2012 Paris OER Declaration.* Paris: UNESCO. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/English_Paris_OER_Declaration.pdfUT dalam angka. 2021. www.ut.c.id.

Vajjhala, N.R., Strang, K.D., Sun, Z. 2015. Statistical modeling and visualizing open big data using a terrorism case study. In: 3rd international conference on future Internet of things and cloud (FiCloud), 2015. IEEE. 489–96.

Waller, M. A., Fawcett, S. E. 2013. Data science, predictive analytics, and Big Data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*. 34, 77–84.

Ward, J. S., Barker, A. 2013. Undefined by data: a survey of Big Data definitions. CoRR, arXiv:1309.5821.

Weihs C, Ickstadt K. 2018. Data Science: the impact of statistics. *International Journal of Data Science and Analytics*. 6:189–194. https://doi.org/10.1007/s41060-018-0102-5.