

The Role of Big Data and Machine Learning in COVID-19

Mustafa ABABNEH¹,
Aayat ALJARRAH²,
Damla KARAGOZLU³

¹Faculty of Computer Information System,
Near East University, Mersin 10, Turkey,
20194017@std.neu.edu.tr

²Faculty of Computer Information System,
Near East University, Mersin 10, Turkey,
20194007@std.neu.edu.tr

³Faculty of Computer Information System,
Near East University, Mersin 10, Turkey,
damlakaragozlu@neu.edu.tr

Abstract: *The big rise in the existence of digital data contributed to creating many good chances, especially related to corporations, institutions and firms. Also, it gives the capability to scrimp data regarding its major or area, where the countries have benefited from the analysis of big data (BD) greatly in the face of epidemics and diseases, especially COVID-19 since BD is now available everywhere around us, from official reports and scientific studies related to virology and epidemiology. The general aim of this study is to clarify how the conjunction among both BD and machine learning (ML) created huge differences in data science and a big influence on the applications related to a lot of fields chiefly in COVID-19. The method which is used in this study 'relevance tree' by identifying papers related to ML and BD, especially in COVID-19. The results have been shown that the use of reinforcement learning in analyzing BD provides effective and tremendous results, although it faces many challenges and restrictions that have been explained in detail in this study. In addition, the results showed that most of the countries in the time of Corona turned into smart cities, totally dependent on smart applications based on the analysis of BD using ML, and one of the most important applications that were circulated around the world global positioning system. In addition to the results that have been found, data privacy is one of the most important challenges facing data analysis. Consequently, it recommended future researchers to focus on studying the challenges faced by ML in analyzing medical data in the COVID-19 era.*

Keywords: *Big Data; Machine Learning; Artificial Intelligence; COVID-19.*

How to cite: Ababneh, M., Aljarrah, A., & Karagozlu, D. (2020). The Role of Big Data and Machine Learning in COVID-19. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 11(2Sup1), 01-20.
<https://doi.org/10.18662/brain/11.2Sup1/89>

1. Introduction

In today's era of information and technology, the rapid explosion of data is quickly spreading everywhere, such as websites, social applications and smartphones in general. This huge role of data enforces us to think about the algorithms related to online learning and its frameworks. However, to pick an appropriate tool is something very sensitive and hard. For example, this new learning system needs a lot of preparations and requires many methods. This new globe has depended on data which became the main source to acquire knowledge and experience among people and countries. This new type of information has caused a digital war, especially when it comes to organizing information and protecting it. Therefore, the concept of big data (BD) becomes very popular to show both the arrangement of data and its big existence at random. The complexity of this BD obligates us to face all the difficulties. Relying on four kinds of Vs: velocity, variety, volume and veracity, we can consider them as the basic characteristics. Compared to other approaches such as traditional ways of computing, the method faces many problems where some examples are De Mauro et al. (2016) and Barker and Ward (2013). It is undeniable that this topic took the attention of researchers, although it has some problems such as access, process and storage. For instance, to know more about this method, researchers can use some works related to the works by Gandomi and Haider (2015), Khan et al. (2014) and Oussous et al. (2018).

The new technological devices create new types of developed approaches, such as machine learning (ML) and other file systems. To illustrate, Hadoop makes it easy to spread ML in outer libraries like the library of the Scikit Teaching resulting in controlling and mastering data (Shvachko et al., 2010). In these libraries, the methods of ML mostly depend on being classified in specific algorithms that are not suitable for BD. Nevertheless, other methods are used very well for BD in order to enhance the teaching such as deep learning that proved its skills.

The perfect field for BD is its participation in ML research (Al-Jarrah et al., 2015; Ma et al., 2014; Zhou et al., 2017). Both fields share the same major in computer science and data science. Both fields work side by side and combine their usefulness. The main goal of BD is to collect data, store it and analyze it. It aims to discover the invisible patterns and to help in making decisions. However, ML is done with the help of BD by analyzing it with algorithms that help in computing, and taking into consideration the

previous experiments as mentioned by Zhou et al. (2017); many interests are shown towards ML by the support of BD.

The catastrophe of COVID-19 is considered as an emergency case that needs help from both national and international sectors. According to world reports, it caused more than 5 million infected cases with about 300,000 dead people almost in a hundred cities, precisely over May 2020 (NHC, 2020; WHO, 2020). This epidemic had a dangerous impact on both fields: social and economic improvements. In addition, on the 28th of February, the UN Secretary-General Guterres asked the governments, all over the world, to take steps towards controlling this virus, especially by making BD analysis (New. Cn, 2020). Therefore, to keep the life moving, a lot of companies, universities and research teams tend to build many information systems, for instance, 'Fever Clinical Queries', 'Passenger Information Queries' and 'Epidemic Map Displays'. They were built relying on a commercial kind of software that provided significant contributions to stop this virus, basically by depending on the artificial intelligence like ML (CAICT, 2020).

The huge increase in making data digitally participated is creating a huge number of chances. Firms and institutions can minimize data regardless of its area or field. The integration between ML and BD has caused a big difference in data science which was easier to work a lot of applications in many majors. This paper is presented in order to get a full reflection of the new research papers related to this topic. The aim of this review to clarify the difficulties faced by many ML methods in order to produce an affordable framework that can easily fit in the sector of BD analysis. In addition, it clarifies the role of BD and the importance of using ML in its analysis in defeating the COVID-19. We wish that the study can present the significance of ML in BD analysis in order to pass all the difficulties that it faced.

2. Related Studies

To solve problems in artificial intelligence, especially teaching without programming, ML is the best choice (Goodfellow et al., 2016; Murphy, 2012; Shalev-Shwartz & Ben-David, 2014). This multi major has dominated our life from all aspects. Many algorithms were put in ML in order to be precisely presented in three types: supervising, half-supervising, non-supervising and reinforcing learning. The methods of this major are depending on BD to participate in completing other sciences like engineering (Yi et al., 2014). It took the attention of most of the researchers increasingly from 2012, especially in science such as mathematics and

computer science in addition to statistics where its publishing has increased in the year 2016. The scientific research studies use tools that were provided by the Bibliometric (Cadez, 2013; Sweileh et al., 2017; Van Eck & Waltman, 2009). This new method proved its efficiency in helping decision-makers, researchers, and other jobs to find a good view of all the wanted sides. Some useful research was done especially related to the Bibliometrics specialized in big data (Singh et al. 2015; Mishra et al. 2016; Sivarajah et al. 2017). Also, some other majors too got use of the big data like; industry 4.0 (Muhuri et al., 2019), IOT publishing (Nobre & Tavares 2017), technological maps (Gerdri et al., 2013), cloud computing (Heilig & Voß 2014), IT (Khaparde & Pawar, 2013), data mining or a big one (Tseng et al., 2016), and other fields like cybersecurity and machine learning (Makawana & Jhaveri 2018).

Several studies and research studies have worked to clarify the importance of using ML of all kinds in analysing BD in all health, scientific, economic and political fields, and to develop our paper, many previous studies and research studies will be shown related to both BD analysis and ML.

Table 1 provides a summary of the existing studies regarding ML for data analysis in BD.

Table 1. The aims and results of articles related to ML for data analysis in BD.

Authors and Year	Aim of the Study	Results of the Study
Alloghani et al. (2020)	The main aim of the methodical revision was analysing the research papers which had been published differently during the years 2015 to 2018. It was done to achieve the perfect application for the ML strategies related to finding solutions for many issues in the models.	It can be seen from the research papers, the decision making, the machine of supporting vectors, besides the algorithms of Naïve Bayes that they are considered as the most applied, augmented, constructed and observed in BD for learners. In addition, K_means, clustering hierarchy and the analysis of principal components are also considered as the most popular tools for learners in an uncontrolled or unobserved way.
Qiu et al. (2020)	The SE schemes are given a new definition in a practical shape with the help of safe outer resources of the electrocardiogram (ECG) data	These SE schemes are not considered as the right models where they must be designed relying on a specific ranking of protecting the ECG data. This

	in which the mistrust atmosphere of BSN is relying on the ML.	must be done in order to save it from any unlawful attack. This kind of protection is much required for saving the privacy of the sick person. Thus, many exams and tests were done to prove the efficiency and the practicability of the SE model.
Calderón et al. (2019)	In this study, the relatedness of analysis for both: observed sentimentality and the flow of communicating in public studies are discussed. Also, it reveals the application for the analysis of divided watched sentimentality.	This tool tackles the cons of any other principles related to the computational management of small data, especially in the digital environment such as social media. Also, it gives both communicating processes and public research a just view of the cutting-edge technology that can be applied in the field of social computation.
Cavalcante et al. (2019)	In this study, a crossbred method was created, besides applying the ML technique and mimicry. This applying is checked on the data drivers used in supporting decision-makers precisely in soft selecting providers.	The main result of this study is increasing awareness about both ML and simulation in order to know the way of combining them and the right time to do such a thing. This combination is done to make digital equipping the Gemini series where it works to develop this softness or resilience.
Sughasiny & Rajeshwari (2018)	This study provides a full understanding of the significance of the feature selection techniques, the observed ML tools, the unobserved ML tools, and the BD for the healthiness' fields.	Relying on the exploration of many research papers, a recent model is created in order to expect the strictness and the rigidity of the illnesses through BD, data science, and the ML method.
Mohammadi & Al-Fuqaha (2018)	This paper highlights the difficulties of applying BD produced by ML in smart cities. Also, it describes the process of wasting the unclassified data.	As a result, the study aims on creating a half observed profound learning model that aims to present the difficulties and spotlight on the different fields' applications. Also, many challenges were presented to be

		supported by the most trending fields which can be worthy to make research studies that contain LM for the goal of creating a new smart city serving.
Kibria et al. (2018)	This paper contains the basic drivers for BD analysis by revealing the application process of ML, computational intelligence and AL. They play a very significant role, especially in the data analysis and precisely for the recent models of wireless networks.	The advantages and difficulties of operating the AL, LM and Wireless networks in the BD analysis are all discussed in these results.
Chang et al. (2018)	This paper was created for the ML techniques in order to increase their solitude capacity.	As a result, numerical studies were done to describe the various performances that can probably be verified.

3. The Method

In this section, the literature was reviewed in order to achieve the purpose of this paper, and this was done by identifying papers related to ML and BD in order to identify challenges and obstacles for further studies and research. The method which is used in this study ‘relevance tree’ (Saunders et al., 2009). This method assists to determine which keywords are relevant to the objectives and research question (Saunders et al., 2009). The following databases were researched: Google Scholar, Scopus, Web of Science and Wiley. When searching in databases, the search keywords option was chosen by using advanced search related to research questions and research goals. The applied terms are ‘ML’ combined with the words ‘BD’, ‘artificial intelligence’, ‘Unsupervised ML’, ‘Supervised ML’, ‘Reinforcement ML’ and ‘COVID-19’.

During the search, the terms used were ‘title, abstract and keywords’ search in scientific journal papers. The last year has presented a plethora of studies on ML and BD that debate definitions, scopes, advantages, disadvantages and challenges. When studying the research, the focus was on the following group of topics: the first one is the definition of ML and BD and the second one is BD analysis and finally kinds of ML.

4. Results

In this section, the research questions were answered in detail.

4.1 Machine Learning and Big Data Analysis

The term BD was defined in many steps starting from being defined as ‘volume’. Then, the words ‘velocity’ and ‘variety’ were put to express it. However, later on, it was known as ‘veracity’ (Fan & Bifet, 2012), and the term ‘value’ (Fan & Bifet, 2012; Demchenko et al., 2013) was also added. The definition is not an easy point where it really takes a huge effort, especially that it needs some processes such as making the visibility of detection easier and helping in also making decisions besides helping in data processing. However, the word ‘value’ has properties related to the wanted point which is mainly domineering the BD (Uddin et al., 2014). The goal behind the use of BD became well known, but its results rely on the development of the old ways or the new ones to control this data.

As a part of artificial intelligence, ML includes two steps: ‘practicing’ and ‘experimenting’ (Al-Jarrah et al., 2015). The main step presents the learning methods relying on the famous properties of the datasets. In addition, the second step tends to create some new expectations for the unknown properties relying on the information that was obtained in step number 1. Both the above-mentioned processes, ‘practicing and experimenting’ became known as ‘learning’ and ‘expecting’. In fact, the job of ML aims to use a specific algorithm for learning through creating a specific sample to be exercised in predicting.

Thus, this whole process becomes a matter of prediction (Kolisetty & Rajput, 2020). Recently, a lot of researchers explained the ML difficulties related to BD (Najafabadi et al., 2015; Qiu et al., 2016), whereas others explained that are due to a specific technique (Najafabadi et al., 2015). The algorithms of ML are capable to enhance many types of learning strategies, such as ‘Rule Learning’, ‘Instance-based Learning’, ‘Decision Tree Learning’ and ‘Collective Learning’ (Qiu et al., 2016). The whole concept of algorithms is considered a reflection of the advancement.

4.1.1 Big Data Analysis

Business Intelligence is defined as a tool that is used basically for getting advantage of the BD strategies. These new strategies have already affected our current utilization. For instance, its suitability in realizing its classes' lineaments, the features of the parameters and the observing. All these aspects can help to address any problem facing this new technique.

In this research, Assunção et al (2015) tried to show the enhancement of this strategy and to express the best area to apply the BD in the cloud computing platform.

They classified BD analysis' solutions depending on the previous models of customers relying on the existing data models besides other models resulting in helping the decision-making.

Both personalized and no agreements may create many challenges in BD. Every agreement will extremely affect BD which creates some troubles in the acceptance through using three spaces of dimensions. However, improving the BD in classes asks for dividing some categories, which means that the process of enhancement will be very complicated. Therefore, any rise in the shapes of these levels or classes relies on users' learning and experience. As a result, dividing the BD into classes cannot be expected which makes the application of ML very hard, especially when it comes to the algorithms (Kolisetty & Rajput, 2020).

In addition, the agreement of properties participates in the BD hardness. It is built by dividing the class forms in order to drop down the hardness, especially in the area of data increase of dimensions.

Thus, they are extremely basic found principles that can solve issues related to scalability of BD forms besides its conformity which participates in controlling the data and analyzing it. It will contribute to increasing the size of data and increase the hardness in processing it precisely with all the modern technology used forms (Kolisetty & Rajput, 2020).

However, methods for BD processing data are summarized in some points: Some algorithms have done based on the UCI ML repository that aims to make enhancements like having a very elastic algorithm for having fast unobserved data learning. This algorithm was done by Xiang et al. (2018) using the approach of a two-stage unobserved multiple kernel learning machine. However, this experiment faced many difficulties like the high computational needed overhead (Xiang et al., 2018). On the contrary, using the UCI and biomedical repositories, algorithms can build a true computational form with high efficiency, especially by using the approach of 'Predictive modelling, Decision tree, Bayesian and Instance-based' (Liu et al., 2017). However, its limits are recorded by creating a big variability in data presentation and a huge variety in performances and honesty.

Also, some of the algorithms required solutions, especially for the hard categorized data with relationships of coupling classification and frequency such as the one that was done by Zhu specialized in harmonious metrical learning with couplings classification relying on 30 datasets taken from various fields (Zhu et al., 2018). On the other hand, this kind of

algorithm has faced some limitations too such as its disability towards some data properties and controlled the knowledge.

Finally, other algorithms may work in this area like one done by reading using a strategy of deep learning (Read et al., 2015). This algorithm used real-world types of datasets. They were applied to improve the truthiness of famous existing shapes of data. Unfortunately, it has no obvious explanation of the higher dimensions of datasets when it comes to the properties of both reducing and division of labels.

4.2 Types of Machine Learning

ML is considered as a minor form of artificial intelligence that concentrates on learning methods on computers. This process of learning is classified into three main classes: supervised, unsupervised and reinforcement. In the next texts, those three basic tools are explained in detail besides some famous and common techniques in every division:

4.2.1 Supervise Learning

Referring to its title, in this type of learning, there must be an observer or a supervisor who gives the algorithms of learning their ideas precisely on good or bad decision-making or even actions. In this type, data are completely straining where the tools of learning can know if a specific tool action or decision is right or not with a correct percentage. The common algorithms related to this type are discussed in the following.

- First, the Support vector machine: This tool contributes to finding a volume of N-dimensional hyper inside the N space of dimension in order to divide the whole data sets in a group of N properties (Al-Zoubi et al., 2018).

- Second, the Random forest: This tool consists of multi decision trees that were made and mixed by using a factor of merit in order to provide a true division with the expected percentage (Alian et al., 2014).

- Third, the Neural network: This tool was done from unpretentious neurons organised inside many classes and related to each other through a group of weights. It mimics the way that the natural and biological neurons normally work in a full imitation (Azzini & Tettamanzi, 2011).

4.2.2 Unsupervised Learning

This kind of learning is not categorized which indicates that the algorithm needs to work hard in order to identify itself. The algorithms in this type must know the structures and shapes of data and their relations besides their shapes and properties.

First, the K-means clustering (Barabasi & Albert, 1999) is defined as the process of dividing the data to many groups by algorithms where these clusters have some related properties in common.

- Self-organizing Neural Networks (Chawla, 2009): This is the NN type. This strategy works to arrange neurons for the goal of decreasing any mistake or error in their function designing for each trouble and problem.

4.2.3 Reinforcement Learning

This tool contains an algorithm that distinguishes between both cases correct or wrong, to be awarded in case correct and punished in case of being a false one. Such a kind of learning mimics the same way that creatures or humans precisely get knowledge depending on two types of processes: rewarding and punishing. Here are some examples to make my point crystal clear. First, the Q learning (Chen & Chen, 2014), which depends on Bellman balance and reduces the Q-rate. Second, the Deep Q Network (DQN) (Chen et al., 2005), similar to the previous one, has the same process but with some differences such as the capability of being generalized. Third, Deep Deterministic Policy Gradient (DDPG) (Chowdhury, 2010) is the same as DQN, unlike it can find solutions for troubles that have a continuous space of actions.

4.3 The Restrictions of Big Data Analytics

Applying the BD can include many looking forward wishes. Unfortunately, it is not a method that has indefinite features because to make a lot of analysis means that the limits of such data abilities can be known and taken into considerations (Wang et al., 2015). Here are some limits for done experiments of some users with a data explorer used for the first time.

First, Data Misinterpretation where this data can discover the behaviors of the users. On the other hand, it may not know the reason behind these actions and behaviors. However, the way the data are represented in a wrong way can lead the users in a very wrong direction, especially when it comes to their way of making their jobs and business-like getting used of the beneficial information in business areas. Also, relying on the current data to create a new formula of possibilities can probably lead firms to take steps against the right actions relying on a miss-taken relevance. Thus, to clarify the expected engagement and aiming to find solutions for the true troubles with a highlight of the data supporting can be much different in collecting and explaining the data (Kolisetty & Rajput, 2020). Second, Security Limitations where BD has some challenges related to

security. Firms that gather data hold a really important responsibility to secure and save this data. The results of the data breaking can contain lawsuits, paying fines or gaining a bad reputation. The issues related to security and protection can affect the capability of processing data. For instance, the analysis of data done by any other organization may be very complicated because the data can be affected by a firewall to hide it or any other private server of the cloud ones. This creates many problems, especially related to involving and moving data to be processed in a reliable way (Kolisetty & Rajput, 2020). Third, the Outlier Effect is considered to be the third main part which is common in processing data, especially if the user failed or if he/she searched for something new in the searching engine which makes some partial results. In fact, technology is incapable to gather data fully and truly. Nevertheless, the algorithms related to Google and its limited expectations for research and their results contributed to making this project a failure one (Kolisetty & Rajput, 2020). Fourth, Organizations that possess huge data facing a big challenge, which is the extent to which these organizations are able to control the diverse and unorganized BD, as storing, managing and utilizing these data in an optimal way is a real problem (Sharma et al., 2020).

4.4 The Importance of Machine Learning in Big Data

ML hires an algorithm in order to reveal the undiscovered knowledge but without any programming. Using ML includes having frequent combinations where these new forms and samples aim to highly and separately adapt any insecure or opening to BD. Thus, with the help of technology, especially computers, ML now developed a modern shape better than the past one. Nowadays, ML algorithms have become capable of doing many complicated and hard performances related to computation in order to handle the BD analysis.

To illustrate, the focus of ML in BD is capable of being discovered in: 'Google's self-driving car'. In addition, the ML apps using BD can figure out some recommendations and other systems of business online, such as Netflix and Amazon (Kolisetty & Rajput, 2020). Moreover, it is used in the process of the text data especially in many social media inputs, such as FB and Twitter. Finally, ML is able to process BD in order to expect the discovery of faking processes in specific majors like the financial one and in the privacy or security system (Kolisetty & Rajput, 2020).

4.5 The importance of big data and machine learning in the face of COVID-19

The world nowadays faces huge threats because of virus Covid-19 all over the world. This virus is a danger that threatens the whole of humanity. Thus, to stop it, all countries must cooperate. Also, to conquer this epidemic, it was very significant to transform many overpopulated cities to be 'Smart' ones such as China, Kuria, and other cities by using the smart applications in most fields to keep going in this period (Allam & Jones, 2020).

The use of these applications reduces overcrowding which contributes to limiting the spread of this virus. In addition, it must be mentioned that beating this epidemic can be achieved not only by abolishing it but also by maintaining the continuity of the countries' different fields, such as economic, education and trading (Pandey et al., 2020).

Thus, the significance of BD, its analysis, the apps of artificial intelligence and ML has appeared, where the analysis of BD can be used by ML in many cities around the world to stop the spreading of the virus. For instance, one application that was activated is the tracking systems via GPS which works to stabilize people to avoid the places that contain infected persons (Green, 2020).

In addition, these applications can observe the suspected people that are put in a compulsory home quarantine to ensure their commitment (Engle et al., 2020; Wang et al., 2020). Moreover, ML is used to expect the places where the epidemic has spread in (McCall, 2020). Therefore, because of the importance of investing data and data analysis, the OSTP, Office of Science and Technology Policy, in the White House has launched a huge open-source data centre (CORD19).

Many academic and governmental institutions have participated in it, besides many other organizations specialized in artificial intelligence, national health and dozens of other institutions (Lo et al., 2020). One of these companies that are specialized in analyzing and studying medical data using artificial intelligence, precisely the ML, is the company of BLUEDOT. It is located in the Canadian city: Toronto (Tuite et al., 2020).

Eventually, we notice the significance of BD that the world produces every day. Also, it is obvious that ML is important in analyzing this data and in getting advantages from it, especially in providing better medical services and conquering this catastrophe as well. Moreover, analyzing this data correctly will help presidents and leaders to make correct and reasonable decisions in a suitable time to slow the spread of this virus.

5 Conclusion

The BD analysis was defined as a process of explaining and practicing examining huge datasets. The unstructured data provide a very important chance for most aspects and fields. Nevertheless, a lot of this flatness is not efficient computing: scalable or practical. These study findings showed first that analyzing data using ML is of great importance and represents a great reflection of progress in the future since it contributes greatly to decision-making and second it clarified the role of the algorithms used in each type of ML in data analysis. Third, it showed that of the most common constraints faced data analysis using ML is the wrong analysis or erroneous prediction of data as it produces many major problems. Fourth, the results illustrate the most important ML applications that use BD, and last but not least, BD and ML are able to make a lot of efforts to combat the COVID-19, such as creating interactive dashboards, analyzing epidemiological models and suggesting the best vehicles to help access virus treatments. The most important future recommendations proposed by this research are to find good solutions to the challenges and obstacles faced by ML in analyzing BD such as facilitating the process of accessing the data to be analyzed and reducing the time it takes to retrieve data, finding ways to maintain data privacy and also attention to data quality, governance and management. Also, the focus should be on activating the role of ML more in confronting the COVID-19 epidemic, because of its great effectiveness.

References

- Alian, S., & Ghatasheh, N. (2014). Multi-agent swarm spreading approach in unknown environments. *International Journal of Computer Science Issues (IJCSI)*, 11(2), 160.
- Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. *Big Data Research*, 2(3), 87-93. <https://doi.org/10.1016/j.bdr.2015.04.001>
- Allam, Z., & Jones, D. S. (2020, March). On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. *Healthcare*, 8(1), 46. <https://doi.org/10.3390/healthcare8010046>
- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. Berry, A. Mohamed, & B. Yap (Eds.),

- Supervised and Unsupervised Learning for Data Science* (pp. 3-21). Springer, Cham. https://doi.org/10.1007/978-3-030-22475-2_1
- Al-Zoubi, A. M., Rodan, A., & Alazzam, A. (2018, November). Classification model for credit data. In *2018 Fifth HCT Information Technology Trends (ITT), Dubai, United Arab Emirates, 2018* (pp. 132-137). <https://doi.org/10.1109/CTTT.2018.8649549>
- Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79, 3-15. <https://doi.org/10.1016/j.jpdc.2014.08.003>
- Azzini, A., & Tettamanzi, A. G. (2011). Evolutionary ANNs: a state of the art survey. *Intelligenza Artificiale*, 5(1), 19-35. <https://doi.org/10.3233/IA-2011-0002>
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512. <https://doi.org/10.1126/science.286.5439.509>
- Cadez, S. (2013). Social change, institutional pressures and knowledge creation: A bibliometric analysis. *Expert systems with applications*, 40(17), 6885-6893. <https://doi.org/10.1016/j.eswa.2013.06.036>
- China Academy of Information and Communications Technology (CAICT). (2020). Research Report on Data and Intelligent Application in Epidemic Prevention and Control (1.0). <http://www.caict.ac.cn/kxyj/qwfb/ztbg/202003/P020200305495005485729.pdf>
- Calderón, C. A., Álvarez, M., & Mariño, M. V. (2019). Distributed Supervised Sentiment Analysis of Tweets: Integrating Machine Learning and Streaming Analytics for Big Data Challenges in Communication and Audience Research. *Empiria: Revista de metodología de ciencias sociales*, 42, 113-136. <https://doi.org/empiria.42.2019.23254>
- Cavalcante, I. M., Frazzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. *International Journal of Information Management*, 49, 86-97. <https://doi.org/10.1016/j.ijinfomgt.2019.03.004>
- Chang, Z., Lei, L., Zhou, Z., Mao, S., & Ristaniemi, T. (2018). Learn to cache: Machine learning for network edge caching in the big data era. *IEEE Wireless Communications*, 25(3), 28-35. <https://doi.org/10.1109/MWC.2018.1700317>
- Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In O., Maimon, & L. Rokach, (Eds.), *Data mining and knowledge discovery*

- handbook* (pp. 875-886). Springer. https://doi.org/10.1007/0-387-25465-X_40
- Chen, B., & Chen, L. (2014). A link prediction algorithm based on ant colony optimization. *Applied Intelligence*, 41(3), 694-708. <https://doi.org/10.1007/s10489-014-0558-5>
- Chen, H., Li, X., & Huang, Z. (2005). Link prediction approach to collaborative filtering. **Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '05)**, Denver, CO, 2005, pp. 141-142. <https://doi.org/10.1145/1065385.1065415> .
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval* (3rd. ed.). Facet Publishing.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122-135. <https://doi.org/10.1108/LR-06-2015-0061>
- Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, 2013, 48-55. <https://doi.org/10.1109/CTS.2013.6567203>
- Engle, S., Stromme, J., & Zhou, A. (2020). Staying at Home: Mobility Effects of COVID-19 (April 3, 2020). <http://dx.doi.org/10.2139/ssrn.3565703> .
- Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter*, 14(2), 1-5.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gerdtsri, N., Kongthon, A., & Vatananan, R. S. (2013). Mapping the knowledge evolution and professional network in the field of technology roadmapping: a bibliometric analysis. *Technology Analysis & Strategic Management*, 25(4), 403-422. <https://doi.org/10.1080/09537325.2013.774350>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Green, K. (2020). How GPs can contribute to the challenge of covid-19. *BMJ*, 369:m1829. <https://doi.org/10.1136/bmj.m1829>
- Heilig, L., & Voß, S. (2014). A scientometric analysis of cloud computing literature. *IEEE Transactions on Cloud Computing*, 2(03), 266-278. <https://doi.org/10.1109/TCC.2014.2321168>.
- Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, A. K. M., Alam, M., Shiraz, M., & Gani, A. (2014). Big data: survey, technologies, opportunities, and

- challenges. *Scientific World Journal*, 712826.
<https://doi.org/10.1155/2014/712826>
- Khaparde, V., & Pawar, S. (2013). Authorship pattern and degree of collaboration in Information Technology. *Journal of Computer Science & Information Technology*, 1(1), 46-54.
- Kibria, M. G., Nguyen, K., Villardi, G. P., Zhao, O., Ishizu, K., & Kojima, F. (2018). Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. *IEEE Access*, 6, 32328-32338.
<https://doi.org/10.1109/ACCESS.2018.2837692>
- Kolisetty, V. V., & Rajput, D. S. (2020). A Review on the Significance of Machine Learning for Data Analysis in Big Data. *Jordanian Journal of Computers and Information Technology (JJCIT)*, 6(01). <https://doi.org/10.5455/jjcit.71-1564729835>
- Liu, H., Gegov, A., & Cocea, M. (2017). Unified framework for control of machine learning tasks towards effective and efficient processing of big data. In W. Pedrycz, & S-M. Chen (Eds.), *Data science and big data: An environment of computational intelligence* (pp. 123-140). Springer, Cham.
- Lo, K., Wang, L. L., Neumann, M., Kinney, R., & Weld, D. S. (2020). S2orc: The semantic scholar open research corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (4969–4983).
<https://doi.org/10.18653/v1/2020.acl-main.447>
- Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for big data analytics in plants. *Trends in plant science*, 19(12), 798-808.
<https://doi.org/10.1016/j.tplants.2014.08.004>
- Makawana, P. R., & Jhaveri, R. H. (2018). A bibliometric analysis of recent research on machine learning for cyber security. In *Intelligent Communication and Computational Technologies* (pp. 213-226). Springer.
https://doi.org/10.1007/978-981-10-5523-2_20
- McCall, B. (2020). COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *The Lancet Digital Health*, 2(4), Article e166-e167. [https://doi.org/10.1016/S2589-7500\(20\)30054-6](https://doi.org/10.1016/S2589-7500(20)30054-6)
- Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2018). Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*, 270(1-2), 313-336. <https://doi.org/10.1007/s10479-016-2236-y>
- Mohammadi, M., & Al-Fuqaha, A. (2018). Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. *IEEE Communications Magazine*, 56(2), 94-101.
<https://doi.org/10.1109/MCOM.2018.1700298>

- Muhuri, P. K., Shukla, A. K., & Abraham, A. (2019). Industry 4.0: A bibliometric analysis and detailed overview. *Engineering applications of artificial intelligence*, 78, 218-235. <https://doi.org/10.1016/j.engappai.2018.11.007>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.*
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1. <https://doi.org/10.1186/s40537-014-0007-7>
- National Health Commission of the PRC (NHC). (2020). COVID-19 epidemic situation up to 24:00 on March 8th. <http://www.nhc.gov.cn/xcs/yqtb/202005/%20f2c83db9f73d4be5be0dc96af731813c.shtml>
- New.cn. (2020, May 22). COVID-19 is urged by the UN Secretary General to do everything possible to contain the outbreak. *New.cn*. http://www.xinhuanet.com/2020-02/29/c_1125642849.htm/.
- Nobre, G. C., & Tavares, E. (2017). Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study. *Scientometrics*, 111(1), 463-492. <https://doi.org/10.1007/s11192-017-2281-6>
- Pandey, R., Gautam, V., Bhagat, K., & Sethi, T. (2020). A machine learning application for raising WASH awareness in the times of Covid-19 pandemic. *arXiv preprint arXiv:2003.07074*.
- Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Qiu, H., Qiu, M., & Lu, Z. (2020). Selective encryption on ECG data in body sensor network based on supervised machine learning. *Information Fusion*, 55, 59-67. <https://doi.org/10.1016/j.inffus.2019.07.012>
- Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, Article 67. <https://doi.org/10.1186/s13634-016-0355-x>
- Read, J., Perez-Cruz, F., & Bifet, A. (2015, April). Deep learning in partially-labeled data streams. In R. L. Wainwright, J. M. Corchado, (Eds.), *SAC '15: Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 954-959). Association for Computing Machinery. <https://doi.org/10.1145/2695664.2695871>
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th edition). Perntice Hall.

- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Sharma, A., Singh, G., & Rehman, S. (2020). A Review of Big Data Challenges and Preserving Privacy in Big Data. In M. Kolhe, S. Tiwari, M. Trivedi, K. Mishra (Eds.), *Advances in Data and Information Sciences* (pp. 57-65). Springer. https://doi.org/10.1007/978-981-15-0694-9_7
- Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010, May). The hadoop distributed file system. In *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, Incline Village, (pp. 1-10). <https://doi.org/10.1109/MSST.2010.5496972>
- Singh, V. K., Banshal, S. K., Singhal, K., & Uddin, A. (2015). Scientometric mapping of research on 'Big Data'. *Scientometrics*, *105*(2), 727-741. <https://doi.org/10.1007/s11192-015-1729-9>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Sughasiny, M., & Rajeshwari, J. (2018, August). Application Of Machine Learning Techniques, Big Data Analytics In Health Care Sector–A Literature Survey. 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, Palladam, India, 2018, (741-749). <https://doi.org/10.1109/I-SMAC.2018.8653654>
- Sweileh, W. M., Al-Jabi, S. W., AbuTaha, A. S., Zyoud, S. H., Anayah, F., & Sawalha, A. F. (2017). Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016. *BMC medical informatics and decision making*, *17*(1), 72. <https://doi.org/10.1186/s12911-017-0476-7>
- Tseng, S. F., Won, Y. L., & Yang, J. M. (2016). A bibliometric analysis on Data Mining and Big Data. *International Journal of Electronic Business*, *13*(1), 38-69. <https://doi.org/10.1504/IJEB.2016.075333>
- Tuite, A. R., Ng, V., Rees, E., Fisman, D., Wilder-Smith, A., Khan, K., & Bogoch, I. I. (2020). Estimation of the COVID-19 burden in Egypt through exported case detection. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30233-4](https://doi.org/10.1016/S1473-3099(20)30233-4)
- Uddin, M. F., & Gupta, N. (2014, April). Seven V's of Big Data understanding Big Data to extract value. In *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education, Bridgeport, CT, 2014*, (pp. 1-5). <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523-538. <https://doi.org/10.1007/s11192-009-0146-3>

- Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Burdick, D., Eide, D., Funk, K., Katsis, Y., Kinney, R., Li, Y., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., ..., & Kohlmeier, S. (2020). COVID-19: The Covid-19 Open Research Dataset. *arXiv preprint arXiv:2004.10706*.
- Wang, L., Wang, G., & Alexander, C. A. (2015). Natural language processing systems and Big Data analytics. *International Journal of Computational Systems Engineering*, 2(2), 76-84. <https://doi.org/10.1504/IJCSYSE.2015.077052>
- Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*.
- World Health Organization (WHO). (2020, May 22). Coronavirus disease (COVID-2019) situation Re- ports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
- Xiang, L., Zhao, G., Li, Q., Hao, W., & Li, F. (2018). TUMK-ELM: a fast unsupervised heterogeneous data learning approach. *IEEE Access*, 6, 35305-35315. <https://doi.org/10.1109/ACCESS.2018.2847037>
- Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. *Ieee Network*, 28(4), 5-13. <https://doi.org/10.1109/MNET.2014.6863125>
- Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350-361. <https://doi.org/10.1109/ACCESS.2017.2696365>
- Zhu, C., Cao, L., Liu, Q., Yin, J., & Kumar, V. (2018). Heterogeneous metric learning of categorical data with hierarchical couplings. *IEEE Transactions on Knowledge and Data Engineering*, 30(7), 1254-1267. <https://doi.org/10.1109/TKDE.2018.2791525>

Biodata



Mustafa Ababneh, received his B.Sc Computer Information System at Jordan University of science and technology, M.Sc. in Computer Science at Amman Arab University in Amman, Jordan and now. He is Ph.D. student in Computer Information System at Near East University in Mersin 10, Turkey. My research interests focus on Big Data, Social Media, cloud computing and Information Retrieval. Email: 20194017@std.neu.edu.tr



Ayat Al-Jarrah, received her B.Sc Biomedical Engineering at Yarmouk University, M.Sc. in Computer Science at Amman Arab University in Amman, Jordan and now she is Ph.D. student in Computer Information System at Near East University in Mersin 10, Turkey. my research interests focus on Big Data, Social Media, cloud computing and Information Retrieval. Email: 20194007@std.neu.edu.tr



Assist. Professor Dr. Damla Karagozlu received her M.Sc. in Information Technology at Bournemouth University and Ph.D. in Computer Education and Instructional Technology at Near East University. She is currently an assistant professor at the Department of Computer Information Systems at Near East University in Mersin 10, Turkey. Her research interests focus on augmented reality, cloud computing and cybersecurity.

Email: damla.karagozlu@neu.edu.tr