International Journal of Multidisciplinary and Current Educational Research (IJMCER)
ISSN: 2581-7027 ||Volume|| 4 ||Issue|| 5 ||Pages 131-134 ||2022||

Leveraging Big Data Analytics to Power AI and ML (Machine Learning) Automation

Alex Mathew
Department of Cybersecurity, Bethany College

ABSTRACT : This document highlights how big data can be leveraged to power Artificial intelligence (AI) and Machine learning (ML). AI systems can be defined as machine-based systems that have a varying level of autonomy that, for a given set that is defined, objectives be used in the main predictions, decisions, or recommendations. Big data is responsible for feeding machine learning models that utilize such data in order to learn as well as improve the predictability as well as performance automatically through data and experience without any programming done by humans. Machine learning is a powerful as well as an essential tool that can be utilized in conducting many tasks as well as computing problems that are linked to big data. Learning target functions (f) that map input variables (X) to an output (Y) are the terms used to describe machine learning algorithms that incorporate data analytics. It is calculated as Y = f(X). In addition, supervised and unsupervised learning, as well as reinforcement learning techniques, are all part of the machine learning experience. Since data pairs are almost covering various kinds of domains, the collection of big raw data from the environment is quite complex and has great redundancies. ML is key in addressing the challenges that are presented by big data and showing some hidden patterns and information as well as bits of knowledge from great information. Understanding machine learning in the future will focus on how to make it easier for non-experts to specify and interact with various data types in various streams by making it more declarative.

.

KEYWORDS - Big data, Machine learning, Artificial intelligence, predictive analysis, descriptive analysis, prescriptive analysis

I. INTRODUCTION

Artificial intelligence (AI) systems can be defined as machine-based systems that have a varying level of autonomy that, for a given set that is defined, objectives are used in the main predictions, decisions, or recommendations. The use of AI methods is increasingly utilizing a massive quantity of alternative data sources as well as data analytics that are constantly utilizing huge amounts of alternative data sources and data analytics that are called ‘big data.’ Big data is responsible for feeding machine learning models that utilize such data in order to learn as well as improve the predictability as well as performance automatically through data and experience without any programming done by humans. Machine learning considers a category of artificial intelligence, and it is where computers get the ability to learn from data through the use of appropriate algorithms, making it possible for computers to identify the hidden patterns in data without having a user programming it actively in order to do so with the main goal of conducting a solid task. On the other hand, The term “big data analytics” refers to enormous amounts of data that can be generated, processed, and increasingly utilized by digital tools and information systems for predictive, descriptive, and prescriptive analyses [4] [11]. The ability to process unstructured data, increase in data storage capacity, and increased availability of structured data all contribute to the development of this capability. In society today, there is increased availability of novel data sources that can be processed using complex and powerful algorithms that include artificial intelligence systems. That leads to several benefits, but there are some challenges, too [1].

II. PROPOSED METHODOLOGY

Conventional data utilize centralized database architecture where problems that are large and complex are handled by one computer. Such an architecture is costly and has been proven to be ineffective when it comes to processing big data [2] [3]. Moreover, Big data is built on distributed database architecture, which divides large database blocks into smaller data sizes to solve problems. After that, many of the computers in a given computer network compute the problem’s solution. When attempting to find a solution to a problem, these computers also communicate with one another. In comparison to centralized database systems, the distributed database also provides lower prices, is cheaper, and has better performance [6]. This is due to the fact that in distributed database systems, microprocessors are more cost-effective than mainframes. In addition, the distributed database has more computational power than the conventional database, which makes use of conventional data. Machine learning is a powerful and essential tool that can be utilized in conducting many tasks as well as computing problems that are linked to big data. However, several searches today still face great challenges when it comes to big data processing [12]. Therefore, to realize the full potential of big data, the methodology of this paper will attempt to address many challenges as well as open issues that include:

1. AI and machine learning can be used to explore and make use of useful information in big data while also drawing more attention to the field because large amounts of useful data are being lost because novel data is often untagged and unstructured.
2. In order to deal with real-world issues, researchers typically use a single learning algorithm or technique in many of the currently available machine learning applications. However, it is essential to realize that all the approaches taken have some strengths and limitations. As such, there is a need to consider hybrid learning and the existing big data background.
3. The features of big data make the visualization of data become a very challenging task. The visualization techniques used recently, such as dimension reduction, could result in an abstract view of data. As such, the manner in which machine learning techniques could be utilized to provide trued geometric representations for big data should also be investigated.

III. BLOCK DIAGRAM

The user, the system, the big data, and the system all interact with machine learning, which is the primary focus of the big data AI and ML framework. These interactions can go either way. For instance, big data typically serve as learning components' inputs, resulting in outputs that eventually become big data. As such, a user might interact with the learning components through the provision of domain knowledge, usability feedback as well as personal preferences [5] [9]. Moreover, it is going to leverage the learning outcomes to improve the process of decision-making. The domain can be used as both the context in which learned models are applied and a source of information to direct the learning process. As depicted in Figure 1 below, today's systems architecture has an impact not only on simultaneous meetings but also on the optimal operation of learning algorithms as well as their efficiency.

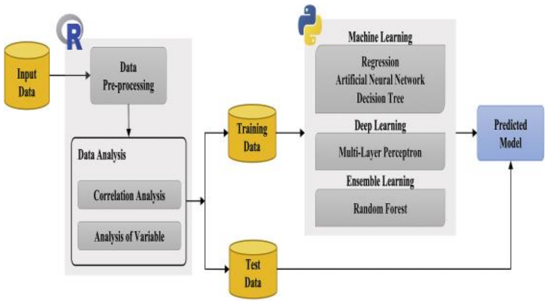

Fig. 1: Block diagram of machine learning environment

IV. ALGORITHM

Learning target functions (f) that map input variables (X) to an output (Y) are the terms used to describe machine learning algorithms that incorporate data analytics. It is calculated as Y = f(X). That is thought of as a general learning task in which new examples of input variables are used to make predictions for the future (Y). The functions' appearance or form is unclear. It wouldn't be necessary to learn it from m big data using ML algorithms if that were possible because it could be used directly. The most widely used form of machine learning for learning the mapping Y = f(X) and predicting Y for novel X. That is referred to as predictive analysis, and its key goal is to come up with the most probable prediction. If big data is utilized in the storage of bulk data as well as manipulation, extracting the proper information will only be enabled by machine learning. Using such machine learning, it will be possible to extract a pattern that is efficient. In addition, supervised and unsupervised learning, as well as reinforcement learning techniques, are all part of the machine learning experience. Unsupervised learning originates from unlabeled data sets, and that means they are directly linked to determining unknown properties in them [7] [10]. Most of the time, properties that are known from the training data are the focus of machine learning. Additionally, the discovery of previously unknown data properties is the primary focus of data mining. For example, the performance evaluation of a classifier is concerned with data selections, measurement of performance, estimation of errors as well as statistical tests.

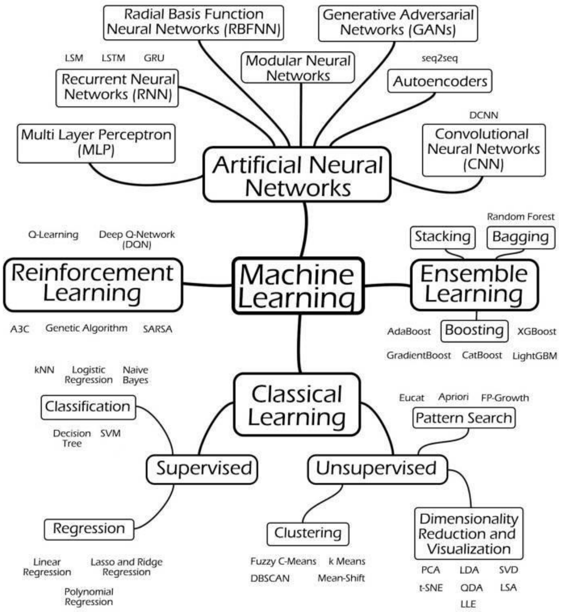

Fig. 2: Machine learning systems algorithm

FLOW CHART : FIGURE 3 BELOW SHOWS A PROCESS FOR APPLYING MACHINE DATA. THE PROCESS SHOWS A PATH THAT INVOLVES DESCRIPTIVE, PREDICTIVE AS WELL AS PRESCRIPTIVE ANALYSIS AND SIMULATION. WHAT IS EVEN MORE NOTABLE IS HOW THE MACHINE LEARNING PROCESS HAS BEEN NOTED EXPLICITLY AS RECURSIVE, AND THAT IS MOSTLY RIGHT IN MODELING LARGE DATA QUANTITIES [8]. IN ADDITION, IT IS IN CHARGE OF DETERMINING THE RELATIVE NUMBER OF RECORDS AT EACH STAGE OF A MACHINE LEARNING TASK.

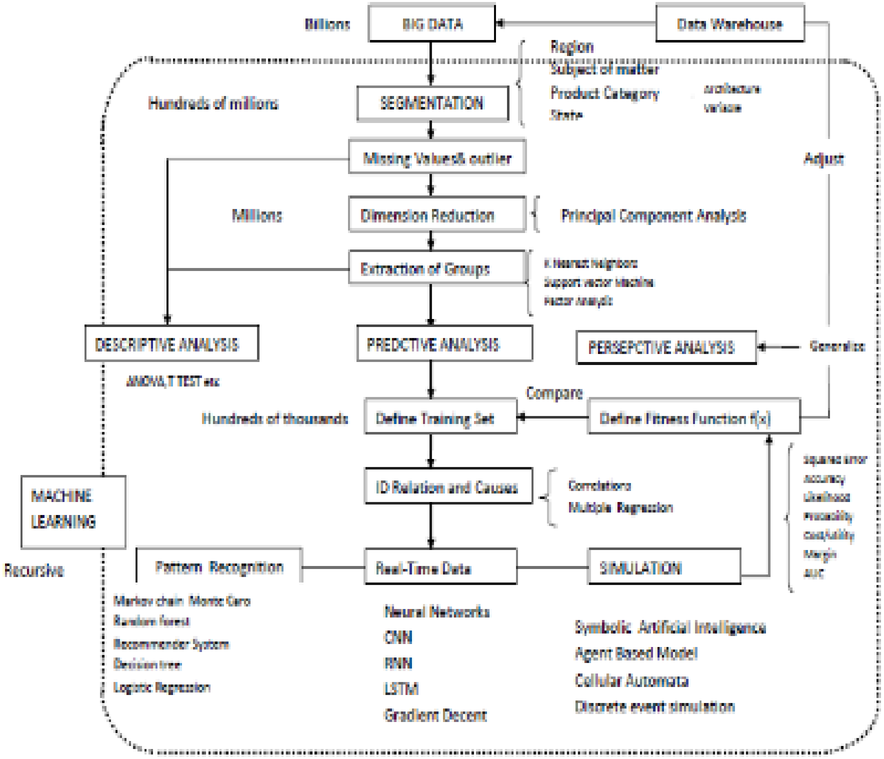

Fig. 3: Flow chart of Big data and machine learning

V. RESULT ANALYSIS

The procedure of processing big data is composed of four key processes that include pre-processing, analysis, model establishment, and model updating phases. Since data spouses are almost covering various kinds of domains, the collection of big raw data from the environment is quite complex and has great redundancies. As such, invalid and dirty data should be deleted in the first pre-processing phase. Additionally, people have to face massive, uncertain, and incomplete data in real life, and some key attributes to improve the predictability of processing are required.

CONCLUSION

ML is key in addressing the challenges that are presented by big data and showing some hidden patterns and information as well as bits of knowledge from great information. Understanding machine learning in the future will focus on how to make it easier for non-experts to specify and interact with various data types in various streams by making it more declarative. Extending machine learning techniques to assess the performance of various types of problems. AI and ML is the best choice to take care of the challenges brought about by big data and uncover hidden patterns, insights, and knowledge.

REFERENCES

[1] L. Zhou, S. Pan, J. Wang, and A.V. Vasilakos, Machine learning on big data: Opportunities and challenges. Neurocomputing, vol. 237, pp.350-361, 2017..

[2] R. Kumar, R. K. Banyal, P. Goswami, and V. Kumar. “Machine learning algorithms for big data analytics.” In Computational Methods and Data Engineering, pp. 359-367. Springer, Singapore, 2021.

[3] K. Divya, Sree, P. Bhargavi, and S. Jyothi. “Machine learning algorithms in big data analytics.” Int. J. Comput. Sci. Eng, vol. 6, no. 1, pp. 63-70, 2018.

[4] B. Rushiti. Aug 15, 2020. “Navigating Into the World of Machine Learning.” Medium. https://medium.com/swlh/navigating-into-the-world-of-machine-learning-1c1b10ae40b (accessed Oct 1, 2022)

[5] J. H. Jeong, J. H. Woo, and J. Park. “Machine learning methodology for management of shipbuilding master data.” International Journal of Naval Architecture and Ocean Engineering, vol. 12, PP. 428-439, 2020.

[6] A. Mahbod, and H. Leung. “A deep learning-based methodology for video anomaly detection in crowded scenes.” In Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II, vol. 11413, pp. 135-143. SPIE, 2020.

[7] S. L. Nita, L. Dumitru, and A. Beteringhe. “Machine learning techniques used in big data.” Scientific Bulletin of Naval Academy, vol. 19, no. 1, pp. 466-471, 2018.

[8] M. G. Kibria, K. Nguyen, G.P. Villardi, O. Zhao, K. Ishizu and F. Kojima. “Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks.” IEEE Access, vol. 6 pp. 32328-32338, 2018.

[9] B. Qolomany, A. Al-Fuqaha, A. Gupta, D. Benhaddou, S. Alwajidi, J. Qadir, and A. Fong. “Leveraging machine learning and big data for smart buildings: A comprehensive survey.” IEEE Access, Vol. 7 pp. 90316-90356, 2019.

[10] K. Rabah. “Convergence of AI, IoT, big data and blockchain: a review.” The lake institute Journal, vol. 1, no. 1, pp. 1-18, 2018.

[11] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, and D. Pothuhera. “Online incremental machine learning platform for big data-driven smart traffic management.” IEEE Transactions on Intelligent Transportation Systems 20, no. 12 (2019): 4679-4690.

[12] Sun, Alexander Y., and Bridget R. Scanlon. “How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions.” Environmental Research Letters, vol. 14, no. 7 pp. 073001, 2019.