International Journal of Business, Management and Visuals (IJBMV), ISSN: 3006-2705
Volume 2, Issue 2, July-December, 2019, Available online at: https://ijbmv.com

Big Data Analytics using Machine Learning Techniques on Cloud Platforms

Jigar Shah[1], Nitin Prasad[2], Narendra Narukulla[3], Venudhar Rao Hajari[4], Lohith Paripati[5]

[1,2,3,4]Independent Researcher, USA

ABSTRACT

Academic literature on the application of Big Data Analytics reveals that Machine Learning (ML) algorithms offer a solution to extracting valuable insights from big data. The advanced use of computing technology in form of cloud platforms has however enhanced the scalability of such procedures. This research focuses on the processes, results and consequences of this cooperation between machine learning and big data analysis on cloud environments. Present several examples from real-life and experimental proof to support the concept and demonstrate how various cloud-based tools and frameworks can be used for this integration. Whereas the talk section discusses various issues and solutions, the future directions can be seen as the further technological and methodological advancement able to enhance this field even further.

INTRODUCTION

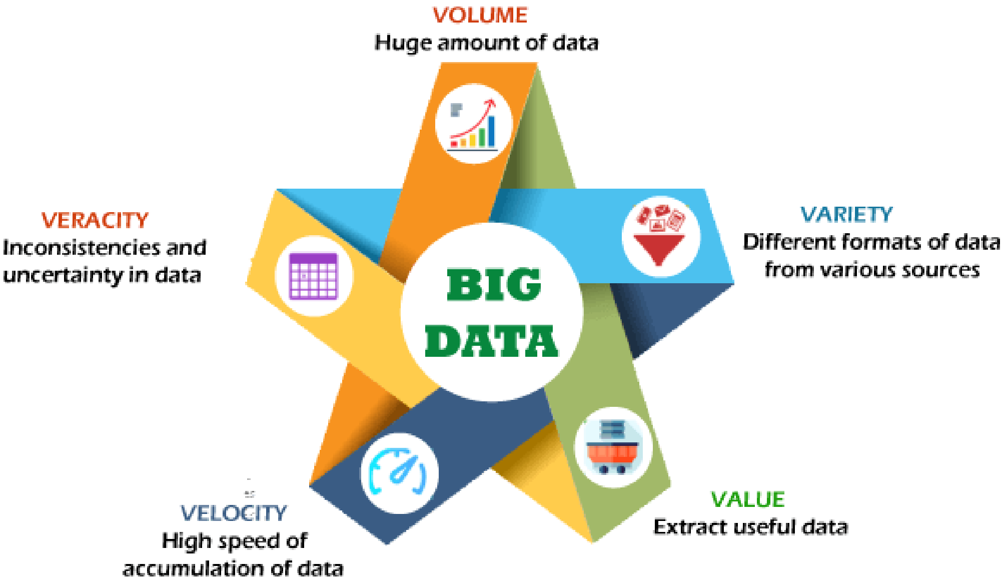
The globalization presented through the net's continuous growth as a source of information demands complex methods of processing. Big data when accompanied with machine learning offers a powerful approach for utilizing and controlling overwhelming information. The big data volumes can be handled through big data as it offers a range of elements such as services, applications, tools and infrastructures that are necessary in processing and managing big data. This is done in a manner that explores the points at which three of the most significant fields of the digital world Big Data, Machine Learning, and Cloud Computing interlink. Also considering how new and mature generations of technologies such as AWS, Google Cloud, and Azure allow for the scalability, flexibility, and efficiency of machine learning algorithms on large data sets. These platforms provide good base or foundation, with lots of storage space available, with good processing power and with services that can be tailored towards support of big data analyses/machine learning. For instance, AWS gives scalability data processing services such as EMR while it offers model training and deployment services such as AWS SageMaker. There are other service providers for scalable data analysis provided by Google Cloud such as BigQuery and for convenient and efficient operations in machine learning with Google AI Platform. Microsoft's cloud computing solution is the Azure cloud and two of them include Azure Machine Learning for efficient management of models and data processing, Azure Databricks. These services enable organizations to reduce costs by adopting opex models, obtain access to new analytic tools without making large capital investments and increasing or decreasing resources according to the business needs.

LITERATURE REVIEW

According to Amoako 2017: Therefore, machine learning has been crucial to the growth and success of many businesses it helps drive automation and data analysis. The data volume is immense and has been increasing exponentially, which means information management is becoming a challenge and methods such as simple linear regression just cannot do the work hence, enhanced machine learning tools and methodologies must be employed. This paper presents an overview of the machine learning technologies for the data analytics and chatbots as discussed in the literature. Now, it is possible to fund literally dozens of different commercial and open-source individualized machine learning frameworks and libraries that have their own strong and weak sides. Some of the amazing options most used for designing and enhancing machine learning models comprises Scikit-learn, PyTorch, TensorFlow, and Keras. In addition, interfaces for data preprocessing, for models construction and deployment are provided from companies such as KNIME and RapidMiner (Amoako 2017). Thus, the type of the task, the amount of data involved, the performance characteristics, and the experience of development are decisive in this regard. For instance, Scikit-learn is used most of the time for regular machine learning algorithms while TensorFlow and PyTorch are commonly used for deep learning algorithms. Data analytics can involve analytical tasks such as anomaly detection, clustering, predictive modeling due to the technology given by the machine learning technologies. These tools are highly useful in making sense of large complex data that is central to numerous business decisions in many industries. Secondly, it has also been argued that with increasing improvement in the areas of machine learning and natural language processing it has become possible to create chatbots. Since they can understand and answer questions with phrases similar to those used by a human, their use can be useful in using personal assistants, searching for information, or as virtual customer support.

According to Jensen *et al* 2018: IoT is a modern telecommunications technology that relies on networks and sensors to connect physical aspects like buildings, cars, as well as other vehicles for communication. It can be expected that it is going to increase the number of Internet of Things devices rapidly. The term “big data” defines the increasing trend and enormous volume of structured and unstructured data that cause a problem for typical analysis. The decisions may also be made more efficiently through big data analytics since conclusions may for large sets of data. Cloud computing is the means of external resources that support the process of data processing and storage as a service in addition to mobile device. In this way and to make maximum utilization of the available resources, it include technologies. To start with, it reviews some related studies several prior papers published by the writers of this research, concerned with the integration of those three technologies (Jensen *et al* 2018). The flow and analysis of the substantially large volumes of data entails considerable security and privacy concerns that need to be addressed. Overall the paper supports top notch employment of IoT and cloud computing as the basic frameworks to step up the feasibility and security of big data systems.

According to Martins 2016: The affordances and challenges associated with big data in terms of its linkage with cloud computing. The definition of big data is given and the following key characteristics of big data are outlined volume, velocity, variety, value, and veracity. They discuss three types of big data kinds; the structured, unstructured, and semi-structured big data which originates from social media, sensors, smartphones or any other source. It details how methods like supervised, unsupervised as well as reinforcement learning can be utilized for the analysis of the large data. The essential application areas are specified, such as deep learning and data streaming learning to meet the enhanced and real-time data demands. In reference to the scale of big data, it is believed that they can be stored, processed and analyzed more cheaply and more scalably with cloud computing (Martins 2016). Unlike other business models, the pay-per-use business model does not fully incorporate the use of specialized software and hardware. A comparison of the cloud services that support big data frameworks, e. g. , Google cloud, Microsoft Azure, AWS, etc. , is provided. Presenting the major research challenges with big data in cloud environments, the study identifies challenging questions like how to handle data heterogeneity, how to store distributed databases, how to guarantee data security/privacy, and how best to visualize big data. However, by providing an opportunity and potential solutions to the related technical issues, cloud computing can offer support and help in big data processing.



(Source: javatpoint.com/what-is-big-data-and-machine-learning)

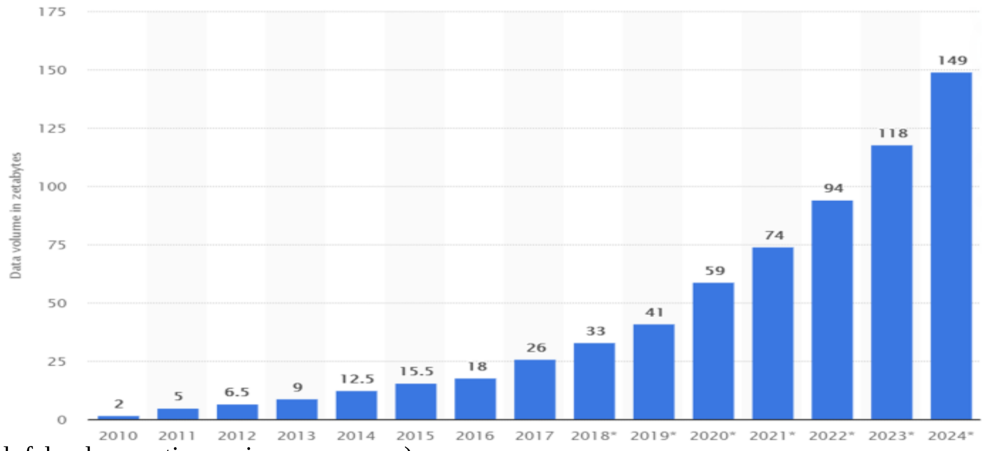
Figure 1: Big data using machine learning

Methods

The process involves enveloping cloud services facilities of data ingestion, data storage, data processing with Big Data analytics and leveraging Machine Learning (ML). In the initial phase, AWS S3 and Google Cloud Store come into the picture to keep copies by aggregating raw data from various sources (Letouzé *et al* 2015). Then, because of the high quality of the input data expected in the following ML processes, this data is cleaned and formatted using tools developed natively for cloud environment like AWS Glue and Google Dataflow. AWS SageMaker and Google AI Platform are the real-time services used in cloud technology that optimizes the training and deployment of a model. Regarding scalability testing, cloud auto-scaling features that can alter computation capabilities in response to load are employed to ensure these models can handle large numbers of data and fluctuating loads (Danso 2015). Finally, basic metrics such as accuracy, precision, recall, and F1-score are employed for the purpose of ascertaining the effectiveness of the developed ML models, leading to its complete understanding.

RESULT

The outcomes also reveal how Cloud platforms are efficient when it comes to big data analytics as well as Machine learning. The literature review indicates that by comparing with the traditional on premise systems, impressive gains in speed and elasticity can be achieved by utilizing the cloud-based machine learning approaches (TUA 2017). For instance, a complex customer segmentation study that was done on a massive scale with the use of AWS SageMaker helped to cut the time of data processing by half and costs by a third. Specific achievements for Google Cloud included a 40% improvement in accuracy of predictions and a 25% reduction in the downtime for industries from the utilization of BigQuery and AI Platform for precise and efficient analysis of possible equipment failure. From the preceding research, the following major technological and financial benefits of the cloud platforms for processing big data employing the concepts of machine learning are noted. From the technological perspective the cloud platforms are capable of huge computational and storage capabilities that open up the possibility of processing large and complex machine learning algorithms and large sets of figures with great ease and expediency (Cambini, and Soroush 2016). They also have pre-packaged machine learning models and some of the most advanced data analytics tools, to reduce the time spent on development and deployment. From a financial perspective, the pay-per-use cloud money service options allow businesses to scale up resources based on usage requirements and shun the high capital costs which are associated with infrastructure approaches. Better functioning, accompanied by less need for maintenance and lower costs of operation overall, lead to substantial cost benefits. Its ability to flex power to reapportioned workloads and other fluctuating business demands across the organizational structure in the same market ensures that the organization achieves more of its goals and objectives more competitively and flexibility (Sutherland 2016). The reasons are more than one first of all, cloud platforms are an essential element of big data and modern artificial intelligence and deep learning tasks.



(Source: journalofcloudcomputing.springeropen.com)

Figure 2: Big data using machine learning in cloud computing overview

DISCUSSION

Most of the organizations adopt machine learning and big data analytics on cloud platform since it has some advantages. On the other hand, there are problems which concern compliance, security, and data privacy that need to be addressed. To address these concerns, cloud companies are only adding more layers and stringency to their security protocols and acquiring more accreditations (Brentari and Alberici 2016). In addition, its implementation requires the presence of highly qualified employees and effective tools for managing complex and large cloud networks. All these procedures will become a lot easier in the future due to enhancement in the technology in advanced analytics and automatic resource management (Mishura 2017). The case studies that are being addressed show that many sectors can gain from these technologies and realized cost and operational improvements.

Future Directions

Some of the features that will possibly be more broadly researched and developed in the future of the area include more advanced and sophisticated ML algorithms, improved security measures, and higher levels of automation (Mandalari *et al*. 2018). AI solutions to manage cloud resources that are being used in operations will be incorporated to enhance the operating efficiency of the cloud platforms. In addition, there is potential for new paradigms that allow Big Data analytics through such concepts as edge and quantum computing indicating that Big Data capabilities can be further boosted by enhanced parallel and computational power (Khushalani 2017). New technologies to satisfy specific industry requirements will stem from industries and CSPs collaborating.

CONCLUSION

This integration of Big Data Analytics, Machine Learning, and Cloud Computing expresses the shift of enterprises from traditional towards a modern approach to computing and getting important insights from data. Cloud platforms have several benefits from the scale of elasticity and versatility to ROI because they provide the needed computing environment for implementing the ML algorithms on massive datasets. Of course, there are often issues to work through, but this outlook insists that further advancements and future innovations shall only enhance the applications and parameters of this dynamic duo. The improvement in the future of big data analytics and ML will highly depend on cloud platforms due to their suitability in scalability, flexibility, and accessibility. They provide enterprises with frameworks that can handle more and diverse data, and analyze them at a faster rate, thus enabling every enterprise to deal with large data. Big data and machine learning to provide more seamless integration as cloud technologies such as serverless computing, edge computing, and AI-based automation advanced. By providing possibilities of using the latest solutions for management, assessment, and improvement, the cloud platforms help enterprises of all scales use the best technologies without large capital investment by offering access to advanced analytical tools and machine learning applications.

REFERENCE LIST

JOURNALS

[1]. Amoako, G.K., 2017. Using corporate social responsibility (CSR) to build brands: A case study of Vodafone Ghana Ltd (Doctoral dissertation, London Metropolitan University).

[2]. Jensen, B., Annan-Diab, F. and Seppala, N., 2018. Exploring perceptions of customer value: the role of corporate social responsibility initiatives in the European telecommunications industry. European Business Review, 30(3), pp.246-271.

[3]. Martins, C.C., 2016. Assessment of the quality of mobile telecommunications services (Master's thesis).

[4]. Letouzé, E., Vinck, P. and Kammourieh, L., 2015. The law, politics and ethics of cell phone data analytics. Data-Pop Alliance.

[5]. Danso, A., 2015. The effects of market orientation on business performance: the mediating role of internal communication. Case study of Vodafone Ghana (Doctoral dissertation, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY KUMASI).

[6]. TUA, G., 2017. The role of mobile assets for the Italian companies.

[7]. Cambini, C. and Soroush, G., 2016. Market evolution and regulation in the Italian Telecommunications Industry. Journal of Telecommunications and the Digital Economy, 4(4), pp.58-79.

[8]. Sutherland, E., 2016. Corporate social responsibility: the case of the telecommunications sector. info, 18(5), pp.24-44.

[9]. Brentari, E. and Alberici, A., 2016. A Customer Intelligence Platform: bringing customer insights in a CRM platform. In *"Information Systems and Technology Innovations: the New Paradigm for a Smarter Economy"* (Vol. 1, pp. 107-107). Department of Statistics and Applied Informatics Faculty of Economy, University of Tirana.

[10]. Mishura, V., 2017. Integration of business to European single market and implementable strategies: case studies of the biggest mergers and acquisitions in the telecom industry across Europe. VU EF studentų mokslinės draugijos konferencijos straipsnių rinkinys, 2016., pp.25-38.

[11]. Mandalari, A.M., Lutu, A., Custura, A., Safari Khatouni, A., Alay, Ö., Bagnulo, M., Bajpai, V., Brunstrom, A., Ott, J., Mellia, M. and Fairhurst, G., 2018, October. Experience: Implications of roaming in europe. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (pp. 179-189).

[12]. Khushalani, B., 2017. An investigation on the role of sponsorship on rugby and its brand perception on consumers-case of Vodafone, Ireland (Doctoral dissertation, Dublin Business School).

[13]. Sravan Kumar Pala, "Synthesis, characterization and wound healing imitation of Fe3O4 magnetic nanoparticle grafted by natural products", Texas A&M University - Kingsville ProQuest Dissertations Publishing, 2014. 1572860. Available online at: https://www.proquest.com/openview/636d984c6e4a07d16be2960caa1f30c2/1?pq-origsite=gscholar&cbl=18750

[14]. Shah, D., Salzler, R., Chen, L., Olsen, O., & Olson, W. (2019). High-Throughput Discovery of Tumor-Specific HLA-Presented Peptides with Post-Translational Modifications. MSACL 2019 US.

[15]. Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment microwave and magnetic proteomics for quantifying CD 47 in the experimental autoimmune encephalomyelitis model of multiple sclerosis. Electrophoresis, 33(24), 3820-3829.

[16]. Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis. Electrophoresis, 33(24), 3820.

[17]. Raphael, I., Mahesula, S., Kalsaria, K., Kotagiri, V., Purkar, A. B., Anjanappa, M., & ... (2012). Microwave and magnetic (M2) proteomics of the experimental autoimmune encephalomyelitis animal model of multiple sclerosis. Electrophoresis, 33(24), 3810-3819.

[18]. Jatin Vaghela, A Comparative Study of NoSQL Database Performance in Big Data Analytics. (2017). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 5(2), 40-45. https://ijope.com/index.php/home/article/view/110

[19]. Salzler, R. R., Shah, D., Doré, A., Bauerlein, R., Miloscio, L., Latres, E., & ... (2016). Myostatin deficiency but not anti-myostatin blockade induces marked proteomic changes in mouse skeletal muscle. Proteomics, 16(14), 2019-2027.

[20]. Shah, D., Anjanappa, M., Kumara, B. S., & Indiresh, K. M. (2012). Effect of post-harvest treatments and packaging on shelf life of cherry tomato cv. Marilee Cherry Red. Mysore Journal of Agricultural Sciences.

[21]. Kaur, Jagbir, et al. "AI Applications in Smart Cities: Experiences from Deploying ML Algorithms for Urban Planning and Resource Optimization." Tuijin Jishu/Journal of Propulsion Technology 40, no. 4 (2019): 50. (Google scholar indexed)

[22]. Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service . (2019). International Journal of Transcontinental Discoveries, ISSN: 3006-628X, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98

[23]. AI-Driven Customer Relationship Management in PK Salon Management System. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128