ISSN 2278-3091
Volume 8, No.5, September - October 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse63852019.pdf
https://doi.org/10.30534/ijatcse/2019/63852019

A Hadoop based Maching Learning Technique for Semantic Indexing of Learning Objects in Big Data Environment

Kamal EL GUEMMAT[1], Sara OUAHABI[2]
[1] Laboratory of Information Processing, Faculty of Sciences Ben M'Sik, Hassan II University, Casablanca, Morocco, k.elguemmat@gmail.com
[2] Modeling and Information Processing Laboratory, Faculty of Sciences Ben M'Sik, Hassan II University, Casablanca, Morocco, sara.ouahabi@gmail.com

ABSTRACT
Today with the big data generated every lapse of time, BIG DATA comes to offer several opportunities to store and manage this huge amount of data. The processing and analysis of these data becomes essential in order to draw the most relevant information and knowledge that can help us to make our decisions. This set of treatment is possible via different machine learning techniques which is interested in automatically learning the treatments so take into account the old experiences. The link between BIGDATA & Machine learning exists today via digital analytics which is interested in taking advantage of these techniques to locate a specific resource or to make managerial decisions concerning the management of universities or to propose new courses answering to the needs of different interlocutors. The problem that persists, there is a lack of solution according to our research that is interested in doing pretreatment in an educational BIGDATA environment in order to extract the semantic metadata through machine learning algorithms. This metadata will be very useful later to generate the knowledge that will improve the teaching process. Our solution is used to answer this problem through various techniques of HADOOP for BIG DATA and MACHINE LEARNING. As a result, our application is largely valid according to the techniques that integrates that are confirmed in this field.

Key-words: Machine Learning, Pretreatment, Educational Bigdata Environment, Metadata, Hadoop.

1. INTRODUCTION

With the advent of the internet and new technologies, there is a huge data source forming a big data [1]. Thus the management and exploitation of this large body of data becomes paramount [2].

Machine Learning techniques can help us in this way, but the problem is that there are many limitations and constraints on traditional machine learning techniques in a BIGDATA environment.

However, several machine learning techniques appeared that allow us to take advantage of BIGDATA with potentialities more adapted to this context, unlike classical techniques [3], [4], [5].

Another question that arises is that we must propose techniques specific to the educational field [6].

There are some works in the literature, [2], [7], [8], [9]. But according to our knowledge there is a lack of application that can preprocess Learning Object (LO) by machine learning techniques [4] to get out good quality metadata [10] to take advantage of these to generate knowledge and make decisions.

The purpose of your article is to propose a new application that remedies these concerns. It is based on the Framework [7] and consists of several modules from HADOOP [11], [12], [13]. For the supervised learning phase, it uses confirmed indexing techniques [14], [15], [16], [17], [18].

This article will be distributed as follows, the next section discusses the state of the art on BIGDATA and Machine Learning and their relation to the world of learning analysis. Then there is our proposal and our application followed by a section that raises the different opportunities and challenges and in the end there is the conclusion.

2. STAT OF THE ART

2.1. Big data & Machine Learning

Today we live in an era in there is a growing increase in data generation without stop. Several large-scale data are collected and studied in many fields, from engineering sciences to social networks, commerce, bio molecular research and security [19]. The term BIG DATA profoundly means this trend of data generation.

With emerging technologies and all associated peripherals, it is expected that huge amounts of data will be created over the next few years. In fact, 90% of the current data was created in the last two years. Continue in the foreseeable future [1].

ABI Research estimates that in 2020 there will be more than 30 billion connected devices [20]. These mega-data have tremendous potential in terms of business value in various areas such as healthcare, biology, transportation, online advertising, energy management and financial services [21].

With the advent of cloud computing mobile computing. Data can be collected from more people and sources at a record cost for extended time intervals without additional effort [22]. Big Data describes fundamentally too big and

fast data, exceeding the processing capacity of conventional database systems [23].

[24] pointed out the definition of Big data has little to do with the data itself, as the analysis of large quantities of data is not new, but rather Big Data refers to emergent suit of technologies that can process mass volumes of data of various types at faster speeds than ever before.

It also covers innovative technologies and techniques for capturing, storing, managing and analyzing a large ensemble with various structures.

Big data are characterized by five dimensions: volume, speed, variety, veracity and value [25] requiring new data processing to enable decision discovery and process optimization.

Big data challenge The effective understanding and use of this new wealth of raw information poses a significant challenge to today's green engineers / researchers (Geo, climate and environment, Bio, medicine, and health, Stars, galaxies, and the universe) [1].

There are three essential stages to unlock the value of BIGDATA in any organization. They include the course of discovery, analysis and visualization and application [2].

However, with traditional machine learning methods are unfortunately not designed to handle a large amount of data [5].

Given the large influx of data, it is absolutely necessary to improve the way conventional computer / analytic data models are designed and developed.

Machine learning makes it possible to build systems that improve automatically thanks to the experiments [26].

The Learning Machine is a research area that focuses on theory, performance, properties of learning systems and algorithms.

The reason is that Learning machine has the ability to learn from data and provide information, decision ideas and data-driven forecasts [27].

In general, the Learning machine is divided into three sub domains: supervised learning, unsupervised learning, reinforced learning [28].

Machine learning allows users to guess hidden structures and make predictions from large masses of data. So the ML is an important element for the analysis of large data [29]. Big data analysis is a mechanism to discover the knowledge behind the data. Large data analytics must use specialized, high-performance analysis techniques and tools for data mining, data optimization, predictive analytics, predictive analytics, and more. [30]. The McKinsey Global Institute has stated that machine learning would be one of the key drivers of the Big Data revolution [31].

In the literature, several studies deal with the relationship between BIGDATA & Machine Learning [3], [4], [5]. [1].

The work of [3] it's a Literature that studies the latest advances in machine learning research for large data processing. He studies the close links of machine learning with signal processing techniques for large data processing. Given the big data growth today, existing machine learning techniques have great difficulty adapting. They are designed for generic and nonspecific cases for large amounts of data.

According to the latter, there are several critical machine learning issues for BIGDATA; Learning for large scale of data; Learning for different types of data; Learning for high speed of streaming data; Learning for uncertain and incomplete data; Learning for data with low density and diversity

However, several techniques appear to deal with these difficulties. Representation Learning, Deep learning, Distributed and parallel learning, Transfer learning, Active learning, Kernel-based learning.

[4] introduces a Machine Learning Framework on Big Data (MLBiD) to guide discussion of its opportunities and challenges. It focuses on machine learning and introduces the preprocessing learning and evaluation phases. In addition, the Framework also includes four other components, namely large data, users, domains and systems.

Big data also present major challenges for Machine Learning, such as high data dimensionality, scalability of models, distributed computing, streaming of data [32], adaptability and usability

Machine Learning can be characterized in several dimensions: nature of learning feedback supervised learning, unsupervised learning, and reinforcement learning, target of learning tasks (representation, task), and schedule of data availability (batch learning, online learning). Preprocessing, evaluation, Big Data and other components.

In order to efficiently handle large data, existing machine learning paradigms and algorithms need to be adapted.

The marriage of ML and Big Data suggests a prosperous future in a new frontier. It is essential to discover BIGDATA's models and hidden knowledge and information and actually exploit them to make decisions and scientifically explore them.

[5] summarizes and organizes the relationship between Machine learning and BIG DATA and deals the challenges according to the different axes of big data volume, speed, variety, veracity and value. It identifies the challenges of machine learning and associates each challenge with a specific dimension of Big Data. He discusses machine learning challenges caused by volume (Processing Performance, Curse of Modularity, Class Imbalance, Curse of Dimensionality, Feature Engineering, Non-Linearity, Bonferonni's Principle, Variance and Bias). He discusses machine learning challenges caused by variety (Data Locality, Data Heterogeneity, Dirty and Noisy Data), It discusses machine learning challenges caused by velocity (Data Availability, Real-Time Processing/Streaming, Concept Drift, Independent and Identically Distributed Random Variables). He discusses machine learning challenges caused by truthfulness (Data Provenance, Data Uncertainty, Dirty and Noisy Data).

It introduces the techniques and methodologies currently being developed and used to address these challenges associated with machine learning with Big Data. Handling techniques used in conjunction with existing algorithms are first introduced (data manipulation, processing manipulation, algorithm manipulations). Secondly, different paradigms of machine learning that are particularly well adapted to the challenges of Big Data are discussed.

[1] deals with the exponential use of nodes in a BIGDATA Machine learning environment and the reduction of energy costs which is one of the top priorities of many energies related companies. It examines the theoretical and experimental literature on data modeling, in large-scale, data-intensive domains, regarding: (1) the effectiveness of the model, including computer-based learning requirements, as well as the structure and design of data-intensive areas, and introduces 2) new algorithmic approaches with the least memory and processing requirements to minimize computational costs, while maintaining / improving its predictive/classification accuracy and stability.

This article provides a comprehensive review of the most recent literature on sustainable / energy efficient machine learning, including theoretical, empirical, and experimental studies of various needs and recommendations.

2.2. Big data & Machine Learning & Learning Analytics

A lot of data is generated from massive open online courses (MOOC), online learning management systems, web-based educational resources, the use of mobile in educational technologies, the use of social networks by the students ...

This mass of data cannot be managed by traditional learning management systems from which comes the need to use big data technologies to capture, manage and process data.

according to the gradual changes taking place in the higher education environment, learning technologies continue to cross all facets of higher education, generating a multitude of useful "data traces". These data can be used to help higher education institutions better adapt to changes occurring inside and outside their environment.

Unlike traditional methods of data mining, most of which aim to automate the process of detecting interpretable models and executing predictions, educational data focuses on the development of a new tool for discovering data models and apply BIGDATA tools and analysis techniques [33], [6].

Big Data integrates emerging fields of inquiry into learning analysis [34], which is already an expanding field in the education sector.

According to Ferguson [35], the formal definition of learning analytics can be “the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.”

However, research in the analysis of learning is limited on the examination of indicators of the particular individual and the performance of the class.

[34] indicated that Big Data represents the most dramatic framework in the effective use of the vast array of data and ultimately shaping the future of higher education.

[36] also discussed the application of BIGDATA in higher education, noting that technological developments allowed some analysts to move towards higher analysis in higher education.

Big data can address some of the key challenges in higher education practice [7]: 1. improving learners’ experience [34], 2. improving learners’ knowledge trough enhanced academic studying, 3. more effective evidence-based decision making, 4. strategic response to changing global trends [2], 5. opportunity for converting complex, often unstructured data into actionable information.

In the literature, several studies deal with the relationship between BIGDATA & Machine Learning & LEARNING ANALYTICS. [2], [7], [8], [9].

[2] addresses the challenges posed to higher education institutions and explores the potential of BIGDATA to address these challenges.

Today there are several trends affecting these institutes (economic, social, technological, educational).

The article dealt with the framework of [37] who proposed a conceptual framework to describe Big Data in higher education through four components (Institutional analytics, Information technology (IT) analytics, Academic/programme analytics, Learning analytics).

According to [2], when used effectively, Big data can help institutions improve their knowledge and improve student performance, reduce dropout rates and increase the number of graduates as well offers several benefits for either students, administration or faculty. The key contribution of Big Data depends on the application of three models (descriptive, relational and predictive) and the utility of each to guide better decision-making.

[39] this work evokes big data for education and the different benefits that can offer this technology and the analytics technologies for different areas.

[7] Provides a core framework architecture to support education research versus big data.

The general process of extracting practical values from large data can be broken down into five steps [38]: 1. Acquisition and recording; 2. Extraction, cleaning, and annotation; 3. Integration, aggregation, and representation; 4. Modeling and analysis; 5. Interpretation.

These five phases can be divided into two main sub processes: data management and data/learning analytics.

The roles and abilities of learners, researchers, teachers, and data scientists are clearly defined and proposed within this framework. The framework, proposed is composed of five modules: Data capture and collection, ETL, Hadoop platform, analysis engines execute, presentation layer).

[8] this work uses collaborative filtering based on recommendation techniques, in order to recommend elective courses to students based on their grades obtained in other subjects. With the use of machine learning libraries from mahout over Hadoop to generate a set of recommendation.

[9] offers a Big Data Approach for Classification and Prediction of Student Results Using Map Reduce. Predictive models will help the instructor to understand how students in the classroom will succeed or not, and the instructor will be able to choose appropriate instructional interventions to improve the student's learning outcomes.

It also helps instructors predict the success and failure of exam students and allows them to give advice to prevent exam failures.

The implementation is done in the Hadoop framework with Map Reduce and Revolutionary R Enterprise RRE.

3. THE APPLICATION

The preprocessing step transforms data from the BIGDATA into a form that can be used as an input for learning through data cleansing, extraction, transformation, and merging.

Some learning methods, including representational learning, can also be used for preprocessing data [4].

The preprocessing phase allows us to extract metadata. These last ones that have a good quality can help us make educational decisions [10].

Our work cf. Figure 1 fits into this perspective in order to propose an application that can preprocess an educational BIGDATA according to machine learning algorithms.

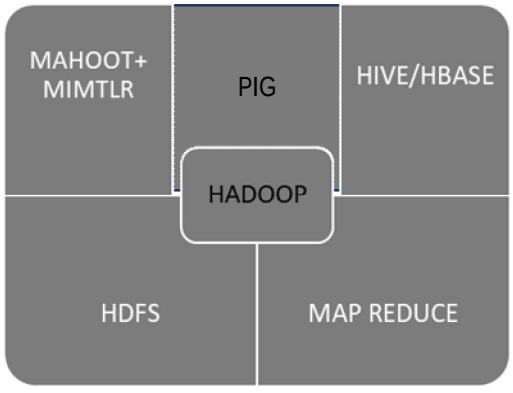

Figure1: Our Application

This application is based on the Framework proposed in the work of [7] and on the Hadoop Framework.

Hadoop, which was published as a large enough ecosystem that allows us to manage large data, pretreated them, then, in our case, the learning phase chooses algorithms dedicated to the automatic processing of the natural language and adjusts the parameters of the model to generate the desired outputs [11], [12], [13]. Hadoop is responsible for the powerful Hadoop Distributed File System (HDFS), encouraged by Google's file system, as well as the parallel programming model using the Map-Reduce paradigm. Thus the execution program is divided into two parts: MAP and REDUCE.

- MAP: a node is responsible for mapping the data to another set of data (key / value pairs). These keys that are spread over several nodes will be grouped by the reduce.
- REDUCE: this second reduction step takes the output of an input map and combines these data tuples into a smaller set of tuple. The reduce task is always performed after the map job.

Hive is a data warehouse framework constructed on top of Hadoop that supports analysis of large datasets stored in Hadoop's HDFS.

HBase is a distributed, scalable, big data store that runs on top of HDFS. HBase enables real-time and random access to large data.

Pig is a platform for evaluating large sets of data consisting of a high level language (Pig Latin) and an implementation framework whose compiler produces a series of Map-Reduce FOR completing within from Hadoop.

Mahout is a library of scalable machine learning algorithms built on top of Hadoop using the Map-Reduce paradigm.

In the supervised learning part and in order to ensure our pretreatment we will enter as an example of learning the indexing results of the approaches, Multilayer Indexing Model for Teaching and Learning Resources (MIMTLR), that are confirmed [14], [15], [16], [17], [18].

4. OPPORTUNITIES AND CHALLENGES

With the advent of Big Data, many of the assumptions on which the algorithms are based have now been broken, preventing the execution of analytic tasks. In response to these pitfalls, along with the need to deal quickly and comprehensively, new approaches and learning paradigms have been developed. However, it is still difficult to find the best tools and techniques to meet specific challenges. There are several challenges such as the source of data from different sources with different formats, the adaptation of different traditional algorithms to big data, updating the existing models when new data arrives, without having to recycle the data. complete model, stream & online learning processing, a combination of different approaches would better cover the problems related to machine learning with Big Data [5].

Our application has this opportunity, they are based on various approaches from the ecosystem Hadoop and the

Framework of [7] to do the preprocessing of the learning objects.

According to [4] There are several challenges either in the process of: data preprocessing aims to solve a number of problems such as redundancy, inconsistency, noise, heterogeneity, transformation, labeling (for (semi) supervised Machine Learning), data imbalance and representation / selection of features; Learning, in taxonomy, they first ranked the studies according to the parallelism is taken into account in their algorithms / platforms. The methods of the non-parallelism category aim to have much faster optimization methods, able to process Big Data without any parallelism; The evaluation of Big Data Machine Learning is not a simple combination of both types of metrics. It must take into account both the tradeoffs in each type of metrics and the complex tradeoffs between them; Big data, user domain, system. Thus, there are several challenges confronting our application.

[39] suggested three cooperative elements to be examined during the collection of data for analytics: timing, population, and location.

Effective understanding and use of this new wealth of raw information poses a significant challenge to today's green engineers / researchers [1].

[40] gives some recommendations to green engineers / researchers on some key mechanisms of sustainable data modeling. Ensemble models, Model complexity problem, Local learning strategy, Semiparametric approximation, Deep learning, Big data computing.

One of the major barriers for implementing big data analytics in higher education could be financial expenses [41].

Most institutional systems are not interoperable, summary of administrative data, class data and online can pose additional problems [37].

Security and privacy issues pose additional challenge to implementing Big Data [42], Thus in higher education [2].

Our application is concerned about the preprocessing part of learning resources in order to prepare them for smarter processing. In order to arrive at a complete system of educational decision in a BIGDATA environment and according to the challenges mentioned above, it is really a big challenge.

5. CONCLUSION

The application we have proposed is largely valid, allows preprocessing, indexing and organizing learning objects according to machine learning algorithms in a BIGDATA environment according to Hadoop's proven techniques.

It can handle a huge number of learning objects and extract the most important metadata using dedicated techniques for information extraction and processing of natural language. This metadata will enable us to benefit from this huge number of learning resources in order to easily locate them, to make decisions about teaching and to improve it and to make decisions on the management of the school and develop it.

Our application removes the pretreatment problem in a BIGDATA context, because we have found a lack in this sense, we organize the resources according to precise information that will help us in the process of knowledge generation. We remain as a perspective work on the decision that can be done using these data pre-processed and refined according to the Hadoop ecosystem.

The mixing of these domains, BIGDATA and Machine Learning, appears very beneficial either in the context of teaching or in other fields in order to benefit from a large mass of data to extract the useful knowledge to take the good decisions.

REFERENCES

1. 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
2. Daniel, B. (2015). **Big Data and analytics in higher education: Opportunities and challenges**. British journal of educational technology, 46(5), 904-920.
3. 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, 2016(1), 67. https://doi.org/10.1186/s13634-016-0355-x
4. 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/10.1016/j.neucom.2017.01.026
5. L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). **Machine learning with big data: Challenges and approaches**. IEEE Access, 5, 7776-7797. https://doi.org/10.1109/ACCESS.2017.2696365
6. Luan, J. (2002). **Data mining and its applications in higher education**. In A. Serban & J. Luan (Eds), Knowledge management: building a competitive advantage in higher education (pp. 17–36). PA: Josey-Bass.
7. Klašnja□Milićević, A., Ivanović, M., & Budimac, Z. (2017). **Data science in education: Big data and learning analytics**. Computer Applications in Engineering Education, 25(6), 1066-1078.
8. Dwivedi, S., & Roshni, V. K. (2017, August). **Recommender system for big data in education**. In 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH) (pp. 1-4). IEEE. https://doi.org/10.1109/ELELTECH.2017.8074993
9. Mohan, M. M., Augustin, S. K., & Roshni, V. K. (2015, December). **A BigData approach for classification and prediction of student result using MapReduce**. In 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) (pp. 145-150). IEEE. https://doi.org/10.1109/RAICS.2015.7488404
10. Moges, H. T., Van Vlasselaer, V., Lemahieu, W., & Baesens, B. (2016). **Determining the use of data**

quality metadata (DQM) for decision making purposes and its impact on decision outcomes—An exploratory study. Decision Support Systems, 83, 32-46.
https://doi.org/10.1016/j.dss.2015.12.006

11. Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). **A survey of open source tools for machine learning with big data in the Hadoop ecosystem**. Journal of Big Data, 2(1), 24.
12. Nesi, P., Pantaleo, G., & Sanesi, G. (2015**). A hadoop based platform for natural language processing of web pages and documents**. Journal of Visual Languages & Computing, 31, 130-138. https://doi.org/10.1016/j.jvlc.2015.10.017
13. Gudivada, V. N., Rao, D., & Raghavan, V. V. (2015). **Big data driven natural language processing research and applications**. In Handbook of Statistics (Vol. 33, pp. 203-238). Elsevier.
14. Kamal EL GUEMMAT, Sara OUAHABI. **A semantic distances-based approach for a deeply indexing of learning objects**. International Journal of Emerging Technologies in Learning (iJET), Vol 14, No 6, 2019, ISSN: 1863-0383.
15. Kamal EL GUEMMAT, Sara OUAHABI. **Understandings of Exploratory Pedagogy for Educationnal search engine**. The Turkish Online Journal of Distance Education, 20 (2), 2019. ISSN 1302-6488. https://doi.org/10.17718/tojde.557857
16. Kamal EL GUEMMAT, Sara OUAHABI. **A Literature Review of Indexing and Searching Techniques Implementation in Educational Search Engines**. International Journal of Information and Communication Technology Education ISSN: 1550-1876-EISSN: 15501337. Issue #54 (Vol 14 Nr 2 2018). https://doi.org/10.4018/IJICTE.2018040106
17. Kamal EL GUEMMAT, Sara OUAHABI. **Issues of E-Learning Search Engine and its Challenges**. The International Journal of E-Learning and Educational Technologies in the Digital Media (IJEETDM), 4(3): 75-85 2018. ISSN 2410-0439 (Online).
18. Kamal EL GUEMMAT, El Habib BENLAHMAR, Mohamed TALEA, ElKhadir LAMRANI. **Implementation and Evaluation of an Indexing Model of Teaching and Learning Resources**. Procedia - Social and Behavioral Sciences-Volume 191, 2 June 2015. ISSN: 1877-0428 https://doi.org/10.1016/j.sbspro.2015.04.654
19. A Sandryhaila, JMF Moura, **Big data analysis with signal processing on graphs: representation and processing of massive data sets with irregular structure**. IEEE Signal Proc Mag 31(5), 80–90 (2014)
20. ABI. (2013). **Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020**, ABI Research. [Online]. Available: https://www.abiresearch.com/press/more-than-30-billion-devices-willwirelessly-conne/].
21. W. Raghupathi and V. Raghupathi, **Big data analytics in healthcare: Promiseandpotential**, HealthInf.Sci.Syst.,vol.2,no.1,pp.1–10,2014. https://doi.org/10.1186/2047-2501-2-3
22. S.Madden,**Fromdatabasestobigdata**,IEEEInternetComput,16 (2012), 4–6. https://doi.org/10.1080/13604813.2012.687881
23. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. et al (2011). **Big data: the next frontier for innovation, competition, and productivity**. Retrieved October 7, October 30, 2014, from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_nex _frontier _for_innovation]
24. Yang, L. (2013). **Big Data Analytics: What Is the Big Deal?** Retrieved December 30, 2013, from http:// knowledge.ckgsb.edu.cn/2013/12/30/technology/big-data-analyticswhats-big-deal/
25. M.A.BeyerandD.Laney**,''Theimportanceof'bigdata' :Adefinition,''** Gartner Research, Stamford, CT, USA, Tech. Rep. G00235055, 2012.
26. M. I. Jordan and T. M. Mitchell, **"Machine learning: Trends, perspectives, and prospects,"** Science, vol. 349, pp. 255-260, 2015. https://doi.org/10.1126/science.aaa8415
27. M. Rouse. (2011). **Machine Learning Definition.** [Online]. Available: http://whatis.techtarget.com/definition/machine-learning.
28. B Adam, IFC Smith, F Asce, **Reinforcement learning for structural control**.J Comput Civil Eng 22(2), 133–139 (2008)
29. C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, **"Big data analytics: a survey,"** Journal of Big Data, vol. 2, pp. 1-32, 2015// 2015
30. G. Siemens, **Learning analytics: envisioning a research discipline and a domain of practice**. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, (2012), 4–8. https://doi.org/10.1145/2330601.2330605
31. M. James, C. Michael, B. Brad, and B. Jacques, **Big Data: The Next Frontier for Innovation, Competition, and Productivity**. New York, NY: McKinsey Global Institute, 2011.
32. M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, **"Deep learning applications and challenges in big data analytics,"** Journal of Big Data, vol. 2, pp. 1-21, 2015
33. Romero, C. R. & Ventura, S. (2010). **Educational data mining: a review of the state of the art**. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 40, 6, 601–618 https://doi.org/10.1109/TSMCC.2010.2053532
34. Long, P. & Siemen, G. (2011). **Penetrating the fog: analytics in learning and education.** EDUCAUSE Review, 46, 5, 30–40.
35. R. Ferguson, Learning analytics: **Drivers, developments and challenges**, Int. J. Technol. Enhan. Learn. 4 (2012), 304–317 https://doi.org/10.1504/IJTEL.2012.051816
36. Wagner, E. & Ice, P. (2012). **Data changes everything: delivering on the promise of learning analytics in higher education**. EDUCAUSE Review, July/August, 33–42.
37. Daniel, B. K. & Butson, R. (2013). **Technology enhanced analytics (TEA) in higher education**, Proceedings of the International Conference on Educational Technologies, 29 Novemebr–1 December, 2013, Kuala Lumpur, Malaysia, pp. 89–96.

38. A. Labrinidis and H. V. Jagadish, **Challenges and opportunities with big data**, Proc. VLDB Endowment, 5 (2012), 2032–2033. https://doi.org/10.14778/2367502.2367572
39. B.Manoj, K.V.K.Sasikanth, M.V.Subbarao, V Jyothi Prakash (Volume 7, No.6, Novembre-Decembre 2018). **Analysis of Data Science with the use of Big Data.** International Journal of Advanced Trends in Computer Science and Engineering. DOI: 10.30534/ijatcse/2018/02762018. https://doi.org/10.30534/ijatcse/2018/02762018
40. B. Becker, **Learning analytics: insights into the natural learning behavior of our students**, Behav. Soc. Sci. Librarian, 32 (2013), 63–67.
41. J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, **Academic analytics: A new tool for a new era**, EDUCAUSE Review, 42 (2007), 40.
42. P Amarendra Reddy, O Ramesh (Volume 8, No.2, March-April 2019). **Security Mechanisms leveraged to overcome the effects of Big Data Characteristics**. International Journal of Advanced Trends in Computer Science and Engineering. DOI: 10.30534/ijatcse/2019/34822019.