



Big Data Analytics for Preschool

Mohamed Bellaj, Abdelaziz El Hibaoui and Ahmed Bendahmane

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

September 23, 2020

Big Data Analytics for Preschool

M.BELLAJ

Computer Science and Systems Engineering
Laboratory
Faculty of Science, Abdelmalek Essaadi
UniversityTetuan, Morocco.
bellajmed@gmail.com

A.EL HIBAOU

Computer Science and Systems Engineering
Laboratory
Faculty of Science, Abdelmalek Essaadi
UniversityTetuan, Morocco.
a.hibaoui@gmail.com

A.BENDAHMANE

Computer Science and Systems Engineering
Laboratory
Faculty of Science, Abdelmalek Essaadi
UniversityTetuan, Morocco.
ab.dahmane@gmail.com

Abstract

This paper provides a depth review of Big Data Analytic (BDT) benefits, process, and challenges in the preschool education sector. BDT plays an important role in optimizing education intelligence by simplifying institutions, organization, educator's approaches, predictive teaching and assessment approach. Furthermore, BDTs are used to examine, identify and predict learners' needs, threat failures and outcomes to progress their knowledge outcomes and to make sure that the academic programmers undertaken are of excellent standards. This paper also identified the phases of employment Big Data and how to analyse it. Although BDT significant contributions for education, there are several challenges regarding its security, privacy, ethics, and lack of skilled professionals, data processing and storage. Consequently we will describe the sources of some challenges of implementing big data analytics in the preschool sector and give some propositions to overcome these obstacles.

I. Introduction

In educational domain, the whole of pedagogical decisions made by a professor to measure a pupils understanding of the material or organize the structure of a course may probably have the highest influence on pupil learning and the degree of graduation. High-grade lessons may decrease the period a learner takes to acquire certain material, permit pupil to acquire further information in the equivalent volume of time and support learner to make efficient decisions about whatever they ought to actually study and drill.

This learning efficiency not simply develops the skill of the pupil, but it is also beneficial to lessening certain demands of the teachers. The tool, most convenient for supporting the supple decision making that educators need to raise the quality of the education experience is Big Data Analytics. It is an efficient to provide educators and apprentices an edge in understanding where and how enhancements can be made in the learning step.

A number of decisions making, frequently touted in the domain of product-innovation has reached to modernize the service industry of education. Big Data Analytics generates new occasions to advance the schooling process by helping teachers and learners make cleverer decisions earlier in the knowledge evolution. Developments, in expending data science to drive practice innovation, are developing fast. Nearly daily, the creation of new tools and smart apps to help learners and educators create an efficient use of their time.

Technology has and will permanently be a vital part of our practice. Still, it is how we, as educators, practice the power of new technology to support our decisions that is most important. What is commonly called "soft skills", it offer the opportunity for a teachers to acquire social and emotional intelligence.is what actually drives the learning procedure.

In the aim to establish the right interrogations, to mark the preeminent use of big data as a device to advance our decision making, it is primordial for educators and pupils to recognize how data analytics can upgrade the learning process. Three conducts we can benefit from using Big Data based on its aptitude to measure understanding, personalize the learning practice, and design additional interesting courses.

II. Big Data & education

The appropriate meaning of Big Data refers to the huge quantity of information that flows over multiple stations and generally online in each brief instant. Its data that is too enormous, composite, and dynamic for any conventional implements to release and manage. The form originated in the open public community, where experts were trying to discover quicker and more accessible solutions to accumulate and process vast quantities of information. Recognitions to progressions in technology, this data can currently be explored and interpreted, providing unlimited benefits to the government, education, engineering, healthcare, and further data-driven activities.

To be familiar with the idea about making Big Data "big" defining by both in dimension and significance, especially in education is that it allows for the exploration and prediction of learners comportment across an enormous variety of demographics, acquisition degree ,individual backgrounds, thoughtful progressions, academic intentions, environmental features, personal potentials, skills –everything can be able of measuring.

In education domain, the majority of these data factors are now being explored to serve design instructional approaches, examine and evaluate the influence of these approaches on both learners and educators, generally all of that can create a revolution in educational domain by creating an efficient learning environment.

III. Value Created from Big Data

- ✓ Creating transparency by making big data openly available for business and functional analysis (quality, lower costs, reduce time to market, etc.).
- ✓ Supporting experimental analysis in individual locations that can test decisions or approaches, such as specific programs.
- ✓ Assisting, based on customer information, in defining market segmentation at more narrow levels.
- ✓ Supporting Real-time analysis and decisions based on sophisticated analytics applied to data sets from customers and embedded sensors.
- ✓ Facilitating computer-assisted innovation in products based on embedded product sensors indicating customer responses.

IV. Big Data dimensions

One view, espoused by Gartner's Doug Laney describes Big Data as having three dimensions: volume, variety, and velocity. Thus, International Data Corporation (IDC) defined it: "Big data technologies describe a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis." Two other characteristics seem relevant: value and complexity. We summarize these characteristics in the next paragraph.

1) Three Vs

in general Most descriptions of big data focus on the size of data in storage and Size matters, though there are other main attributes of big data, specifically data variety and data velocity. The three Vs of big data (volume, variety, and velocity) constitute a comprehensive explanation, and they break the myth that big data is only about data volume. In addition, each of the three Vs has its individual implications for analytics (See Figure 1).

2) The Big Data – Where is it?

Big data surrounds us, although we may not immediately appreciate it. Part of the problem is that, except in unusual situations, the majority of us don't agree with huge amounts of information in our daily lives. Lacking this immediate experience, we frequently fail to understand both opportunities as well challenges presented by big data. Because of these features, there are currently an amount of disputes and challenges in addressing these characteristics going forward.

- ✓ **Data Volume:** that means it measures the amount of data available to an institution, which does not necessarily have to own all of it as long as it can access it. As data volume rises, the value of diverse data records will decrease in proportion to age, richness, type, and quantity among other factors.
- ✓ **Data Velocity:** Data velocity measures the rapidity of data creation, streaming, and aggregation. Organization has rapidly augmented the speed and richness of data used for diverse private transactions (for example, web-site clicks). Data velocity management is much more than a bandwidth issue; it is also an ingest issue (extract transform-load).
- ✓ **Data Variety:** Data variety is a measure of the richness of the information representation – text, images video, audio, etc. From an analytic

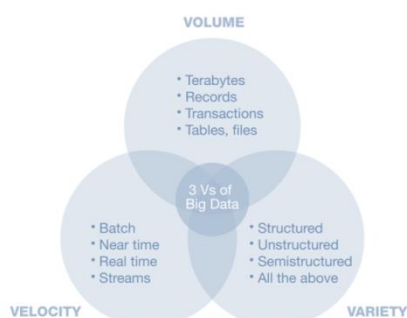


Figure 1: three vs of Big Data

perspective, it is possibly the principal obstacle to effectively using large volumes of data.

Non- structured data, Incompatible data arrangements, and inconsistent data semantics create significant challenges that can lead to analytic extension.

- ✓ **Data Value:** this stage measures the utility of data in making decisions. It has been noted that "the purpose of computing is insight, not numbers". Data science is examining and useful in getting to recognize the data, but "analytic science" involves the predictive control of big data.
- ✓ **Complexity:** Complexity measures the degree of interconnectedness (probably very huge) and interdependence in big data structures such that a small modification (or arrangement of small changes) in one or a few features can yield very large changes or a small transformation that cascade through the system and substantially mark its behavior, or no change at all.

V. Benefits of Big data implementation in preschool

It is obviously that big data with the enormous analytics studies tough efficiently predictions can be lead to deep understanding of educational process and related mechanisms. Generally the big data analytics process used for:

Collaboration: many specialists from several branches and disciplines can change experience and create collaborate centers. These initiatives ameliorate the educational approaches used and encourage collaboration and cooperation and coordinate between researches view.

Understanding the educational process: by identifying the majority of web resources used by the pupils and focus on the specific test, it became less difficult to choice the recommended courses and tests, the relative difficulty degree of exams, the distribution of learning activities per time but the most important is to identify a regular indicator and parameters to correct the previous errors.

Understanding the learning process: for each learner some characteristics, styles and a particular speed of learning. By giving an idea about educational learner's path we can extract different recommendation to develop the learning process.

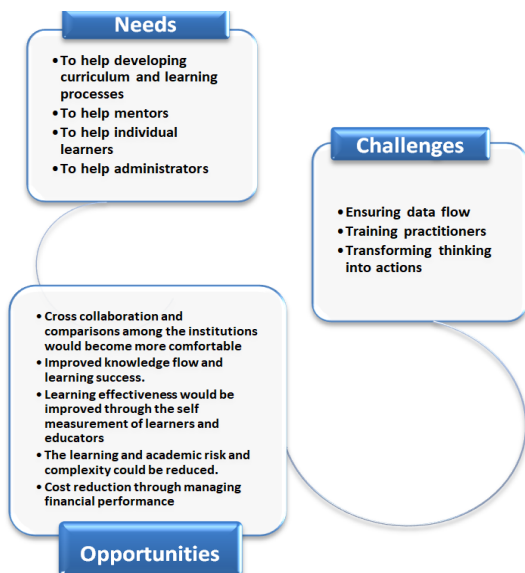
Feedback: among the popular obstacle for educational systems is to identify the sources of problem and fixed it, by the traditional method we can define the reason of learners fails instantly and the learner could fail again.so to overcome the situation, the new approaches suggest to extract the previously information from each learner path to propose alternative methods correction.

Predicting: the specialists can develop a suitable curriculum performance for the next generation based on their personal profiles, by using digital traces saved during the previous interaction with the platform and other available data.

Motivation: by getting the results of implementation of big data, learners potentially observe the benefit and became more and more convinced they see the impact of how this works.

VI. NEEDS, OPPORTUNITIES AND CHALLENGES

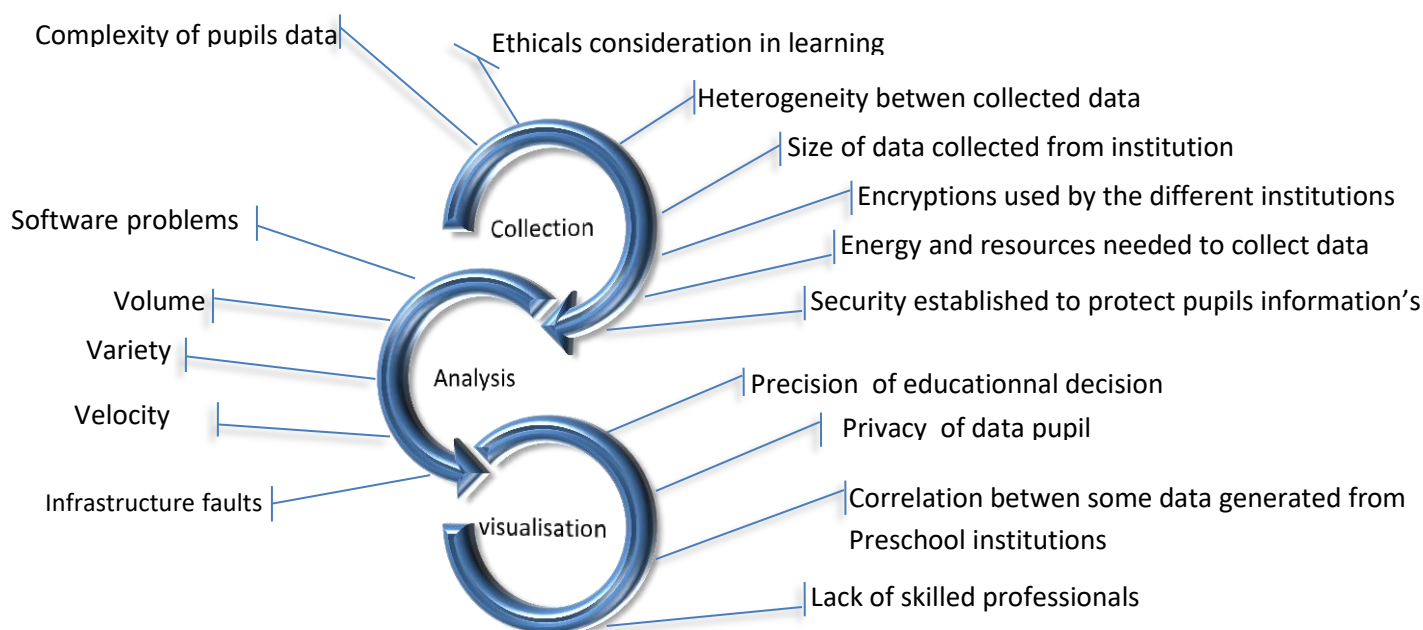
Big data analytics is now becoming a great challenge for the educationalist. People are now concerned about the intelligent outcome of institution to know students learning and academic



progresses.

But it is becoming hard for the management of the institution to just drive for the big data analytics without knowing its aspects. Figure above represents the educational (learning and academic) analytics from three aspects: Needs, Opportunities and Challenges.

VII. Big data process and challenges in education



1) Collection & challenges

Data collection is the first phase in revealing the value accrued from Big Data. This needs identifying data that can expose beneficial and valuable information. Data must be filtered for significance and stored in a form that is useful, as little is gained in investing in huge volumes of data and storage infrastructure if the huge majority of the data in it is not practical.

There are also challenges related with quality of data collected and described. Lack of uniform measures and indicators make inter (national) comparison and evaluation difficult, as the quality of data generated from Big Data is entirely reliant on the quality of data collected and the robustness of the procedures or indicators used.

1. **Ethics:** Ethical defies for Big Data Analytics in Preschool comprise identifying the institute approaches for preserving personal learner data privacy, individual consent, data ownership, and transparency. These challenges arise as educational data collection, and use is not subject to any official ethical examination process. Moreover the difficulties rise when the data sources are complex and sophisticated.
2. **Heterogeneity:** the greatest essential problem in the data collection phase of the data sources is the Heterogeneity. Heterogeneous data complications ascend due to variety, Representation and Semantics of the data sources. Furthestmost of the data produced currently essentially differ from the data types that the original systems were designed for Common errors in representation are caused by date formats and character fields. Database inventors might try to link datasets using name and surname of the pupil to extract some essential information. Since character fields are item sensitive, even small misrepresentation, such as, using changed capital letters will make the joins and search inefficient.
3. **Size of data:** Another great problem is to transfer the data collected. Due to the dimension of the data the rapidity of the transfer may be a bottleneck in the process.
4. **Energy and resources:** Storing and loading such a huge sized data requires massive volume of energy and resources. One of the obstacles of the Big Data is to

find the best situated servers to stock the data. The server locations must also be energy proficient and scalable. The location is essential due to the speed of transfer of the stored data to do the analyses.

5. **Encryption:** A lot of the safe transmissions needs some type of encryption decided on beforehand by scholar institutions and learners parents, in other hand the institutions don't establish a commons laws to explain certain measures and procedures.
6. **Data security:** Perhaps the main threat to personal security is the unregulated accumulation of data by numerous companies (game application, social media ...). This information represents a severe security fear, particularly when many individuals so willingly surrender such information. Questions of accuracy, dissemination, expiration, and access abound who has responsibility for the fidelity and accuracy of the data?

Propositions

- i. Obviously, several big data must be protected with respect to privacy and security laws and procedures. IDC suggested five levels of increasing security: privacy, compliance-driven, custodial, confidential, and lockdown. Further study is obligatory to evidently define these security levels and chart them against both present law and current analytics.
- ii. For these causes, ethical guidelines are compulsory to ensure stewardship and ownership of data are visibly defined and issues of privacy are taken into consideration so that data are protected from abuse. In this sense, educators and organizations can better inform instructional design and pedagogy while enabling pupil to become further conscious of their own learning behaviors.
- iii. users and Developers of datasets must therefore be aware that data should be revised and updated on a periodic basis to be used efficiently.
- iv. Since storing the big data can be expensive, certain companies have tried to store only part of the data collected. by result that will produce a challenge in the interpretation of data extraction results, in other hand that can arise the portion of the false correlations and unidentified data links.
- v. The searchers must Describe the practice of this information, if it is for the benefit of the learners, such as predicting pupils' comportment and a series of recommendations and assistances based on Learning Analytics, or if it is for research aims to achieve Learning Analytics intentions. And describe the time period of keeping learners' data and a description of a deletion process.

2) Analysis & challenges

Once data have been rendered into a practical form, it has to be analyzed to produce actionable data. Though, with the increasing diversity in the nature of information, handling and analyzing varied data set is becoming a very hard process. Analysis needs to include linking, connecting and correlating diverse data sets to be able to grasp the information that is

supposed to be transported by these data. This situation is, therefore, named as the 'complexity' of Big Data.

So how do we make certain that all data of a given type is reliable and accurate? Or, perhaps just approximately, wrong outcomes are not the single problem in big data analytics. The central problem is the velocity of the analysis. This is one of the three V's of the big data. Most define these V's as; Volume, Variety and Velocity.

- a. **Variety and Volume:** Volume and Variety has previously been debated during the data collection and integration process.
- b. **Velocity:** Velocity of big data not simply refers to the flow of data from sources to database but also the flow of the information from database to the finale result in analytics. The quickness of the data extraction and data analysis is the greatest chief advantage that an institution can get over competitors.
- c. **Infrastructure faults:** Storing and analyzing large volumes of data that is crucial for a company to work requires a vast and complex hardware infrastructure. If more and complex data is stored, more hardware systems will be needed.
A hardware system can only be reliable over a certain period of time. Intensive use and, rarely, production faults will most certainly result in a system malfunction.
- d. **Software problems:** Loosing data is not always a hardware problem. Software can as well malfunction and cause irreparable and more dangerous data loss. If one hard drive fails, there is usually another one to back it up, so there is no harm done to data, but when software fails due to programming "bug" or a flaw in the design, data is lost forever .Software solutions are limited by hardware capabilities.

Propositions

- i. Correct and fast planned decisions will growth the return on investment and increase the educational-solution share of the institutions. However this requires to examine the dynamic big data enormously fast.
- ii. For avoiding such catastrophic events they use a backup system that does the simple operation of storing all data. By doing this, companies obtain continuity, even if they are drawn back temporary.
- iii. To overcome this problem, programmers developed series of tools that will reduce the impact of a software failure. A simple example is Microsoft Word, which saves from time to time the work that a user is doing in order to prevent the loss of it in case of hardware or software failure. This is the basic idea of preventing complete data loss.

3) Visualization & challenges

This is the last stage where the analyzed data is made available to users in a form that is interpretable and integrated into existing processes, and ultimately used to guide decision making.

We can resume the data visualization process by several steps noted below:

•**Prediction:** Through prediction, learner activities and future performance can be revealed. Thus, an appropriate intervention would accomplish Learning Analytics goals.

•**Intervention:** determine which students might be at risk, advise students who may need additional assistance and improve students success.

•**Recommendation:** The main goal in the context of Learning Analytics is the aptitude to make recommendations to pupil based on their activities

•**Personalization:** improve educational performance and accelerate educational innovation. For instance, it can be carried out through personalizing e-learning based on learner's ability, or support students by personalizing learning suggestions.

•**Reflection and Iteration:** The objective behind reflection is to evaluate the past work to improve future experience and to turn it into learning according to personalization and adaptation. This iteration can optimize all Learning Analytics stakeholders in the design of its cycle.

•**Benchmarking:** Benchmarking is a learning process, which identifies the best practices that produce superior results. These practices are replicated to enhance performance outcomes (Vorhies & Morgan, 2005).
Challenge:

- a. It is necessary to know level of exactness that the handler requires to identify the enough data to create an estimate or prediction of the particular probability and accuracy of a given result.
- b. At what periodic time data still operational and what does the validity of publicly existing results expire?
- c. To protect learner Privacy Big Data in education needs transparency that make public the identity of the pupil to report decisions. Moreover, using Big Data Analytics implements to predict learners' future educational outcomes, practical performance and commitment possibly will destruct their privacy.
- d. The eliminated information from the prototype due to the high correlation could have been inappropriate predictive ones that revealed the incorrect correlation and false linkage to the designated variable.
- e. The challenge produced by using big data creates at every step of the analytical process for data scientists and management. The enterprises are having bigger defies in employing skilled data scientist that can interact professionally with the big data rather than the difficulties in the analytical process itself

Propositions

- i. Design suitable systems plans to handle the data effectively.
- ii. Establishment of a universal standard with internationally agreed-upon fundamental principles would also be a vast benefit. These changes should include critical consideration existing, conditions and nuances, shaping clear procedures and allowing appropriate conversion periods to apply the required modifications.
- iii. The governments and especially high-tech companies must provide more financial resources in the aim to encourage investment in this domain and facilitate laws by establishing less complicates procedures.
- iv. Sensitize the parents and educator by helping institutions to manipulate their information's and

results for the reason that they have a great part to facilitate the first step for the scientists.

- v. The institutions around the world must decrease the amount of strategies and approaches used in educational domain and seek to produce together international rules which can standardize the Complicate data process and growth the efficacious results.

Common challenges: Build a specialized center over the world to overcome a lack of big data researchers in education; furthermore we can promote the competence of the staff during each process of treatment.

VIII. Conclusion:

although this domain play a primordial part in learning development by focusing efforts on child learning process we can obviously develop and implement an efficient strategies to overcome learning obstacles from preschool to high school, We found a lack of studies that shed the light on big data analytics in preschool(BDTP) issues process including their challenges.

This paper initiates a collaborative exploration effort to begin examining (BDTP) issues and challenges. We indicated several of the major issues in big data storage, management, and application. That we believe must be addressed within the next decade.

Our coming research will focus on developing skillfully a further understanding of the issues related with (BDTP). We will begin to explore more efficient solutions to some of the challenges that we have raised in this paper through our research.

Process of Big data Analytics in preschool	Process issues	Challenge	propositions	comments	
	collection	<i>Ethics</i>	enabling pupil to become further conscious of their own learning behaviors	(+)give learners a self-confident and protect their rights (-)require collaboration	
		<i>Heterogeneity</i>	Establish a periodic revise of DATA stored	(+) simple to update (-)application require a periodic modification	
		<i>Size of data</i>	Storing a part of the data collected	(+)resolve a part of whole problems (-)losing some sensible information	
		<i>Energy and resources:</i>	Looking for an approximate location to minimize distance between stations.	(+) accelerate the process of data acquisition. (-)very costly.	
		<i>Encryption:</i>	Establish an efficient common encryption strategy.	(+) Has huge potential, can protect data versus attackers (-)Has limitations outcomes.	
		<i>Data security</i>	Increasing security levels and standardizes it.	(+)efficient mechanism to protect data (-)more sustained efforts towards implementation are required	
	analytics	<i>Variety and Volume</i>	Follow and examine data process progression	(+)very important to start any analytics process (-)lack of indicators precision	
		<i>Velocity</i>	Corporation between organization to plan the fast decisions	(+)reach a certified results (-)require a cost resources	
		<i>Infrastructure faults</i>	Develop a series of tools to reduce the impact of hardware failure	(+) provide a useful mechanism to gain additional time (-)insufficient budget	
		<i>Software problem</i>	use a backup system	(+) reduce the rate of failure (-)limited update	
	visualization	<i>level of precision</i>	Design appropriate systems and minimize strategies around the Word.	(+)reach an operational results (-)less level of collaboration	
		<i>Data expiration</i>	Allow sufficient transition periods to apply the convenient changes	(+)keep the suitable and useful data (-)Not feasible by the majority of institution.	
		<i>Privacy</i>	Sensitize the parents and learners and demand their permission.	(+)protect pupil's confidentialities (-)need collaboration	
		<i>investment</i>	provide more financial resources	(+)the essential pillar for each changes (-)rare attractive results	
	Common challenge: we can consider lack of skilled big data researchers in education. The most shared factor between all the processes of treatment in big data analytics.				

Table 1: Summarizing the Grid of big data analytics challenges and Solutions

References

- [1] B. Daniel, "Big Data and analytics in higher education: Opportunities and challenges: The Value of Big Data in Higher Education", Br. J. Educ. Technol., vol. 46, no 5, p. 904-920, sept. 2015.
- [2] S. Kaisler, F. Armour, J. A. Espinosa, et W. Money, "Big Data: Issues and Challenges Moving Forward", in 2013 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 2013, p. 9951004.
- [3] U. Sivarajah, M. M. Kamal, Z. Irani, et V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods", J. Bus. Res. vol. 70, p. 263-286, janv. 2017.
- [4] A. Klačnja-Milićević, M. Ivanović, et Z. Budimac, "Data science in education: Big data and learning analytics", Comput. Appl. Eng. Educ., vol. 25, no 6, p. 1066-1078, nov. 2017.
- [5] M. Khalil et M. Ebner, "Learning Analytics: Principles and Constraints", 2015.
- [6] B. Logica et R. Magdalena, "Using Big Data in the Academic Environment", Procedia Econ. Finance, vol. 33, p. 277-286, 2015.
- [7] Wouter Vollenbroek, Knut Jägersberg, Sjoerd de Vries, Efthymios Constantinides, "Learning Education: An „Educational Big Data“ approach for monitoring, steering and assessment of the process of continuous improvement of education", Paper presented in European Conference in the Applications of Enabling Technologies, 20-21 November 2014.
- [8] Katrina Sin and Loganathan Muthu, "Data in Education Data Mining and Learnings Analytics", ICTACT JOURNAL ON SOFT COMPUTING: Social Issue on Soft Computing Models for Big Data, July 2015, Volume: 05, Issue: 041035 Application of Big Data.
- [9] Shikha Anirban, "Big Data Analytics in the Education Sector: Needs, Opportunities and Challenges", International Journal of Research in Computer and Communication Technology, Vol 3, Issue 11, November – 2014
- [10] Mervat A. Bamiah, Sarfraz N. Brohi, Babak Bashari Rad, "Big Data Technology in Education: Advantages, Implementations, and Challenges", journal of Engineering Science and Technology Special Issue on ICCSIT 2018, July (2018) 229 - 241© School of Engineering, Taylor's University.
- [11] Julius Murumba, Elyjoy Micheni, "The International Journal of Engineering and Science (IJES)" Volume || 6 || Issue || 6 || Pages || PP 14-21 || 2017 || ISSN (e): 2319 – 1813, ISSN (p): 2319 – 1805, DOI: 10.9790/1813-0606021421.
- [12] Fourth Quarter 2011 TDWI Best Practises Report Big Data Analytics, By Philip Russom.
- [13] B. Tulasi, "Significance of Big Data and Analytics in Higher Education", Journal of Computer Applications (0975 – 8887) Volume 68– No.14, April 2013 21.
- [14] B. Tulasi, "Significance of Big Data and Analytics in Higher Education", Journal of Computer Applications (0975 – 8887) Volume 68– No.14, April 2013 21.
- [15] by Rebecca Webb, "12 Challenges of Data Analytics and How to Fix Them", published on Wed, Dec 5, 2018, <https://www.clearrisk.com/risk-management-blog/challenges-of-dataanalytics>
- [16] Alex Bekker, "Head of Data Analytics Department, ScienceSoft The „Scary“ Seven: big data challenges and ways to solve them" <https://www.scnsoft.com/blog/big-data-challenges-and-their-solutions>
- [17] by TeachThought Staff, "13 Challenges For Big Data In Education", January 11, 2016 in The Future Of Learning <https://www.teachthought.com/the-future-of-learning/13-characteristics-data-rich-learning-environment/>
- [18] María Jesús Rodríguez-Triana, Alejandra Martínez-Monés, Sara Villagrà-Sobrino, "Learning analytics in small-scale teacher-led innovations: Ethical and data privacy issues", Journal of Learning Analytics, 3(1), 43–65. <http://dx.doi.org/10.18608/jla.2016.31.4>.
- [19] Junpei Naito, Yukino Baba, Hisashi Kashima, Takenori Takaki, Takuya Funo, "Predictive Modeling of Learning Continuation in Preschool Education Using Temporal Patterns of Development Tests", The Eighth AAI Symposium on Educational Advances in Artificial Intelligence 2018 (EAAI-18).
- [20] Carpenter KLH, Sprechmann P, Calderbank R, Sapiro G, Egger HL (2016) "Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach". PLoS ONE 11(11): e0165524. doi:10.1371/journal.
- [21] Antonio Cerrato(&), Fabrizio Ferrara, Michela Ponticorvo, Luigia Simona Sica, Andrea Di Ferdinando, and Orazio Miglino, "DILIGO Assessment Tool: A Smart and Gamified Approach for Preschool Children Assessment", 2017, DOI 10.1007/978-3-319-59451-4_23.