EasyChair Preprint

№ 2143

Comparison and Relationship between Big data Analytics and Machine learning

Mohammed Abdelmoneim and Saad Subair

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

December 12, 2019

Comparison and Relationship between Big data Analytics and Machine learning

Mohammed abdel moneim Elkhier saaed[1*]
Department of Information Technology
International University of Africa,
Khartoum, Sudan
+249900359307
Email:mohammedabdelmoneim410@gmail.com

Professor.saad subair[2]
Department of Computer Science,
International University of Africa,
Khartoum, Sudan
+249912451880
Email:ssubair@gmail.com

ABSTRACT

Big data is a term that applies to large objects of data that vary in nature whether they are structure , unstructured or semi-structured, including from internal or external sources of the organization, and are generated with a high degree of speed with a turbulent model, which does not Fully compatible with traditional and structured data warehouses and requiring a robust and complex ecosystem with a high-performance computing platform and analytical capabilities to capture, process, transform, detect, extract and derive value and deep insights within a timed time. Machine learning is a type of Artificial Intelligence, which allows software applications to become more accurate in predicting results without explicitly programming them. The primary focus of machine learning is to build algorithms that can receive input data and use statistical analysis to predict outputs within an acceptable range. In this paper we will introduce the concept of big data and machine learning and the comparison between them and the techniques used for both and finally we will see the relationship between them.

Keywords

Big data analytics, machine learning, deep learning

1. Big data Analytics

Although the term "Big Data" has become popular, there is no general consensus about what it really means. Often, many professional data analysts would imply the process of extraction, transformation, and load (ETL) for large datasets as the connotation of Big Data. A popular description of Big Data is based on three main attributes of data: volume, velocity, and variety, veracity (or 4Vs). Nevertheless, it does not capture all the aspects of Big Data accurately. In order to provide a comprehensive meaning of Big Data, we will investigate this term from a historical perspective and see how it has been evolving from yesterday's meaning to today's connotation. Historically, the term Big Data is quite vague and ill defined. It is not a precise term and does not carry a particular meaning other than the notion of its size. The word "big" is too generic; the question how "big" is big and how "small" is small is relative to time, space, and circumstance. From an evolutionary perspective, the size of "Big Data" is always evolving. If we use the current global Internet traffic capacity as a measuring stick, the meaning of Big Data volume would lie between the terabyte (TB or 10^{12} or 2^{40}) and zettabyte (ZB or 10^{21} or 2^{70}) range. Based on the historical data traffic growth rate, Cisco claimed that humans have entered the ZB era in 2015.[1]

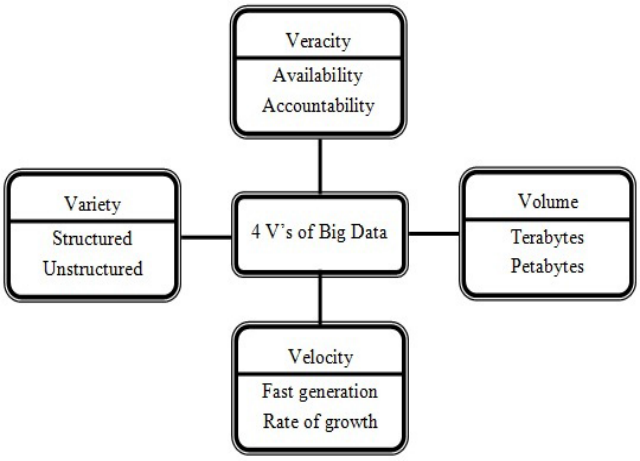

Fig. 1: Characteristics of Big Data

2. CHALLENGES IN BIG DATA ANALYTICS

The challenges of big data analytics are classified into four broad categories [2]

A. Data Storage and Analysis

In recent years the size of data has grown exponentially by various means such as mobile devices, aerial sensory technologies, remote sensing, radio frequency identification readers etc. These data are stored on spending much cost whereas they ignored or deleted finally because there is no enough space to store them. Therefore, the first challenge for big data analysis is storage mediums and higher input/output speed [2].

B. Knowledge Discovery and Computational Complexities

Knowledge discovery and representation is a prime issue in big data. It includes a number of sub fields such as authentication, archiving, management, preservation, information retrieval, and representation. There are several tools for knowledge discovery and representation such as fuzzy set rough set, soft set, near set, formal concept analysis, principal component analysis etc.

C. Scalability and Visualization of Data

The most important challenge for big data analysis techniques is its scalability and security. In the last decades researchers have paid attentions to accelerate data analysis and its speed up processors followed by Moore's Law. For the former, it is necessary to develop sampling, on-line, and multiresolution analysis techniques. Incremental techniques have good scalability property in the aspect of big data analysis. As the data size is scaling much faster than CPU speeds, there is a natural dramatic shift in processor technology being embedded with increasing number of cores. This shift in processors leads to the development of parallel computing. Real time applications like navigation, social networks, finance, internet search, timeliness etc. requires parallel computing. The objective of visualizing data is to present them more adequately using some techniques of graph theory. Graphical visualization provides the link between data with proper interpretation.

D. Information Security

In big data analysis massive amount of data are correlated, analyzed, and mined for meaningful patterns. All organizations have different policies to safe guard their sensitive information.

Preserving sensitive information is a major issue in big data analysis. There is a huge security risk associated with big data. Therefore, information security is becoming a big data analytics problem. Security of big data can be enhanced by using the techniques of authentication, authorization, and encryption.

Various security measures that big data applications face are scale of network, variety of different devices, real time security monitoring, and lack of intrusion system.

3. BIG DATA APPLICATIONS

Big data applications are being used in different industries. Fed BP Disaster response was an application for the government's [USA government] response in the disaster situation. This was built in 2010 when oil rate flow was a key issue. BP and independent groups presented changeable estimates preventing efforts to manage the level and range of the U.S. Government's response. NIST analyzed the estimates and formed actionable intelligence on which to support the final reaction. Applications in oil and gas business could be Equipment maintenance to prevent failure, production optimization, price optimization, safety and compliance. Oil and gas companies have big competition in their field, and facing regular change. Firms need to increase their production volumes and at the other hand they also want the healthy, safety and low risk environment. From exploration and production of the oil, leading companies are using big data for new business values, reduce cost and increase production [3].

4. Tools for Big Data

Table 1 Some List and Comparison of Big Data Tools [4]

Tools	Description	
Apache Hadoop	A software framework employed for clustered file system and handling of big data. It processes datasets of big data by means of the MapReduce programming model	
Knime	KNIME stands for Konstanz Information Miner which is an open source tool that is used for Enterprise reporting, integration, research, CRM, data mining, data analytics, text mining, and business intelligence. It supports Linux, OS X, and Windows operating systems.	
Datawrapper	Datawrapper is an open source platform for data visualization that aids its users to generate simple, precise and embeddable charts very quickly.	
MongoDB	MongoDB is a NoSQL, document-oriented database written in C, C++, and JavaScript. It is free to use and is an open source tool that supports multiple operating systems including Windows Vista (and later versions), OS X (10.7 and later versions), Linux, Solaris, and FreeBSD.	
Lumify	Lumify is a free and open source tool for big data fusion/integration, analytics, and visualization	
HPCC	HPCC stands for High-Performance Computing Cluster. This is a complete big data solution over a highly scalable supercomputing platform. HPCC is also referred to as DAS (Data Analytics Supercomputer). This tool was developed by LexisNexis Risk Solutions.	
Storm	Apache Storm is a cross-platform, distributed stream processing and fault-tolerant real-time computational framework. It is free and open-source. The developers of the storm include Backtype and Twitter. It is written in Clojure and Java.	
Cassandra	Apache Cassandra is free of cost and open-source distributed NoSQL DBMS constructed to manage huge volumes of data spread across numerous commodity servers, delivering high availability. It employs CQL (Cassandra Structure Language) to interact with the database.	

5. Machine learning

Is a paradigm that may refer to learning from past experience (which in this case is previous data) to improve future performance. The sole focus of this field is automatic learning methods. Learning refers to modification or improvement of algorithm based on past "experiences" automatically without any external assistance from human [5]. Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets. These procedures operate by construction of a model from example inputs in order to make data-driven predictions or choices rather than following firm static program instructions [6].

Machine learning involves two types of tasks [6]:

A. Supervised machine learning:
On a pre-defined set of "training examples", which then facilitate its ability to reach an accurate conclusion when given new data.

B. Unsupervised machine learning:
This program given bunch of data and must find pattern and relationship therein.

6. DEEP LEARNING

A new area of machine learning research, which has been introduced with the objective of moving machine learning closer to one of its original goals: Artificial Intelligence. Deep learning draws its roots from Neocognitron; an Artificial Neuron Network (ANN) introduced by Kunihiko Fukushima in 1980. An ANN is an interconnected network of processing units emulating the network of neurons in the brain. The idea behind ANN was to develop a learning method by modeling the human brain. Deep learning is a method to train multi-layer (and hence the word "deep") ANN using little data. This is the reason why ANN is back in the game. Using an example to compare Machine Learning with Deep Learning, we can say that if a machine learning algorithm learns parts of a face like eyes and nose for face detection tasks, a deep learning algorithm will learn extra features like the distance between eyes and the length of the nose [6].

Application of machine learning [7]:

A. DNA classification: Understanding genomics

Figure 2 shows a DNA microarray data, the colors, red, green, gray and so on, show the degree to which different individuals do or do not have a specific gene. The idea is to form a group of different individuals such that each of them has a certain gene. So a clustering algorithm can be run to group individuals into different categories or into different types of people. So this is Unsupervised Learning because the algorithm is not given any information in advance whether there are type 1 people, type 2 persons, and type 3 persons and so on. Instead a bunch of data is given and the algorithm automatically finds structure in the data into these types of individuals.

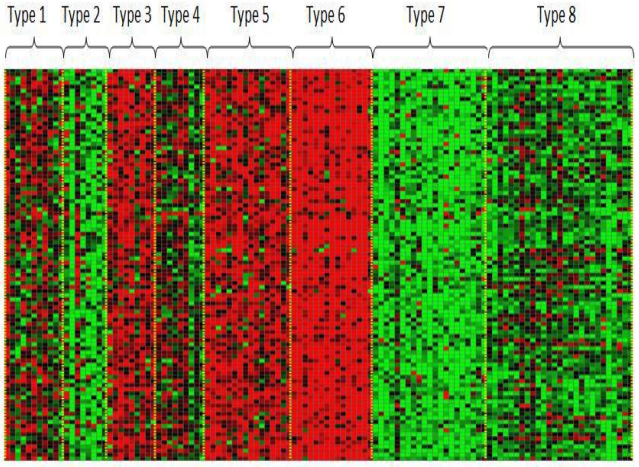

Figure 2: DNA microarray data

B. Organizing large computer clusters

At large data centers that are large computer clusters, unsupervised learning helps to figure out which machines tend to work together, so that if those machines are put together or if there is some crisis, then the data centers can work more efficiently.

C. Social network analysis:

Unsupervised Machine learning algorithms can automatically identify the friends within a user circle in Facebook or Google, or it can identify the maximum number of mails sent to a particular person and categorize into collective groups. It also identifies which are groups of people that all know each other.

D. Market segmentation:

Many companies have huge databases of customer information. So, Unsupervised Machine learning algorithms can look at this customer data set and automatically discover market segments and automatically group customers into different market segments so that the company can automatically and more efficiently sell or market the different market segments together. Again, this is Unsupervised Learning because it is not known in advance what the market segments are, or which customer belongs to which segment.

E. Astronomical data analysis:

These clustering algorithms give surprisingly interesting useful theories of how galaxies are born .

F. Analysis of gene expression data: cancer diagnosis:

Cancer can be defined as a class of diseases that is characterized with out of control cell growth. There are about a 100 different types of cancer claiming the lives of innumerable people across the world. Thus identifying the type of cancer is a crucial step in its treatment. It is done through classification of patient samples. The classification process and results may be improved by analyzing the gene expression of the patient which may provide additional information to the doctors.

G. Anomaly/Novelty detection in astronomical data:

Modern astronomical observatories are very advanced and can produce massive amount of data which the researchers don't even have time to look at. Sometimes the researchers even lack the adequate knowledge; experience and training to 33 deduce the exact significance or meaning of these data sets. It is not unusual that these large-scale astronomical data sets can contain anomalies/novelties. Thus the need for machines which can be trained to go through the data generated and in the process detect any anomalies that may be present in the data set (at a much faster rate and in most cases with better accuracy) becomes evident. Anomaly/Novelty Detection is the process of finding unusual things or characteristics which are different from our prevalent knowledge about the data. Anomalies detection problems are primarily of two types: 1) point anomaly - anomalies of this kind are individual celestial objects that present unusual characteristics. 2) group anomalies - this is an unusual collection of points. A group of points can be considered abnormal either because it is a collection of anomalous points, or because that the way its member points aggregate is unusual, even if the points themselves are perfectly normal .

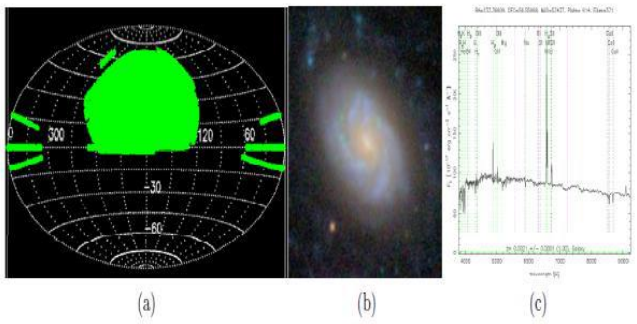

Figure 3: Summary of the Sloan Digital Sky Survey (SDSS) data set. (a) The coverage map of SDSS. (b) A sample imaging data. (c) A sample spectroscopic data i.e. the spectrum.

H. Computational biology:

Computational biology also known as bioinformatics is the use of biological data to develop algorithms and establish relations among various biological systems. With automation again, biologists are collecting lots of data about gene sequences, DNA sequences, gene expression array analysis, combinatorial chemistry and so on, and machines running algorithms are providing a much better understanding of the human genome, and what it means to be human.

I. Handwriting recognition:

It turns out one of the reasons it's so inexpensive today to route a piece of mail across the countries, is that when an address is written on an envelope, it turns out there's a learning algorithm that has learned how to read the handwriting so that it can automatically route this envelope on its way, and so it costs less.

J. Face recognition:

Human face is not unique, rigid object and numerous factors cause the appearance of the face to vary. There are numerous application areas where face recognition can be exploited such as security measure at an ATM, areas of surveillance, closed circuit cameras, image database investigation, criminal justice system, and image tagging in social networking sites like Face book etc [6].

K. Speech recognition:

All speech recognition software utilizes machine learning. Speech recognition systems involve two distinct learning phases: one before the software is shipped (training the general system in a speaker-independent fashion), and a second phase after the user purchases the software (to achieve greater accuracy by training in a speaker -dependent fashion).

L. Information retrieval:

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). The user provides an outline of their requirements—perhaps a list of keywords relating to the topic in question, or even an example document. The system searches its database for documents that are related to the user's query, and presents those that are most relevant. The information retrieval process can be divided into four distinct phases: indexing, querying, comparison, and feedback. All phases of information retrieval can be performed manually, but automation has many benefits—larger document collections can be processed more quickly and consistently, and new techniques can be easily implemented and tested. The instant availability of enormous amounts of textual information on the Internet and in digital libraries has provoked a new interest in software agents that act on behalf of users, sifting through what is there to identify documents that may be relevant to users' individual needs .

M. Intrusion detection:

Intrusion detection is the process of monitoring the events that are occurring in the systems or networks and analyzing them for signs of possible incidents, which are violations or threats to computer security policies, acceptable use policies, or standard security practices It is mainly of two types based on the intrusions first is Misuse or signature based detection and the other is Anomaly detection.

7. Tools for machine learning

Table 2-7 Some List and Comparison of machine learning Tools [8]

tools	platform	cost	language	Algorithms or feature
Scikit learn	Linux Mac os windows	free	Python Cython C,c++	classification Regression Clustering Preprocessing Model... its
PyTorch	Linux Mac os windows	free	Python C++ Cuda	Autograd Module Optim module nn module
TensorFlow	Linux Mac os windows	free	Python C++ Cuda	Provides a library for dataflow programing
weka	Linux Mac os windows	free	java	Data preparation classification Regression Clustering Visualization Association
KNIME	Linux Mac os windows	free	java	Can work with large data volume support text mining ..its
colap	Cloud service	free		Support libraries of pytorch,keras,tensorflow and OpenCV
Kears	Cross-platform	free	python	Python

8. Algorithms of machine learning

Table 3-8 some machine learning Algorithms [5]

Algorithms	Descriptions
Regression Algorithms	: Linear Regression, Logistic Regression, Stepwise Regression , Ordinary Least Squares Regression (OLSR), Multivariate Adaptive Regression Splines (MARS) , Locally Estimated Scatterplot Smoothing (LOESS) etc.
Instance-based Algorithms	K-Nearest Neighbour (KNN), Learning Vector Quantisation (LVQ), Self-Organising Map (SOM), Locally Weighted Learning (LWL) etc.
Regularization Algorithm	Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO) , Elastic Net, Least-Angle Regression (LARS)
Decision Tree Algorithms	Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0, Chi-squared Automatic Interaction Detection (CHAID) , Decision Stump, M5, Conditional Decision Trees etc.
Bayesian Algorithms	Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Averaged One-Dependence Estimators (AODE), Bayesian Belief Network (BBN), Bayesian Network (BN)
Support Vector Machine (SVM)	Based on the kernel in use, SVM can perform both linear and nonlinear classification.
Clustering Algorithms	K-Means, K-Medians, Affinity Propagation, Spectral Clustering, Ward hierarchical clustering, Agglomerative clustering. DBSCAN, Gaussian Mixtures, Birch, Mean Shift, Expectation Maximisation (EM)
Association Rule Learning Algorithms	Apriori algorithm, Eclat algorithm
Artificial Neural Network (ANN) Algorithms	Perceptron, Back- Propagation, Hop-field Network, Radial Basis Function Network (RBFN)
Deep Learning Algorithms	Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN), Stacked Auto-Encoders

9. Literature review

Jafar Raza Alam[1], Asma Sajid[2], Ramzan Talib[3], Muneeb **Niaz[4]** [3] **,** A Review on the Role of Big Data in Business , Big data is a game changing thing. Successful organizations are achieving business advantages by analyzing big data. It has received significant attention in recent years but some challenges are one of the major causes in diminishing the growth of organizations. The main issues why these organizations are not begin their planning stage to implement the big data strategy because they don't know enough about the big data and they don't understand the benefits of big data. In this study, an attempt is made to review the role of big data in the business. Result Big data must be integrated in the organization's architecture, even the organization have their well established and large businesses. Countries in the world, IT companies and the relevant departments have started working on big data

Ms. Komal[1] [9], A Review Paper on Big Data Analytics Tools, Everyday billions of people fetch, upload and share information on social media and other platforms through mobile phones, laptops, PDAs. The information comprises of pictures, blobs, goggle map locations, videos, text, voice messages that are collection of structured, unstructured and complex data objects. Traditional data processing techniques are insufficient to handle this enormous, heterogeneous and fast-paced data. E-commerce and digital marketing has gained so much popularity over these years that business industry has become more dependent on online transactions and services. Big data analytics has proven to be a boon for such an industry as it helps to extract useful patterns and unknown correlations of potential consumer market, client preferences, buying attributes and lot of other information from intricate data sources. The result, the rate of development of information processing tools is comparatively much slower than the rate of development of information. Currently available tools in the market do not address all the issues of Big Data analytics. Even the most high-tech tools and techniques like Hadoop, Cassandra and Ignite can't justify real-time analysis in true sense.

D. P. Acharjya, Kauser Ahmed P[2], A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, A huge repository of terabytes of data is generated each day from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data challenges, open research issues, and various tools associated with it. As a result, this article provides a platform to explore big data at numerous stages.

Annina Simon[1], Mahima Singh Deo[2], et al [6], An Overview of MachinemLearning and its Applications, The possibility of this research paper is to create attentiveness among upcoming scholars about recent advances in technology, specifically deep learning an area of machine learning which finds applications in big data analytics and artificial intelligence. Result Deep learning techniques have been criticized because there is no way of representing causal relationships (such as between diseases and their symptoms), and the algorithms fail to acquire abstract ideas like "sibling" or "identical to." Not much theory is available for most of the methods which is disadvantageous to beginners. Deep Learning is only a small step towards building machines which have human-like intelligence. Further advancements must be made in order to achieve our ultimate goal.

Kajaree Das[1], Rabi Narayan Behera[2] [5] , A Survey on Machine Learning: Concept, Algorithms and Applications

Over the past few decades, Machine Learning (ML) has evolved from the endeavor of few computers Enthusiasts exploiting the possibility of computers learning to play games, and a part of Mathematics (Statistics) that seldom considered computational approaches, to an independent research discipline that has not only provided the necessary base for statistical-computational principles of learning procedures, but also has developed various algorithms that are regularly used for text interpretation, pattern recognition, and a many other commercial purposes and has led to a separate research interest in data mining to identify hidden regularities or irregularities in social data that growing by second. This paper focuses on explaining the concept and evolution of Machine Learning, some of the popular Machine Learning algorithms and tries to compare three most popular algorithms based on some basic notions. Sentiment140 dataset was used and performance of each algorithm in terms of training time, prediction time and accuracy of prediction have been documented and compared.

Ayon Dey[10] , Machine Learning Algorithms: A Review

In this paper, various machine learning algorithms have been discussed. These algorithms are used for various purposes like data mining, image processing, predictive analytics, etc. to name a few. The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically.

10. Comparison between big data and machine learning

Table 4-10 Comparison

Big data	Machine learning
For data use : Big data can be used for variety of purpose: including :financial research ,collecting sales data	**For data use :**Machine learning is the technology behind self-driving cars and advance recommendation engines
Foundation for learning : big data analytics pulls from existing information to look emerging patterns that can help shape our decision making processes	**Foundation for learning :** Machine learning can learn from the existing data and provide the foundation required for machine to teach itself
For pattern recognition :Big data analytics can reveal some patterns through classification and sequence analysis	**For pattern recognition** Machine learning takes this concept a one-step ahead by using the same algorithms that big data uses to automatically learn from collected data
Data volume : Big data as name suggest tends to be interested in large-scale datasets where the problem is dealing with the large volume of data	**Data volume :** Machine learning to be more interested in small datasets where over-fitting is the problem
Purpose big data is to store large volume of data and find out pattern in data	**Purpose** To learn from trained data and predicted or estimate future results
Big data discussions include storage, ingestion and extraction	Machine learning is sub field of computer science and/or artificial intelligence that gives computers the ability to learn without explicitly programmed
Big data has got more to do with high performance computing	Machine learning is part of data science

11. Relationship between big data and machine learning

Table 4 -11 Relationships

Machine learning is the process in which machine uses the data provided by big data and then try to respond accurately accordingly thus leading to improve business operation, service quality, customer relationship and more.
Machine learning gets data from big data and learns more data they have at their disposal.
Big data analytics provide different variety of data to machine to in order to show and present better results in performance of the task.
The breakthroughs in big data analytics have already seen through machine learning in sectors like healthcare, retail, financial and auto industries...its
Both are related in sense that both complement each other .one cannot exist without another
Analyzing big data using machine learning algorithms help organization forecast future trends in the market

12. Conclusion

This paper discussed the comparison and relationship between big data analytics and machine learning and their algorithms and techniques used for each other.so big data with machine learning algorithms can increase the efficiency of business operations

13. References

[1] **Rajkumar Buyya[1],Rodrigo N. Calheiros[2],Amir Vahid Dastjredi[3]** ,"Big Data Principles and Paradigms", Rajkumar Buyya "The University of Melbourne and Manjrasoft Pty Ltd", Australia, Rodrigo N. Calheiros "The University of Melbourne", Australia, Amir Vahid Dastjerdi "The University of Melbourne", Australia ,2016

[2] **D. P. Acharjya[1], Kauser Ahmed P[2]**:"A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools", *(IJACSA)* International Journal of Advanced Computer Science and Applications, "School of Computing Science and Engineering VIT University Vellore", India 632014, *2016*

[3] **Jafar Raza Alam[1], Asma Sajid[2], Ramzan Talib[3], Muneeb Niaz[4],** "A Review on the Role of Big Data in Business" International Journal of Computer Science and Mobile Computing, [1,2,3,4] Colleges of Computer Science and Information Studies, Government College University, Faisalabad, Pakistan, 2014

[4] **Big data tools**

https://www.google.com/amp/s/www.softwaretestinghelp.com/big-data-tools/amp/

[5] **Kajaree Das[1], Rabi Narayan Behera[2]"** A Survey on Machine Learning: Concept, Algorithms and Applications" **International Journal of Innovative Research in Computer and Communication Engineering** ,B.Tech, Dept. of I.T., Institute of Engineering and Management, Kolkata, India[1] Asst. Professor, Dept. of I.T., Institute of Engineering and Management, Kolkata, India [2] 2017.

[6] **Annina Simon[1], Mahima Singh Deo[2] Venkatesan[3], D.R. Ramesh Babu[4]** "An Overview of Machine Learning and its Applications, International Journal of Electrical Sciences & Engineering (IJESE), [1,2]Prefinal Year Dept. of Computer Science and Engineering, [3,4] Professor, Dept, of Computer Science and Engineering,2015

[7] **Sumit Das[1], Aritra Dey[2], Akash Pal[2], [3] Nabamita Roy,"** Applications of Artificial Intelligence in Machine Learning: Review and Prospect", International Journal of Computer Applications (0975 – 8887), JIS College of Engineering, Kalyani, India, 2015

[8] **Machine learning Tools**
"https://www.google.com/amp/s/www.softwaretestinghelp.com/machine-learning-tools/amp/"

[9] **Ms. Komal[1],**" A Review Paper on Big Data Analytics Tools" International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)", Department of Computer Science, Amity University Haryana, 2018

[10] **Ayon Dey**," Machine Learning Algorithms: A Review (IJCSIT) International Journal of Computer Science and Information Technologies, Department of CSE, Gautam Buddha University,Greater Noida, Uttar Pradesh, India,2016

[11] **big data vs. machine learning**
https://www.educba.com/big-data-vs-machine-learning/