A comparison study Big Data Analytics Methods for Selecting Suitable Method

Samah I. Abdel Aal
Faculty of Computers and Informatics, Zagazig University, Egypt
Email: dr.samahibrahim2018@gmail.com
eng_samah2013@yahoo.com

Abstract

Nowadays, Big Data become very critical and can help organizations to make decisions by analyzing feedback and reviews from customers. There are large amounts of data that is growing because of the extensive use of networks, social media, and other sources. Big data analytics can enhance a company's understanding of a customer's needs and preferences. By analyzing data organizations can personalize and relevant product or service; So that, there is a need to analyze such huge amounts of data. There are many types of analytics methods and each one has its own objectives. Therefore, this work highlights two ways for classifying big data analytics methods. Big data models and their methodologies were clearly outlined in this paper. Additionally we compare different categories of big data analytics techniques depending on data kinds like audio, video, social media, and predictive analytics. Also, this work intended to study the big data analytics methods for selecting the suitable method.

Keywords: Information Extraction (IE); Question Answering (QA); Natural Language Processing (NLP).

1. Introduction

The big data concept may have originated during lunch at Silicon Graphics Inc. (SGI) in the mid-1990s 2011[1].The main characteristic of large data is its size. According to Laney, there are three dimensions of data management problems: volume, variety, and velocity [1,2, 3]. Big data analytics enhances a company's understanding of a customer'sneeds and preferences. By analysing data organisations can personalizeand relevant product or service offers [10].After organizations have collected large amounts of data, the next critical step is to begin using analytics. Many businesses are unsure where to start, what categories of analytics will help them expand, or what these various types of analytics represent. Data is rapidly changing and become more complicated, and variable, so that such huge amounts of data need to be managed and analyzed using current methods and technology [4].Due to the difficulties of computational efficiency and the unique qualities of big data, new approaches to manage massive data and get considerable insight should be created. There is a need to analyze such huge amounts of data. For analyzing such huge amounts of data, there are many types of analytics methods and each one has its own objectives. Therefore, this work highlights two ways for classifying big data analytics methods. Also, this work intended to study the big data analytics methods for selecting the suitable method. This paper is organized as the follows: the first section presents the introduction for this work; the second section introducesProcesses for analyzing big data; the third section introduces Benefits of big data analytics; fourth sectionhighlights the main Types of Big Data Analytics Methods; the fifth section gives the conclusion; finally gives references.

2. Processes for analyzing big data

According to [5][7][8] Big data characteristics involves: Volume, Velocity, Veracity, Variety, and Value.The following highlights these characteristics

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

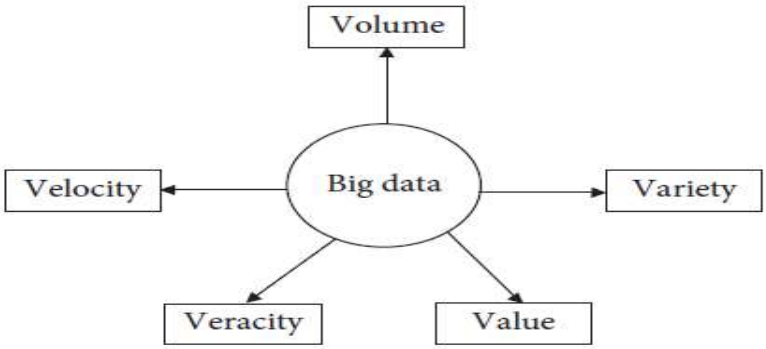

Figure 1: Big data characteristics

Data Analytics is a science, which examine data and draw decisions from it. Itenables us to understand what occurred and what will occur. Data analytics is applied by organizations to make better business decisions. The big data processes involves data management and analytics shown in the following figure

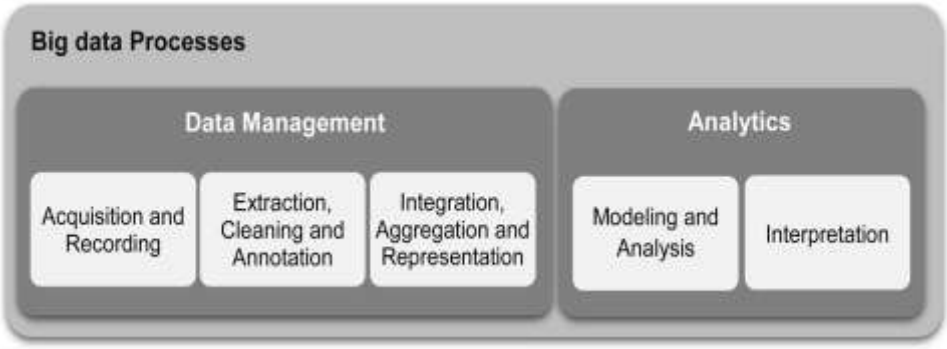

Figure 2: Processes for analyzing big data

3. Benefits of big data analytics

Big Data analytics help organizations to identify new opportunities [6] and can enhance day to day business operations to achieve customer satisfaction. The following figure can summarize the benefits of big data analytics

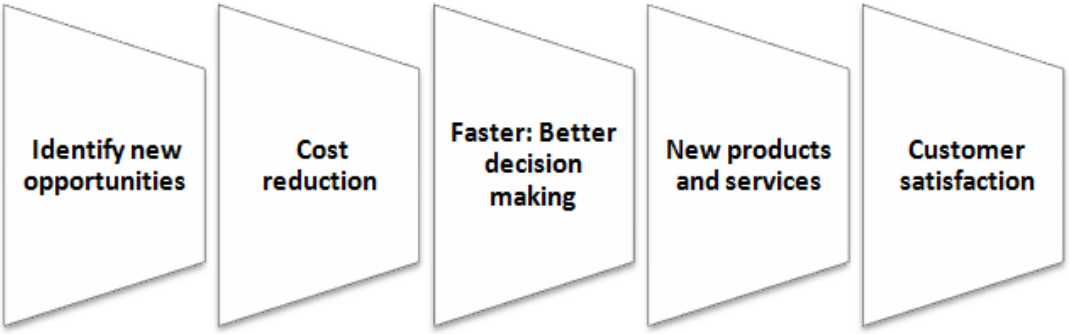

Figure 3: Benefits of big data analytics

4. Types of Big Data Analytics Methods

This paper highlights two ways for classifying of big data analytics methods as the follows:

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

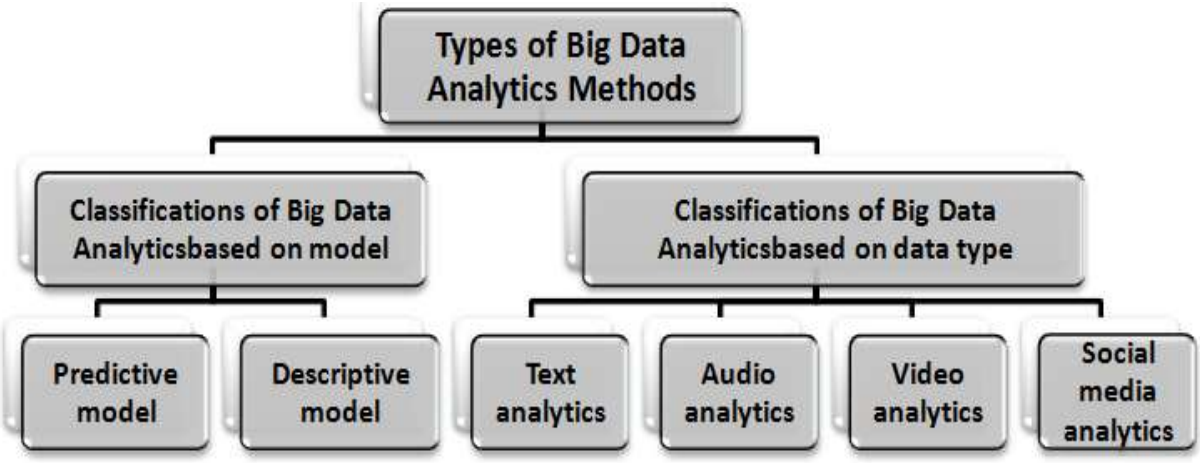

Figure 4: Types of Big Data Analytics Methods

4.1. Classifications of Big Data Analyticsbased on model

This work will address predictive model and descriptive model as the follows:

A. Predictive Model

The most of predictive analytics approaches depend on two methodologies: *regression techniques and machine learning techniques* to prediction based on historical and present data (e.g. neural networks). Big data is so large and has properties like heterogeneity, noise building, misleading correlation, and unintended, new statistical methods are needed [20]. The predictive model involves:

- Multiple regression
- Time serious analysis
- Classification
- Machine Learning

The following table introduces a comparison between: regression analysis classification, time serious analysis, and prediction analysis

Table 1: A comparison between regression analysis classification, time serious analysis, and machine learning

Classification	Identify a certain class of items, many qualities can be employed. Identify what will occur within a class, classification allocates objects to target groups or classes. Classification involves the process of identifying a collection of models that characterize and differentiate data classes or ideas based on a set of training data [12].
Regression Analysis	It tries to express the interdependence of variables. Estimating and prediction are common applications
Time Serious Analysis	It consists of a set of well-defined data points collected at regular intervals.
Machine Learning	Use machine learning techniques to predict future.

B. Descriptive Model

This type interprets data without focusing on a specific variable. The descriptive model includes [11]:

- Cluster analysis
- Association
- Summarization
- Sequence Discovery Analysis

The following table introduces a comparison between: association, summarization, clustering, and sequence discovery analysis

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

Table 2: a comparison between: Association, Summarization, Clustering, and Sequence Discovery Analysis

Association	Clustering	Summarization Analysis	Sequence Discovery Analysis
In order to find a pattern, association establishes a link between two or more objects. A supermarket, for example, could see that consumers frequently to discover common product patterns. .	Clustering is the process of grouping similar records together. Clustering is the process of dividing a dataset into groups. Items in a dataset are grouped or clustered according to the idea that objects in one cluster are extremely similar to one another but quite different from objects in other clusters. [13].	Methods for obtaining a brief description of a dataset are covered. For example, summarizing a huge number of items linked to Christmas season sales offers a basic overview of the data that sales and marketing managers can find valuable data.	It is all about knowing how to accomplish things in a certain order. It all comes down to the user's purchase order, which the business owner may then organize [14].

4.2. Classifying big data Analytics Techniques based on type of data

The big data Analytics Techniques can be classified based on type of data [9]. These types include:

1- Text analytics
2- Audio analytics
3- Video analytics
4- Social media analytics

The following table introduces a comparison between big data Analytics Techniques based on type of data

Table 3: A comparison between big data Analytics Techniques based on type of data

	Text Analytics	Audio analytics	Video analytics	Social media analytics
TYPE OF DATA	Text mining	Speech analytics	Video content analysis (VCA)	Social medial data
FIELDS USED	Emails, survey replies, business papers, news	Medical FeedbackArea	Social media Facebook and other Fields	Social networks, , social news, social media sharing
METHODS	Text summarization techniques Opinion mining	Transcript based approach Phonetic-based approach	Server-based and edge-based	Content-based analytics Structure-based analytics
ADVANTAGES	Can be used to make evidence-based decisions.	Analyze hundreds, recorded calls using audio analytics.	Can execute surveillance activities efficiently and identifying things	Provide an umbrella word for a number of internet platforms that enable users to

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

	Can be used to forecast the stock market using data gleaned from financial news[10] .	Evaluate conversation Provide suggestions based on the customer's previous. Real-time feedback [5].	that have been removed [5]	produce and share information.
DIS-ADVANTAGES	1. It has nothing to do with technology and everything to do with copyright laws. 2. No ways to protect copyright holders' rights.	1. It's simpler to lose track of time. 2.Pay more attention to the images than to the audio.	1. Additional considerations about data management 2. A higher total cost 3. Aggregation can be difficult[19] .	1. Results from real-time processing might be skewed. 2. Many analytics companies claim to be able to examine this stream of information in real time.

A. Text analytic

Text analytic methods focus on extracting information from textual data. Textual data whin organizations include social media feeds, emails, blogs, online forums, survey responses, corporate documents, news, and call center records[14]. The following table introduces a comparison between four text analytics methods

Table 4: A comparison between fourText analytics methods

Information Extraction (IE)	Text Summarization Techniques	Question Answering (QA)	Sentiment Analysis (opinion mining)
There are Two sub-tasks in IE are Entity Recognition (ER) [14]and Relation Extraction (RE)[15]. 1.ER extracts names in text and classifies them into predefined categories 2.REextracts and extracts semantic relationships between entities.	This method provides a summary of a single or more documents automatically. It include: 1.Extractiveapproach identifying and connecting the most important text units [15]. 2. Abstractive summarization the text and creates the summary using Natural Language Processing (NLP).	There are three approach of this technique : • Information Retrieval • knowledge-based • Hybrid approach[14].	examine opinionated material, which comprises people's feelings about things like products, organizations, customers, and events.. [17]. Document-level, sentence-level, and aspect-based sentiment analysis approaches are the three sub-groups of sentiment analysis techniques.

B. Audio analytics

Speech analytics or audio analytics focus on extracting data from audio resources. The following table introduces a comparison between two speech analytics methods as the follows

Table 5: A comparison between vocabulary continuous speech recognition, and phonetic-based

vocabulary continuous speech recognition	phonetic-based approach

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

systems involves two steps: indexing and searching	Focus on sounds and phonemes. It distinct types of sound in a specified language which distinguish one word from another.

C. Video analytics

Video analytics includes a number of techniques for analyzingand findingmeaning data from videos. The following table introduces a comparison between two video analytics as the follows:

Table 6: A comparison between two video analytics methods

	Server-based architecture	Edge-based architecture
Feature set	Videos captured through camera to a centralized server that performs the video analytics [18]**.**	Analytics are applied at the 'edge' of the system.
performance	medium to high performance due to limited quality of the videos	Low to high due to limited processing resources.

D. Social media analytics

Social media analytics is the study of data from social media. It involves opinions, pictures, videos, and relationships between network entities in social media (e.g., individuals, companies, and goods).The following table introduces a comparison between three social media analytics methods

Table 7: A comparison between three social media analytics methods

Community detection	Social influence analysis	Link prediction
Community detection, focus on finding implicit communities within a network. Applying Bayesian network and graph mining technique etc. [18]	Social influence focus on modeling and assessing the impact of social network participants [17][19]	Forecast the occurrence of interaction, cooperation, and between network participants. [21]

5. Conclusion

In this study, the fundamental methods of big data analysis methods are investigated. Big data analytics needs to be studied extensively in order to manage firms and understand customer needs. The various big data models and their methodologies were clearly highlighted in this paper. Additionally, we compare various categories of big data analytics techniques depending on data kinds like audio, video, social media, and predictive analytics.This work tented to highlight different types of the big data Analytics methods. This work highlights two ways for classifying big data analytics methods. Also, this work intended to study the big data analytics methods for selecting the suitable method.

References

[1] F. X. Diebold, "A personal perspective on the origin(s) and develop-ment of "big data"."*ScholarlyPaper No. ID 2202843)*, Diebold, F. X. (2012)..

[2] Chen, " Business intelligence and analytics:From big data to big impact.,"*MIS Quarterly, 36(4), 1165–1188.*, (2012)..

[3] O. L. N. &. S. B. Kwon, " Data quality management, data usage expe-rience and acquisition intention of big data analytics.,"*International Journal ofInformation Management*, pp. 34(3), 387–394., 2014.

[4] G. I. G. (n.d.). [Online]. Available: http://www.gartner.com/it-glossary/big-data/.

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023

[5] M. H. Amir Gandomi, "Beyond the hype: Big data concepts, methods, and analytics,,"*International Journal of Information Management*, no. 0268-4012, pp. 137-144,.

[6] AmirGandomiMurtazaHaider, "Big data concepts, methods, and analytics,"*International Journal of Information Management*, 2014.

[7] "Vendor Independent Big Data Qualifications," [Online]. Available: https://www.bigdataframework.org/data-types-structured-vs-unstructured-data/.

[8] "THE FOUR V'S OF BIG DATA," [Online]. Available: https://opensistemas.com/en/the-four-vs-of-big-data/.

[9] A. &. H. M. &. S. M. Unnisabegum, "Big data analysis Techniques For Big Data," vol. 6, no. 10.13140/RG.2.2.25408.07686., 2019.

[10] K. H. T. Y. L. &. Y. K. T. Yuanzhu Zhan, "Unlocking the power of big data in new product development,"*Springer*, 2018.

[11] B. Leventhal, " An introduction to big data analysis and other techniques for advanced analytics,"*J Direct Data Digit Mark Pract.*

[12] J. J. P. a. M. K. Han, "Big data analysis : concepts and techniques, Elsevier.," (2011). Big data analysis : concepts and techniques, Elsevier..

[13] M. A. H. ,. M. S. Ahmed UnnisaBegum, "Big data analysis Techniques For Big Data,"*International Journal of Advanced Research in Science,*, vol. 6, no. 2350-0328, August 2019.

[14] H. &. B. C. &. U. S. &. M. M. &. Y. Hassani, "Text Mining in Big Data Analytics. Big Data and Cognitive Computing.,"*researchgate*, January 2020.

[15] M. O. Udo Hahn1, "Medical Information Extraction in the Age of Deep Learning," 2020 Aug 21..

[16] M. AmirGandomi, "Beyond the hype: Big data concepts, methods, and analytics,"*International Journal of Information Management*, vol. 35, no. 2, pp. 137-144, April 2015.

[17] C. C. Aggarwal, "Opinion Mining and Sentiment Analysis," March 2018.

[18] I. E. Olatunji, "Video Analytics for Visual Surveillance and Applications: An Overview and Survey,"*researchgate*, July 2019.

[19] Ahmad Arinaldi∗, Jaka Arya Pradana, Arlan Arventa Gurusinga, "Detection and classification of vehicles for traffic video analytic," 2018.

[20] G. Yilma, "Social Sentimental Analytics using Big Data Tools,"*researchgate*, January 2016.

[21] J. H. F. &. L. H. 3Fan, " Challenges of big data analysis. National ScienceReview," pp. , 1(2), 293–314., .2014.

Doi: https://doi.org/10.54216/IJAACI.040201
Received: April 13, 2023 Revised: June 22, 2023 Accepted: August, 10, 2023