Data Analytics Tools in Big Data

[1]*Reena Mehla,* [1]*Seema*

Sharma

[1]Assistant Professor

Abstract

In the information era, enormous amounts of data have become available on hand to decision-makers. Big data refers to datasets that are not only big but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle andextract value and knowledge from these datasets. Furthermore, decision-makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of advanced analytics techniques onbig data. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains. Data analysis in big data refers to the process of examining and interpreting large and complex data sets to uncover patterns, trends, and insights that can drive strategic decision-making. To handle the sheer volume, variety, and velocity of big data, specialized tools are used. These tools provide the necessary capabilities to process, analyze, and visualize large datasets to extract meaningful information.

Keywords: big data, data mining, analytics, decision making

1. Introduction

Big data analysis tools are designed to handle massive amounts of structured, semi-structured, and unstructured data. They leverage advanced algorithms, distributed computing, and parallel processing to efficiently process and analyze data across distributed systems or clustersof computers.

These tools offer a wide range of functionalities, including data integration,data cleansing, data transformation, statistical analysis, predictive modeling, machine learning, and data visualization. They allow organizations to extract valuable insights from large datasets that can

improve decision-making, optimize operations, enhance customer experiences, and drive innovation.

The choice of a data analysis tool for big data depends on various factors such as the nature of the data, the analytical requirements, scalability, performance, ease of use, and cost. Open-source frameworks like ApacheHadoop, Apache Spark, Apache Storm, and Apache Flink have gained significant popularity in the big data analytics space due to their scalability,flexibility, and wide community support.

Commercial tools like Splunk and Tableau provide user-friendly interfaces and powerful capabilities for data visualization and exploration. They enable business users to easily analyze and interpret large datasets without extensive programming skills.

Data analysis refers to the process of inspecting, cleaning, transforming, and modeling data to identify patterns, draw conclusions, and make informed decisions. It involves the systematic examination of data to revealmeaningful insights, relationships, and trends.

Data analysis can be performed on various types of data, including structured, semi-structured, and unstructured data. It utilizes statisticaltechniques, mathematical models, and algorithms to extract valuable information from raw data.

The goal of data analysis is to uncover insights or answers to specific questions, solve problems, or generate hypotheses for further investigation.It involves processes like data cleaning, where outliers and errors are corrected, data transformation and aggregation, where data is organized and prepared for analysis, and statistical analysis and modeling, where patterns and relationships are identified.

Data analysis can be descriptive, aiming to summarize and present data ina meaningful way, or inferential, where statistical analysis is used to make predictions or draw conclusions about a population based on a sample.

The insights gained from data analysis can have numerous applicationsacross various industries and domains. It can be used for business intelligence, marketing analysis, financial forecasting, healthcare research,fraud detection, risk assessment, customer behavior analysis, and much more.

2. **Data Analysis Tools**

There are several data analysis tools available for big data processing.Some of the commonly used ones are:

1. **Apache Hadoop:** Hadoop is an open-source framework that allows distributed processing of large data sets across clusters of computers usingsimple programming models. It provides the Hadoop Distributed File System (HDFS) for storage and the MapReduce programming model for processing.

2. **Apache Spark:** Apache Spark is an open-source cluster computing framework that provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. It supports various programming languages like Java, Scala, and Python, and has libraries forSQL, machine learning, and graph processing.

3. **Apache Storm:** Apache Storm is a distributed real-time computation system for processing large volumes of real-time streaming data. It is fault- tolerant and can process data at high speed, making it suitable for real-timeanalytics and stream processing.

4. **Apache Flink:** Apache Flink is an open-source stream processing framework with batch processing capabilities. It provides high-throughput and low-latency processing of big data with support for event time and out-of-order processing. It also has integrated machine learning libraries for advanced analytics.

5. **Apache Kafka:** Apache Kafka is a distributed streaming platform that isoften used as a data ingestion tool for big data processing. It can handle high-volume, real-time data streams and provides fault-tolerance and scalability.

6. **Splunk:** Splunk is a data analytics tool that allows you to collect, analyze,and visualize machine-generated data. It can ingest and process data froma wide range of sources and provides interactive dashboards and reports for real-time analysis.

7. **Tableau:** Tableau is a data visualization tool that allows you to connect to various data sources, including big data repositories. It provides a drag-and-drop interface for creating interactive dashboards and visualizations, making it easy to explore and analyze large datasets. These are just a few examples of the many data analysis tools available forbig data processing. The choice of tool depends on the specific use case, data volume, processing requirements, and

the skills and preferences of the data analysts.

In conclusion, data analysis in big data requires specialized tools that canhandle the volume, variety, and velocity of large data sets. These tools empower organizations to uncover actionable insights from big data, leading to improved decision-making and a competitive advantage in today's data-driven world.

Overall, data analysis plays a crucial role in understanding and making sense of the vast amounts of data we generate and collect. It enables organizations, researchers, and individuals to make informed decisions andimprove their understanding of the world.

In this research, a literature review on big data analytics, deep learning andits algorithms, and machine learning and related methods has been considered. As a result, a conceptual model is provided to show the relation of the algorithms that helps researchers and practitioners in deploying BDA on IOT data.

The process of discussing over DL and ML methods has been shown in Figure 1.
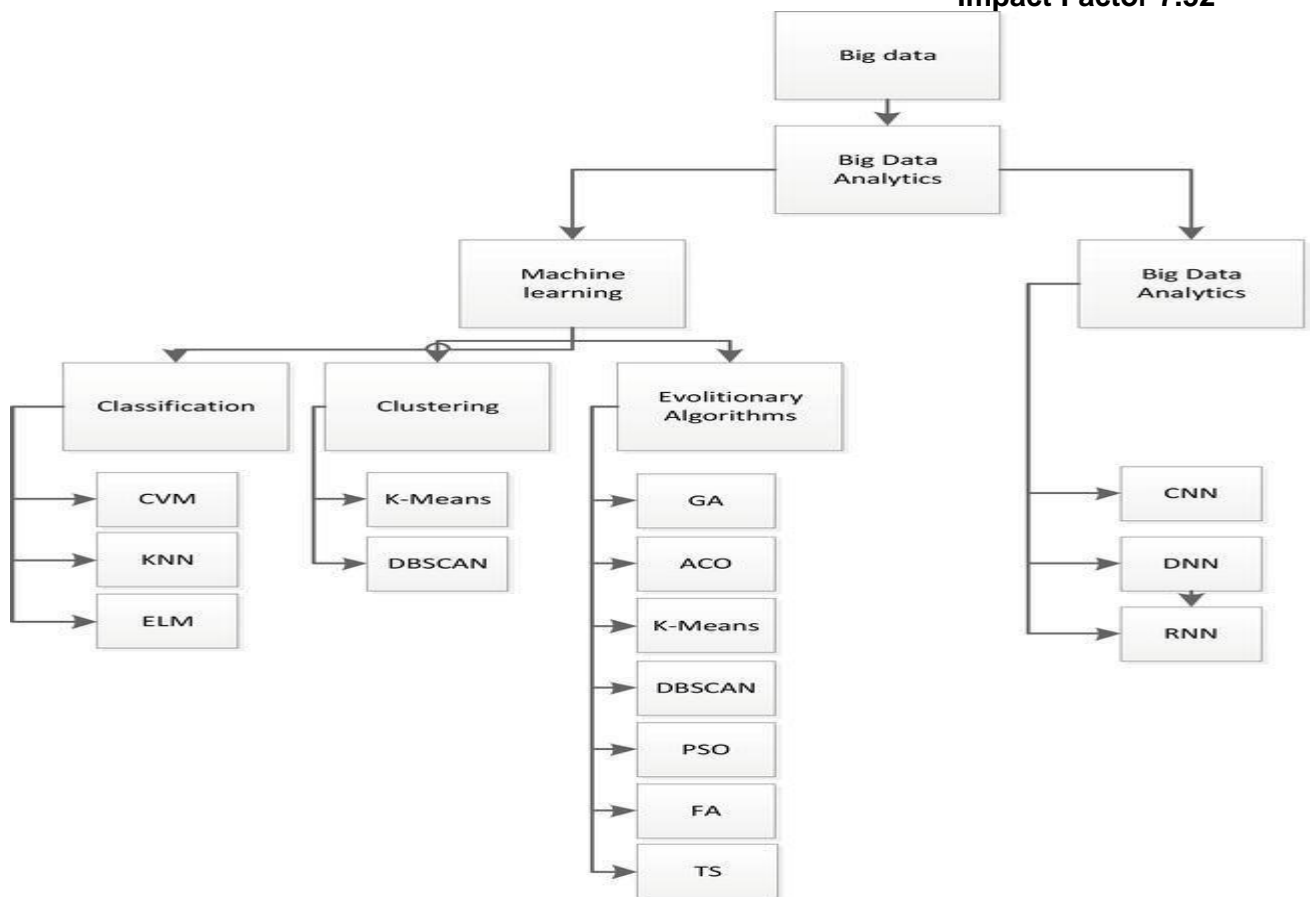
Figure 1. The big data analytics methods in this research

3. APPLICATIONS

Big Data is gradually turning popular. Each area in marketing, healthcare, automation, manufacturing industries can now implement big data analytics.

a. Improving Security and Law Enforcement: In improving security and enabling law enforcement, big data is applied greatly. In U.S, The National Security Agency (NSA) makes use of big data analytics to outwit terrorist strategies. Cyberattacks are detected and prevented using big data techniques, even police servicesuse big data tools to catch criminals and still predict criminal activities. Big data techniques are also used by credit card companies to spot fake dealings.

b. **Healthcare Sector:** Inspite, of the fact that the healthcare sector have access to vast quantities of data. The health care sector has beenoverwhelmed by failures in utilizing the data to restrain

theamount of rising healthcare and by unproductive systems that throttle faster and better healthcare benefits athwart the board. It is majorly because of the aspect that the electronic data is out of stock, scarce, or unfeasible. Furthermore, all the databases that seize healthcare-interrelated information has made those information complex to link data that candemonstrate patterns helpful in the medical field.

Few hospitals like Beth Israel makes use of data composed from a phone application, as of millions of patients, to permit doctors to make use of fullproof based medicines as conflicting to administering many medical orlab tests to all patients that go to the hospital. A series of tests can be resourceful, but they can also be costly and frequently turns out to beunproductive. The University of Florida uses free public health data andGoogle maps, to build visual data that will allow for quicker identification and proficient analysis of healthcareinformation that can be used in tracking the increase of chronic disease.

C. **Improving Sports Performance:** Big data analytics have now been embraced by nearlyall privileged sports. The IBM Slam Tracker tool is used for tennis tournaments, also we make use of video analyticsthat helps in tracking the quality performance of each player in games like football or baseball game. Also, to get feedback (via smart phones and clo/ud servers) on our game and how to improve it, sensor technology in sports equipments such as basket balls or golf clubs is used. Smart technology is used to track nutrition and sleep, as well as social media conversations to monitor emotional wellbeing. Numerous privileged sports teamsalso track athletes outside of the sporting environment using these smart technologies. The NFL has developed personal platform of applicationsto aid all the thirty two teams in making the finest decisions, based on everything from the weather, to the situation of the grass on the field, to statistics of each andevery player's performance in university. This is entirely inname of tactics and also as plummeting player injuries.

d. **Civilizing and Optimizing City and Countries:** Various aspects of different cities and countries can be progressed by using big data. Big data can help cities to optimize traffic flows based on actual time traffic information and also social media and weather data. Many cities at present are monitoring big data analytics with the intention of turning themselves into Smart Cities,in which the transport infrastructure and convenience processes all are tied up. Smart water meters is being used by the city of Long Beach, California to identify illegitimate

watering in actualtime and is being used to aid some homeowners cut theirwater usage by as much as 80%. The above becomes crucial whilst the state is going through the most terrible droughtin recorded history ever and governor has enacted the first state wide water limitations In Los Angeles, to manage traffic lights and to also controlthe congestion of traffic about the city makes use of data from magnetic road sensors and also traffic cameras. Approximately, forty thousand traffic signals around thecity is controlled by computerized system which has reduced traffic jamming by 16%.

d. **Financial Trading:** The area where big data finds a lot of use today is HFT (HighFrequency Trading). Algorithms of big data are used to formulate trading decisions. Nowadays, the greater part of impartiality trading currently takes place by means of data algorithms that rapidly take in account signals from the social media networks and also news websites to create, buy and selldecisions in few seconds. The markets can be scanned by computers which are programmed with intricate algorithms for a set of customizable circumstances and explore for trading opportunities. Depending onthe needs and desires of the client all these programs can be designed to function with zero human interaction / with human interaction, The most sophisticated of these programs rather than being hardcoded are also designed to modify as markets transform.

4. CONCLUSION

In this literature survey, we have discussed Big Data in detail until its existing state. It elaborates on the concepts of big data followed by the technologies and methodology used in this field. In this paper we have alsodiscussed regarding the obstacles faced by it and also about the applications. Lastly we have discussed topic related to future facilities which can be harnessed in big data. Big data is a progressing field, and alot of research is however to be done. Currently, Hadoop is the software that handles the big data. Neverthless, the invariable growth in volumes of data is making Hadoop inadequate. In future, extensive research needs to be carried out and also revolutionary technologies need to be developed totackle the potential of big data completely.

References

1. Adams, M.N.: Perspectives on Data Mining. International Journal ofMarket Research 52(1), 11–19 (2010)

2. Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In:ACM International Conference on Web Intelligence and Intelligent AgentTechnology, vol. 1, pp. 492–499

3. (2010)

4. Bakshi, K.: Considerations for Big Data: Architecture and Approaches.In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)

5. Sreedhar C, Dr. D. Kavitha, K. Asha Rani, Big Data and Hodoop, International Journal of Advanced Research in Computer Engineering &Technology (IJARCET) Volume 3 Issue 5, May 2014 .

6. Ms. VibhavariChavan, Prof. Rajesh. N. Phursule, JSPMs Imperial College of Engineering and Research, Pune, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7932-7939.

7. Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data:The Big Data Revolution! In: Proceedings of the ACM International Workshop on DataWarehousing and OLAP, pp. 101–104 (2011)

8. Economist Intelligence Unit: The Deciding Factor: Big Data & DecisionMaking. In:Capgemini Reports, pp. 1–24 (2012)

9. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp.1–44 (2012)

10. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDBEndowment 2(2), 1481–1492, (2009)