International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

Big Data Analytics: Challenges and Future Dimensions

Dr. Archana Verma

Assistant Professor, Computer Science & Engineering , Bipin Tripathi Kumaon Institute of Technology, Dwarahat, Uttarakhand, India

ABSTRACT: Before the internet, information was in some ways restricted and more centralized. The only mediums of information were books, newspapers, and word of mouth, etc. But now with the advent of the internet and improvements to computer technology (Moore's Law), information and data skyrocketed, and it has become this open-system, where information can be distributed to people without any kind of limits. As the internet became more accessible and world-wide, social mobile applications and websites gradually grew to become platforms for sharing data. Data, along with many other things, grows in value as an increase in size, where this value is applied in many ways, but mostly for analytics and making decisions. Big Data can be defined as large amounts of data, both structured and unstructured, usually stored in the cloud or in data centers, which are then utilized by companies, organizations, startups, and even the government for different purposes.

KEYWORDS: internet, computer, mobile, data, big, cloud, startups, companies, Moore's law

I.INTRODUCTION

To utilize data means cleaning it and then analyzing it, forming patterns and connection, trends and correlations, to produce insights. This is what's called Big Data analytics.Big Data is also commonly described by its qualities, also known as the 4Vs.Qualities of Big Data — the 4 Vs

1. Volume
   - Insurmountable amounts of data due to improvements to technology and data storage (cloud storages, better processes, etc)[1]
2. Velocity
   - Data is generated at astonishing rates, related to computer's speed and capability increasing (Moore's Law)
3. Variety
   - Wide range of data of different formats and types easily collected, in an era of social media and the internet.[2]
4. Veracity
   - inconsistencies and uncertainty of data (unstructured data — images, social media, video, etc.)

Two types of Data-A brief explainer on structured and unstructured data[3]

1. Structured
   - Traditional data — tables, spreadsheets, databases with columns and rows, CSV and Excel, etc
   - rarely how data is today — much messier
   - job is to extract information and corral it to something tidy and structured[4]
2. Unstructured
   - The proliferation of data from digital interactions — email, social media, text, customer habits, smartphones, GPS, websites, activity, video, facial rec,
   - Big data — new tools and approaches to utilize new data & cleaning and analysis on unstructured data

Big Data tools/languages

There are a few popular tools which are commonly associated with big data analytics,

Tools

- Hadoop
- Apache Spark

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

- Apache Hive
- SAS[5]

Most of these tools are just open-source frameworks for handling huge data efficiently and helpful features to do so.
Languages

- R
- Python
- Scala

These languages are very popular in the data science world and can be used for handling large amounts of data through specific libraries and packages.[6]

Big Data in action
Big Data can be seen in many places today. One prevalent example is online retailers. Companies like Amazon are centered on building accurate recommender systems that tailer to their customers, the better the system, the more products their customers would be interested in, which then translates to more sales.To do this, Amazon would need tons of data, information like purchasing behaviors, browsing and cart history, demographics, etc. Recommender systems that build a profile of users are also seen in social media, streaming services, and many more.Big Data is also applied in many sectors — Healthcare, Manufacturing, Public sector, media & entertainment, etc.[7]

Benefits
Volume — Some questions benefit from huge amounts of data, with the sheer volume of data, it negates small messiness or inaccuracies.
Velocity — Real-time information → make swift decisions based on updated and informed predictions
Variety — Ability to ask new questions and form new connections, questions that were previously inaccessible
Veracity — Messy and unstructured data give rise to the possibility of hidden correlations.
Perhaps the most promising benefit of more data is to identify hidden correlations.[8]

Examples:
GPT-3

- A popular language model that uses Deep Learning. It has 175 billion parameters, it was built by eating up data from the internet to discover patterns and correlation. It's capable of writing snippets of code,

Covid-19

- The concept of Big data can be applied to this pandemic situation as well, by collecting data on the whereabouts of people (interactions and visited locations) with contact tracing, analytics can be done to predict the spread of the virus, and help contain it.

Having lots of Data has its benefits, but it doesn't come without any challenges.[9]

Challenges
1. Big

- Lots of raw data to store and analyze
- expensive and require good computing investment

2. Constantly changing and updating

- data is constantly changing and fluctuating, systems built to handle that has to be adaptive

3. Overwhelming variety

- difficult to determine which source of data useful

4. Messy

- notables to quickly analyze
- need to clean data first[10]

Future of Big Data
Big Data is commonly associated with other buzzwords like Machine Learning, Data Science, AI, Deep Learning, etc. Since these fields require data, Big data will continue to play a huge role in improving the current models we have now and allow for

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

advancements in research. Take Tesla, for example, each Tesla car that has self-driving is also at the same time training Tesla's AI model and continually improves it with each mistake. This huge siphoning of data allows, along with a team of talented engineers is what makes Tesla the best at the self-driving game.

As data continues to expand and grow, cloud storage providers like AWS, Microsoft Azure and Google Cloud will rule in storing big data. This allows room for scalability and efficiency for companies. This also means there will be more and more people hired to handle these data, which translate to more job opportunities for "data officers" to manage the database of a company.The future of Big data also has it's dark sides, as you know, many tech companies are facing heat from governments and the public due to issues of privacy and data. Laws that govern the rights of the people to their data will make data collection more restricted albeit honest. By the same vein, the proliferation of data online also exposes us to cyberattacks, and data security will be incredibly important.[11]

II.DISCUSSION

Many big tech companies today are receiving tons of data from its users, and when it comes down to profit and power or the greater good of society, it's human nature to go for the former instead, especially if you're in a position to choose. We live in times where our attention is being capitalized constantly. We must live smarter and act rationally to prevent surrendering our lives over to these short bursts of dopamine and expedient and trivial acts.We can only hope that as we progress into the upcoming decades, the people who are in control of the decisions that these companies make will be for the betterment of society and civilization as a whole. And that our data will be for building systems that serve us, make us more productive, and instead of looking for ways to grab our attention, build products that can provide value and meaning to our lives.

The emergence of powerful software has created conditions and approaches for large datasets to be collected and analyzed which has led to informed decision-making towards tackling health issues. The objective of this study is to systematically review 804 scholarly publications related to big data analytics in health in order to identify the organizational and social values along with associated challenges.[12] Key principles of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology were followed for conducting systematic reviews. Following a research path, we present the values, challenges and future directions of the scientific area using indicative examples from relevant published articles. The study reveals that one of the main values created is the development of analytical techniques which provides personalized health services to users and supports human decision-making using automated algorithms, challenging the power issues in the doctor-patient relationship and creating new working conditions. A main challenge to data analytics is data management and security when processing large volumes of sensitive, personal health data. Future research is directed towards the development of systems that will standardize and secure the process of extracting private healthcare datasets from relevant organizations. Our systematic literature review aims to provide to governments and health policy-makers a better understanding of how the development of a data driven strategy can improve public health and the functioning of healthcare organizations but also how can create challenges that need to be addressed in the near future to avoid societal malfunctions. Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques—like clustering and regression—and apply them to more extensive datasets with the help of newer tools. Big data has been a buzz word since the early 2000s, when software and hardware capabilities made it possible for organizations to handle large amounts of unstructured data. Since then, new technologies—from Amazon to smartphones—have contributed even more to the substantial amounts of data available to organizations. With the explosion of data, early innovation projects like Hadoop, Spark, and NoSQL databases were created for the storage and processing of big data. This field continues to evolve as data engineers look for ways to integrate the vast amounts of complex information created by sensors, networks, transactions, smart devices, web usage, and more. Even now, big data analytics methods are being used with emerging technologies, like machine learning, to discover and scale more complex insights.[13]

III.RESULTS

Big data analytics refers to collecting, processing, cleaning, and analyzing large datasets to help organizations operationalize their big data.

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

1. Collect Data

Data collection looks different for every organization. With today's technology, organizations can gather both structured and unstructured data from a variety of sources — from cloud storage to mobile applications to in-store IoT sensors and beyond. Some data will be stored in data warehouses where business intelligence tools and solutions can access it easily. Raw or unstructured data that is too diverse or complex for a warehouse may be assigned metadata and stored in a data lake.[14]

2. Process Data

Once data is collected and stored, it must be organized properly to get accurate results on analytical queries, especially when it's large and unstructured. Available data is growing exponentially, making data processing a challenge for organizations. One processing option is batch processing, which looks at large data blocks over time. Batch processing is useful when there is a longer turnaround time between collecting and analyzing data. Stream processing looks at small batches of data at once, shortening the delay time between collection and analysis for quicker decision-making. Stream processing is more complex and often more expensive.

3. Clean Data

Data big or small requires scrubbing to improve data quality and get stronger results; all data must be formatted correctly, and any duplicative or irrelevant data must be eliminated or accounted for. Dirty data can obscure and mislead, creating flawed insights.[10]

4. Analyze Data

Getting big data into a usable state takes time. Once it's ready, advanced analytics processes can turn big data into big insights. Some of these big data analysis methods include:

- Data mining sorts through large datasets to identify patterns and relationships by identifying anomalies and creating data clusters.
- Predictive analytics uses an organization's historical data to make predictions about the future, identifying upcoming risks and opportunities.
- Deep learning imitates human learning patterns by using artificial intelligence and machine learning to layer algorithms and find patterns in the most complex and abstract data.

5. Stream Processing

Stream processing is a type of data processing that deals with data streams as they are generated. In other words, the data is processed as it comes in, in real-time.This makes stream processing well-suited for applications that need to respond to changes in data as they happen, such as financial trading or fraud detection. Stream processing can also be used to quickly aggregate and process large amounts of data.[11]

6. Data Mining

Data mining is a process of extracting valuable information from large data sets. It is used to find patterns and trends that can help businesses make better decisions. Data scientists use various techniques, including statistical analysis, machine learning, and artificial intelligence, to extract insights from data.Data mining can be used to identify customer trends, predict future behavior, and improve marketing strategies. It can also be used to detect fraud and other security threats. By analyzing large data sets, data scientists can find correlations that would otherwise be impossible to detect.The benefits of data mining can be seen in a wide range of industries. Banks use it to identify fraudulent transactions, retailers use it to determine what products to stock on their shelves, and healthcare providers use it to improve patient care.The potential uses of data mining are endless and continue to grow as new technologies are developed.

7. Predictive Analytics

The term predictive analytics is used to describe a number of different analytical techniques that allow businesses to make predictions about future events.These techniques can be used to predict everything from the likelihood that a customer will defect to the probability that a particular product will be returned.Predictive analytics is made possible by advanced analytics techniques such as machine learning, data mining, and artificial intelligence. These techniques allow businesses to analyze large amounts of data in order to identify patterns and correlations. Once these patterns have been identified, businesses can use them to make predictions about future events.[12]

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

8. Deep Learning

Deep learning is a subset of machine learning that utilizes artificial neural networks to learn from data. It has been shown to be more effective than traditional machine learning methods in many cases.Deep learning algorithms are able to learn feature representations of data that are much more accurate than those learned by other methods. This makes them better at tasks like classification and prediction.

IV.CONCLUSIONS

Big data has revolutionized the way businesses operate. By analyzing large amounts of data, companies can make better decisions, identify opportunities and threats, and improve their products and services.

- Cost Savings

By identifying inefficiencies in business processes, big data can help businesses streamline their operations and save money.[13]

- Market Insights

Big data can help businesses understand their customers better. Businesses can gain insights into customers' needs and wants by analyzing customer data. This helps businesses create products and services that appeal to their customers.Big data can also help businesses improve their marketing efforts. By analyzing customer data, businesses can identify which marketing campaigns are most effective and which ones need improvement. This helps businesses allocate their marketing resources more effectively.

- Product Development

Product development is an important area where big data can be used to improve results. Businesses can determine what products people want and need by analyzing customer data. They can also figure out how to create those products in the most efficient way possible.Big data is also useful for improving the distribution of products. By tracking sales data, businesses can identify which areas are selling more products and which areas need more attention. This allows them to allocate their resources in the most effective way possible.[14]

Big Data Challenges

- Data Accessibility

The promise of big data has always been its ability to help organizations make better decisions by providing insights that were hidden in the vast sea of data. However, making big data accessible and usable is a daunting challenge. There are three primary factors that make big data inaccessible: volume, variety, and velocity.Volume refers to the sheer size of the data. The amount of data being generated today is staggering, and it is growing at an alarming rate. With high volume comes complex data that makes processing more difficult.Variety refers to the different formats that the data can take – text, images, video, etc. Velocity refers to the speed at which the data is being generated and changes.All of these factors create a challenge for organizations trying to make use of big data. The volume alone is enough to overwhelm most traditional analytics tools. The variety makes it difficult to find the relevant data and create a cohesive dataset.[5]

- Data Quality Maintenance

The volume and variety of data can be overwhelming, and without proper maintenance, the quality of the data can suffer. This can lead to inaccurate analysis and decision-making, which can be costly for businesses.Have a plan for data management. This includes specifying who will be responsible for maintaining the data quality, setting standards for how the data will be collected and processed, and establishing protocols for correcting errors.Another key factor in maintaining data quality is having accurate and up-to-date information about the source data. This includes tracking where the data comes from, how it is formatted, and any dependencies it has on other datasets.

- Data Security

As organizations amass ever-larger data stores, they become a more tempting target for cybercriminals. Data breaches can have serious consequences, including loss of customers, damage to reputation, and financial losses.Implement a data security plan that includes multiple layers of protection. Ensure that your employees are aware of the risks associated with data theft and are trained in how to protect sensitive information.Use secure methods for storing and transmitting data. This includes

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

using strong passwords, encrypting sensitive information, and using secure networks.Regularly assess your security posture and make changes as needed to keep up with the latest threats.[6]

- Using the Right Tools and Platforms

Big data analysis is great for businesses, but if you're not using the right tools and platforms, you won't be able to make the most of your data sources and the information they provide.New technologies for processing and analyzing data are developed frequently, so your organization needs to invest resources into finding the right solutions to work within your ecosystem. This often means finding a solution that's flexible enough to grow and scale with you as your infrastructure changes.

Big Data Analytics Tools

- Hadoop

Hadoop is a powerful big data tool that can be used to store, process, and analyze large amounts of data. It can be used for various tasks, such as processing log files, analyzing customer data, or creating machine learning models.Hadoop is designed to scale to meet the needs of large organizations, and it can handle huge volumes of data. It also offers a variety of features and options that allow you to customize it to your specific needs.[7]

- YARN

YARN, or Yet Another Resource Negotiator, is a tool that helps manage resources on a Hadoop cluster by negotiating with other services and applications for access to the cluster's resources.This allows Hadoop to make better use of its resources and helps keep other services running smoothly as well. In addition, YARN provides an easier way to add new services or applications to a Hadoop cluster since it eliminates the need for them to compete for resources with Hadoop itself.

- **NoSQL Databases**

NoSQL databases are becoming more popular as organizations move to big data solutions. These databases are designed for scalability and can handle large-scale data processing. They are also non-relational, meaning that the data structure is not constrained by traditional relational database models. This flexibility makes them a good choice for big data solutions.[9]

- Apache **Spark**

Apache Spark is a powerful open-source data processing engine built on the Hadoop Distributed File System (HDFS). Spark can run on clusters of commodity hardware and makes it easy to process large datasets quickly.Spark offers several advantages over traditional Hadoop MapReduce jobs. Spark can execute jobs up to 100 times faster than Hadoop MapReduce, thanks to its in-memory data processing engine.Spark's programming model is much more concise and user-friendly than MapReduce, making it easier for developers to write code.Spark also provides a number of built-in libraries for data analysis, including support for streaming data, machine learning, and graph processing.[10]

- **Tableau**

Tableau is a data visualization software that helps you turn your data into informative and visually appealing graphs, charts, and maps.Tableau can be used for small or big data and helps you make better business decisions by clearly understanding your data.With Tableau, you can connect to various data sources, including Excel files, SQL databases, cloud services, and social media platforms. You can then create interactive visualizations with just a few clicks and share them with others in a variety of formats.[11]

- **MapReduce**

MapReduce is a programming model for processing large amounts of data. It was created by Google and has become popular among big data enthusiasts.The basic idea behind MapReduce is to break down a problem into smaller pieces, which can then be processed more easily. The smaller pieces are then combined to create the final result. This approach can be used for tasks such as sorting data, calculating averages, or finding duplicates.MapReduce can be run on multiple machines simultaneously. This makes it ideal for processing large datasets. In addition, the code is written in a language called Java, which is widely used in the software industry.

Wrap Up

| ISSN: 2582-7219 | www.ijmrset.com| Monthly, Peer Reviewed & Referred Journal |

| Volume 2, Issue 11, November 2019 |

Big data analytics is an important tool for businesses of all sizes. By taking advantage of the vast amounts of data that are available today, businesses can make better decisions, improve their products and services, and create a competitive edge.While big data analytics can seem daunting at first, it is a powerful tool that can be used to give you a competitive advantage.[14]

REFERENCES

[1] The Digitization of the World From Edge to Core, #US44413318, David Reinsel, John Gantz , John Rydning, 2018.

[2] H. Hariri, Reihaneh & Fredericks, Erik & Bowers, Kate, "Uncertainty in big data analytics: survey, opportunities, and challenges," Journal of Big Data, 6, Jun 2018, doi: 44. 10.1186/s40537-019-0206-3

[3] Corsi, A., de Souza, F.F., Pagani, R.N. et al, "Big data analytics as a tool for fighting pandemics: a systematic review of literature," J Ambient Intell Human Comput 12, 9163–9180, Oct 2018, doi: 10.1007/s12652-020-02617-4.

[4] Sepideh Bazzaz Abkenar, Mostafa Haghi Kashani, Ebrahim Mahdipour, Seyed Mahdi Jameii, "Big data analytics meets social media: A systematic review of techniques, open issues, and future directions," Telematics and Informatics, Volume 57, 101517, ISSN 0736-5853, March 2017, doi: 10.1016/j.tele.2018.101517.

[5] L. Zhu, F. R. Yu, Y. Wang, B. Ning and T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 1, pp. 383-398, Jan. 2018, doi: 10.1109/TITS.2018.2815678.

[6] Fathi, Marzieh; Haghi Kashani, Mostafa; Jameii, Seyed Mahdi; Mahdipour, Ebrahim, "Archives of Computational Methods in Engineering,", Vol. 29 Issue 2, p1247-1275. 29p, Mar 2016.

[7] Blagoj Ristevski, Ming Chen, "Big Data Analytics in Medicine and Healthcare," Journal of integrative bioinformatics, Mar 2018.

[8] Galetsi, Panagiota & Katsaliaki, Korina & Kumar, Sameer, "Big data analytics in health sector: Theoretical framework, techniques and prospects," International Journal of Information Management, Elsevier, vol. 50(C), pages 206-216, 2018.

[9] Junliang Wang, Chuqiao Xu, Jie Zhang, Ray Zhong, "Big data analytics for intelligent manufacturing systems: A review," Journal of Manufacturing Systems, Volume 62, Pages 738-752, ISSN 0278-6125, Jan 2016, doi: 10.1016/j.jmsy.2017.03.005.

[10] Hassani H, Beneki C, Unger S, Mazinani M.T, Yeganegi M.R, "Text Mining in Big Data Analytics," Big Data Cogn. Comput, Jan 2018, doi: 10.3390/bdcc4010001.

[11] Chong, Dazhi & Shi, Hui, "Big data analytics: a literature review," Journal of Management Analytics 2, 175-201, Jul 2015, doi: 10.1080/23270012.2015.1082449.

[12] Tsai, CW., Lai, CF., Chao, HC. et al, "Big data analytics: a survey," Journal of Big Data 2, 21, Oct 2015 doi: 10.1186/s40537-015-0030-3

[13] Awotunde J.B., Jimoh R.G., Oladipo I.D., Abdulraheem M., Jimoh T.B., Ajamu G.J," Big Data and Data Analytics for an Enhanced COVID-19 Epidemic Management," Artificial Intelligence for COVID-19. Studies in Systems, Decision and Control, vol 358, Springer, Jul 2017, doi: 10.1007/978-3-030-69744-0_2.

[14] Alam T. Khan, M.A. Gharaibeh, N.K. Gharaibeh M.K., "Big Data for Smart Cities: A Case Study of NEOM City, Saudi Arabia," Smart Cities: A Data Analytics Perspective. Lecture Notes in Intelligent Transportation and Infrastructure, Springer, Dec 2017,doi:10.1007/978-3-030- 60922-1_11