Implications of AI and Big Data

Sukhdeep Kaur[1], Sukhdeep Kaur[2]
[1]*Assistant Professor, Department of Computer Science and Engineering Chandigarh Engineering College-CGC, Landran*
[2]*Assistant Professor, Department of Computer Science and Engineering College of Engineering-CGC, Landran*

Abstract-The term "Big Data" describes data that is too big to handle with typical applications because of the difficulties involved in gathering, storing, transporting, accessing, processing quickly, and updating such vast volumes of data. Deep learning, machine learning, and artificial intelligence (AI) analytics are frequently used in conjunction with the big data notion. The impact of big data on the use of AI concepts and approaches is examined in this study.

1. INTRODUCTION

Three topics are covered by a big data system:
Three V's define the dimensions: Volume (the amount of data), Variety (the amount of data in various formats or from many sources) & Velocity (the rate at which information is produced and must be accessible for processing). These days, the 5 V's—Veracity (reliability of the data sources) and Value (usefulness)—are the most important features of big data since they focus on the quality of the collected data rather than its technological aspects . Since the term "big data" alludes to the quick and easy processing and storing of vast volumes of data, analytics involving AI, machine learning, and deep learning are frequently used in this context. Although these names are frequently used synonymously, they are not the same thing .

The goal of artificial intelligence (AI), a subfield of computer science, is to build intelligent machines that can accomplish tasks that humans can, such as learning, comprehending difficult ideas, drawing conclusions, or conversing with people. However, machine learning, which is a subfield of artificial intelligence, refers to the ability of machines to access data and learn on their own .

Deep Learning is a subset of machine learning techniques that, as opposed to task-specific algorithms, rely on learning data representations with an emphasis on neural networks. Artificial Neural Networks (ANNs), often known as Neural Networks, are computer systems that draw inspiration from biological networks.The neurons and nerves that make up the human brain. Due to this technique, computers are now able to learn from their experiences and comprehend the world through a conceptual hierarchy. Deep Learning works well for recognising words from audio and things from photos. In the future, researchers hope to apply this idea to a wide range of more complicated jobs, including marketing, medical diagnosis, automatic language translation, and many other significant social and commercial issues. Big data analytics is the frequently difficult process of looking through vast and diverse sets of data, or "Big data," to find information that might assist businesses in making wise decisions. This information might include hidden patterns, unidentified correlations, market trends, and client preferences.

Big data analytics makes information that might have otherwise remained hidden visible, assisting businesses and organizations in making better decisions. It can be challenging to derive meaningful insights from huge data regarding the trends, correlations, and patterns that exist without a lot of processing capacity. The statistical algorithms and prediction models used in big data data visualization are more sophisticated than those used in simple business intelligence inquiries. As compared to standard business intelligence techniques, the answers are almost instantaneous. With the development of social media, artificial intelligence, and the Internet of Things, which has a vast array of sensors and gadgets, big data is only going to get greater. The "3Vs" of

variety, volume, and velocity are used to measure data. More of it than ever before is present, frequently in real time. If this deluge of data cannot be questioned, it is useless and worthless.

However, the big data analytics model analyzes text, statistics, and language using machine learning to uncover previously undiscovered insights. It is possible to mine every data source for insights and value. The range of unstructured data that may be mined for information, including texts, social media posts, clickstream data, and the abundance of sensors from the Internet of Things, has expanded the significance of big data analytics. The torrent of unstructured data that characterizes today's world is too much for relational databases and warehouses to handle. They work best with data that is structured. Additionally, they are unable to handle real-time data demands. The increasing need for real-time understanding of unstructured data is met by big data analytics. This is especially crucial for businesses that depend on the volume of website or mobile activity and the swiftly changing financial markets.

2. LITERATURE SURVEY

2.1 Definitions of Big Data

The four characteristics of big data, which are large data scale, fast data flow, diversified data types, and low-value density, were summed up by McKinsey as the "4V characteristics." IBM

The definition of big data, which is the 5V features of big data that are very prevalent in the industry, was later developed by adding the fifth characteristics.

- Volume: the quantity of data created and kept.
- Velocity: describes the rate at which data is produced and analyzed in order to satisfy the obstacles and requirements that arise during the course of expansion and advancement.
- Variety: as a result of data type and nature.
- Veracity: the acquired data's data quality can differ significantly, which impacts the precise analysis.
- Variability: pertains to data that exhibits changes in value or other attributes based on the context in which they are produced.

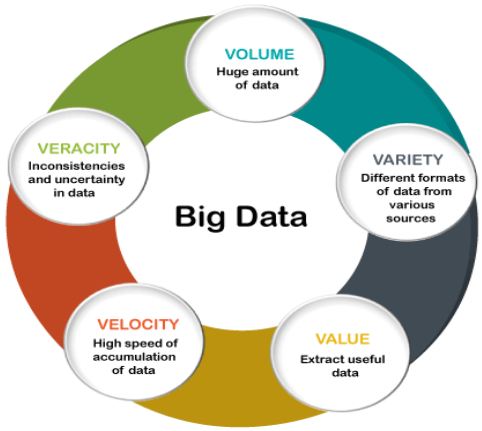

Fig 1: Characterstics of Big Data

2.2 Big Data and AI

Two of the most important domains of computer science nowadays are big data and computational intelligence (AI). Research on artificial intelligence and big data has been ongoing in recent years. Artificial intelligence and big data are inseparable: (1) Big data technology is dependent on artificial intelligence because it employs numerous ideas and techniques related to artificial intelligence; (2) Big data technology is also necessary for the development of artificial intelligence since it needs an enormous amount of data in order to work. Every data input is used by AI, machine learning, and deep learning to create new rules for business analytics in the future.

Five ways AI enhances understanding:

- AI is developing new techniques for data analysis: "Machine learning and AI are the result of the convergence of statistical models and computer science."
- The labor-intensive nature of data analytics is decreasing: This means that managing and analyzing data no longer requires as much time-consuming manual labor as it did in the past.While humans are still essential to data management and analytics, AI is speeding up procedures that previously could have taken days, weeks, or even longer.
- Humans are still very important: "Companies that have effectively integrated human and technological capabilities have been able to broaden the pool of individuals who can access critical analytics insights beyond data scientists and business analysts. This has allowed them to

save time and minimize the possibility of bias arising from business users' interpretation of data. This leads to faster data-driven insights, more effective business operations, and ultimately higher enterprise productivity.

- AI/ML can be utilized to address standard data difficulties: “ML algorithms can find duplicate records that describe the same entity with slightly different terminology, detect outlier values and missing values, normalize data to a common terminology, etc.”
- Prescriptive and predictive analytics are becoming more common: "Big data decisions were traditionally based on historical and current data points, which typically resulted in linear ROI." This has gotten to epic and exponential levels with AI. By using artificial intelligence, prescriptive analytics could deliver forward-looking strategic insights the fact that will assist a business prosper.

3. METHODOLOGY

When it comes to identifying new revenue opportunities and enhanced efficiencies that give a competitive edge, businesses recognize the value of big data analytics. As more large companies recognize this value, they benefit from reduced costs through the discovery of more efficient ways to conduct business, faster and better decision-making through the ability to analyze information immediately and act on the learning, new products, and the ability to more fully understand customers through the use of data.

3.1 Big Data Analytics and AI

AI and big data go together hand in hand. As more data is given to AI, it gets better. It's making it practical for companies to comprehend their clients far better than they could in the past. Nevertheless, without tools to evaluate it, large data is worthless. Humans are terrible at it.

AI uses big data as illustrated :

- Finding anomalies: Artificial intelligence (AI) can examine data to find oddities or anomalies. As an illustration, consider a network of sensors with a planned for, appropriate range. An anomaly refers to anything that occurs outside of that range.
- Probability of future outcome: AI can calculate the likelihood of a future outcome given a known situation that has a known probability of influencing that outcome.
- AI is able to recognize patterns – AI is able to spot patterns that humans aren't able to.
- Data Bars and Graphs: Artificial Intelligence can search for trends in data bars and graphs that human inspection would notice.

3.2 AI methods and techniques

Big data analytics has been enhanced by artificial intelligence (AI) techniques. There are three categories of IoT (Internet of Things) data that we can choose from: (1) Raw data, which is unprocessed and unstructured; (2) Meta data, which is data about data; and (3) Transformed data, which is value-added data. In order to recognize, classify, and make decisions on each of these data kinds, artificial intelligence will be helpful. The ability to make unstructured data understandable and valuable as information for decision-making is made possible by AI in conjunction with sophisticated big data analytics. Effective and efficient decision making is going to rely on the integration of AI to IoT and data analytics, particularly in the areas of streaming data and real-time analytics connected to edge computing networks. For every use case, customer segment, and set of solutions, real-time data will be an important competitive point.

The capacity to gather real-time data, identify valuable characteristics, and make conclusions will give service logic an entirely novel form of intricacy. The service will often be the data itself and actionable insights.

3.2.1 Big Data Management

Strategic placement of massive data centers is a key component of effective Big Data management. The primary advantage of deploying redundant processing systems with physical locations in different regions is that any localized failure, such one brought on by a natural disaster, does not bring down the system as a whole. For 99.9% uptime performance, business-critical gear must have at least triple redundancy. Foundational IT structures prepare for disasters that may affect an IT network by taking disaster risk into account.

Real-time load balancing reroutes processing to the

servers in the network that are still up and running in order to handle any partial systemic failure of some of the servers.

3.2.2 Big Data Storage Management

Big Data has significant storage requirements. One method is to gather and handle the localized data before sending it to a larger cloud-maintained storage system for storage. Implementing a "virtualized" data system, which creates a virtual layer of the data, is a further approach. This virtual layer is aware of the network's data storage locations. The data required for a particular calculation is the only data obtained when implementing the AI algorithm in a virtual system. There is no need to copy data files since the original data store is still there.

This method makes use of a protocol for network-wide data management. It promotes computational processing costs and lessens the need for data storage memory.

3.2.3 Artificial Intelligence

An info graphic chart illustrating the 2019 Data & AI Landscape was created by First mark. It is indisputable that Big Data computational processing is trending toward the cloud [32].

According to Forbes, cloud-based Big Data storage and following mining of data using SaaS AI applications are rapidly supplanting local data storage. Through a significant digital transformation, businesses can get insights that fuel innovation and make better management decisions. Deep learning is responsible for the change.

Deep learning is a method of leveraging AI programming to improve machine learning capabilities. It does not require human intervention or extensive human programming for every possible scenario.

3.2.4 Internet of Things (IoT)

By 2025, around 64 billion gadgets globally are expected to be connected to the Internet of Things (IoT), based to estimates from Techjury [34]. Approximately 24 billion IoT devices are in operation today. Every device gathers data. Big Data is growing exponentially as a result of this trend.It is only beneficial to gather this much data from multiple "smart" tools if it can be actively processed and data mined.

3.2.5 Big Data and AI Strategies

DZone points out that several applications of AI for big data analytics include:

- Discovering Anomalies: AI can examine large amounts of data to find unexpected events, or anomalies, within the collection. Networks of sensors and parameters with a predetermined suitable range can use this. Any network node that is outside of the range is recognized as a possible issue that requires investigation.
- Risks of Future Outcomes: The Bayes theorem can be used by AI to assess Big Data. It is possible to calculate the probability of an event arising by considering previous circumstances that have a fixed chance of impacting what will happen in the future.
- Identifying Patterns: AI is able to analyze enormous quantities of data to find patterns that human monitoring could miss.
- Data Bars and Graphs: Artificial Intelligence has the ability to examine Big Data to recognize patterns in bars and graphs derived from the underlying data set.

4. DISCUSSIONS AND CONCLUSION

The demand for AI stays high for some time to come. AI and data are coming together to form a synergistic connection in which data is unsurmountable without AI and AI is useless without data. "By creating links between these data sets, a complex problem can be viewed in its entirety, facilitating the discovery of new AI-driven insights."

With Big Data, AI has grown into a constant, cyclical process; the two are now unbreakable.

Over time, their gather utilization will go up significantly. This megatrend is being pushed by the exponential growth of the Internet of Things (IoT), which is generating more Big Data, and the compelling value proposition of AI analytics applied to Big Data.

Smaller businesses are using AI and Big Data on a daily basis. They have access to data centers' IT hardware resources. They follow up with the AI tools that come along as cloud services to deal with the Big Data they have accumulated

REFERENCE

[1] R. Magoulas, B. Lorica, "Introduction to Big Data," Release 2.0. Sebastopol CA: O'Reilly Media (11), February 2009.
[2] J. R. Mashey, "Big Data ... and the Next Wave of InfraStress," (PDF). Slides from invited talk. Usenix, 25 April 1998.
[3] S. Lohr, "The Origins of 'Big Data': An Etymological Detective Story," The New York Times, 1 February 2013.
[4] I.T. Hashem, A. Yaqoob, and I. B. Anuar, “"big data" on cloud computing: Review and open research issues," Information Systems. vol. 47, pp. 98–115, 2015. doi:10.1016/j.is.2014.07.006.44.
[5] Laney. D. (2001), "3D data management: Controlling data volume, velocity and variety," META Group Research Note, vol. 70, No. 6.
[6] “Big data Tutorial,” 2013. http://www. slideshare.net/markogrobelnik/big-datatutorial.