IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

Role of Artificial Intelligence in Big Data Analytics

Mrs. C. Radha[1], Mr. R. Midunkumar[2], Mr. S. Muralibabu[3], Mr. V. Partheeban[4], Mr. T. Praveenkumar[5]
Associate Professor, Department of MCA[1]
Student, II MCA, Department of MCA[2,3.4.5]
Muthayammal Engineering College, Namakkal, India

Abstract: *The integration of Artificial Intelligence (AI) in data analytics to enhance efficiency and insights. We discuss how AI techniques, such as machine learning and automation, simplify the analytics process, enabling organizations to extract valuable information from data quickly and effectively. The paper highlights the practical impact of AI in various industries and emphasizes the potential for streamlined decision-making and trend prediction. Overall, the focus is on the simplicity and effectiveness of incorporating AI into data analytics workflows for improved outcomes.*

Keywords: Artificial Intelligence, Big Data Analytics, Predictive Analytics, Narrow AI

I. INTRODUCTION

Big data analytics refers to the process of examining and uncovering patterns, trends, associations, and insights within large and complex datasets, commonly known as big data. the term "big data" refers to datasets that are characterized by their volume, velocity, and variety, which traditional data processing systems may struggle to handle effectively.It involves the use of advanced technologies and analytics techniques to make sense of data that is characterized by its volume, velocity, and variety.

The big data is nothing but a data , available at heterogeneous, autonomous sources, in extreme large amount, which get updated in fractions of seconds.big data analytics is a form of advanced analytics, which involve complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by analytics systems. Today, there are millions of data sources that generate data at a very rapid rate. these data sources are present across the world. some of the largest sources of data are social media platforms and networks. let's use facebook as an example—it generates more than 500 terabytes of data every day. this data includes pictures, videos, messages, and more.

II. HISTORY OF BIG DATA ANALYTICS

The history of Big Data analytics can be traced back to the early days of computing, when organizations first began using computers to store and analyze large amounts of data. However, it was not until the late 1990s and early 2000s that Big Data analytics really began to take off, as organizations increasingly turned to computers to help them make sense of the rapidly growing volumes of data being generated by their businesses. Most enterprises now understand the significant value of applying analytics to captured data that streams into their businesses. Today, there are new benefits that big data analytics brings to the table; they are speed and efficiency.

III. IMPORTANCE OF BIG DATA ANALYTICS

Big data analytics is important because it helps companies leverage their data to identify opportunities for improvement and optimisation. Across different business segments, increasing efficiency leads to overall more intelligent operations, higher profits, and satisfied customers. Big data analytics helps companies reduce costs and develop better, customer-centric products and services.

A. Benefits of big data analytics

Analysis of large volumes of data from disparate sources in a variety of forms and kinds in a timely manner.

Quickly making well-informed judgments for successful strategizing to enhance the supply chain, logistics, and other tactical decision-making sectors.

Savings due to the increased efficiency and optimization of business processes.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

More informed risk management techniques based on large data sample sizes
Greater knowledge of consumer behavior, demands, and sentiment can result in better product development data and strategic management processes.

B. Characteristics of Big Data Analytics

1. Volume

Big data volume is greater than the volume of processed data in a normal system of an enterprise. This results in newly-designed systems. The reason for such volumes of data varies with developments.

2. Veracity

The ingestion and processed data of different systems result in veracity challenges regarding data accuracy. For example, if different records show the same data with different dates and timestamps, it is hard to determine which record is the correct one.

3. Velocity

As the business models of enterprises are depending on IoT data more and more, IoT data is continuously resulting in the increasing speed of data generation. Data generation is not static records in a database solely; a continuous stream of data generation is necessary.

4. Variety

Along with the different source systems, the data that was not logged and overridden before can be stored in big data scenarios. The data is like record updates, and history changes, and can allow for new use cases, such as time-series analytics, that are otherwise impossible on old override data.

IV. TYPES OF BIG DATA ANALYTICS

A. Descriptive analytics

Descriptive analytics refers to data that can be easily read and interpreted. This data helps create reports and visualise information that can detail company profits and sales. Example: During the pandemic, a leading pharmaceutical company conducted data analysis on its offices and research labs. Descriptive analytics helped them identify consolidated unutilised spaces and departments, saving the company millions of pounds.

B. Diagnostics analytics

Diagnostics analytics helps companies understand why a problem occurred. Big data technologies and tools allow users to mine and recover data that helps dissect an issue and prevent it from happening in the future. Example: An online retailer's sales have decreased even though customers continue to add items to their shopping carts. Diagnostics analytics helped to understand that the payment page was not working correctly for a few weeks.

C. Predictive analytics

Predictive analytics looks at past and present data to make predictions. With artificial intelligence (AI), machine learning, and data mining, users can analyse the data to predict market trends.Example: In the manufacturing sector, companies can use algorithms based on historical data to predict if or when a piece of equipment will malfunction or break down.

D. Prescriptive analytics

Prescriptive analytics solves a problem, relying on AI and machine learning to gather and use data for risk management. Example: Within the energy sector, utility companies, gas producers, and pipeline owners identify factors that affect the price of oil and gas to hedge risks.

V.ARTIFICIAL INTELLIGENCE (AI)

AI, or Artificial Intelligence, refers to the development of computer systems that can perform tasks that typically require human intelligence. This includes tasks such as visual perception, speech recognition, decision-making, and language translation. AI is the backbone of innovation in modern computing, unlocking value for individuals and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

businesses. For example, optical character recognition (OCR) uses AI to extract text and data from images and documents, turns unstructured content into business-ready structured data, and unlocks valuable insights.

A. Types of Artificial Intelligence

1. Narrow AI (Weak AI):

This type of AI is designed and trained for a particular task. It excels in performing specific functions but lacks the broad cognitive abilities of a human. Examples of narrow AI include virtual personal assistants (like Siri or Alexa), image recognition software, and recommendation algorithms**.**

2. General AI (Strong AI):

This is a theoretical form of AI that possesses the ability to understand, learn, and apply knowledge across diverse domains, much like a human being.

VI. AI IN BIG DATA ANALYTICS

AI for big data represents a cutting-edge approach to data analytics, providing a means to empower decision-makers within the business with a level of insight previously impossible to deliver. Programs that allow computers to simulate 'learning' through algorithmic and statistical data assessment to identify patterns and draw inferences from these patterns. Artificial intelligence is a different subject from Big data, it is not a segment of big data, yet it operates on data. Data is the fuel for artificial intelligence. Artificial intelligence is the technology that was designed to mimic human intelligence. The role of artificial intelligence in big data is that artificial intelligence facilitates stages of the big data workflow. The steps include aggregating, storing, and retrieving varied data types from disparate sources. Artificial intelligence plays a crucial role in detecting trends or patterns in the data. This is attributed to the AI's ability to learn from the data. This helps in use cases such as providing feedback to customers. Artificial intelligence in big data is also significant. With artificial intelligence, such a huge amount of data would be valuable. Unless it can use the data and drive insights out of the data by molding it into intelligence.

A. Automated Data Processing:

Data Cleaning and Preprocessing: AI algorithms can automatically clean and preprocess large datasets, handling tasks such as missing data imputation, outlier detection, and normalization. This ensures that the data used for analysis is accurate and reliable.

B. Pattern Recognition and Insights:

Identifying Patterns and Trends: AI, especially machine learning models, excels at recognizing patterns and trends within vast datasets. This capability is crucial for extracting meaningful insights and identifying correlations that may not be apparent through traditional analytics approaches.

C. Predictive Analytics:

Forecasting and Predictions: AI models, particularly those based on machine learning, enable predictive analytics by learning from historical data and making predictions about future trends. This is valuable for businesses in planning and decision-making processes.

D. Real-time Analytics:

Streaming Data Analysis: AI facilitates the analysis of streaming data in real-time. This is essential for applications like monitoring, fraud detection, and other scenarios where immediate insights are required for timely decision-making.

E. Natural Language Processing (NLP):

Text and Sentiment Analysis: AI-driven NLP techniques allow organizations to analyze and understand unstructured data, such as text. This is beneficial for extracting insights from sources like customer feedback, social media, and textual documents.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

VII. ACCESSIBILITY OF BIG DATA ANALYTICS USING AI

Data and artificial intelligence are closely intertwined and dependent on each other. AI and machine learning help address common data issues, including data quality and value. High-quality data is crucial as it loses its importance and meaning when its quality is low. Data scientists spend about 80% of their time cleaning and preparing the data. So, artificial intelligence acts as the litmus test for the data. The integration of AI with Big Data extends the capabilities of traditional analytics methods. AI technologies enhance the efficiency, accuracy, and depth of data analysis, enabling organizations to derive actionable insights, make informed decisions, and gain a competitive edge in today's data-driven landscape. The symbiotic relationship between AI and Big Data is essential for unlocking the full potential of data analytics.

Big data and AI have become much more accessible for organizations of all sizes and types. Data science solutions are now accessible to organizations that cannot afford to hire their own experts but still want to unlock their business's full potential. There are companies providing these technologies in user-friendly formats for specific functions, such as understanding customer behavior or recruiting. Big data service companies take the guesswork out of setting up artificial intelligence and Big data, ensuring that there are no missed steps.AI into Big Data analytics enhances accessibility by automating complex tasks, providing user-friendly interfaces, enabling predictive analytics for non-experts, leveraging natural language processing, offering cloud-based solutions, providing pre-built models, and supporting low-code/no-code development. These advancements democratize access to Big Data analytics, empowering a wider range of users to harness the benefits of data-driven insights.

VIII. AI TOOLS USED IN BIG DATA ANALYSIS

A. Tensor Flow:

It is Machine Learning Library Developed by Google, Tensor Flow is an open-source library used for building and training machine learning models. It is widely employed for tasks such as deep learning, neural networks, and natural language processing.

B. PyTorch:

Its also a Machine Learning Library. PyTorch is an open-source machine learning library primarily developed by Facebook. It is known for its flexibility and dynamic computation graph, making it suitable for various machine learning tasks, including deep learning.

C. Apache Spark MLlib:

It is the type of Distributed Machine Learning Library Part of the Apache Spark framework, MLlib provides scalable machine learning capabilities for big datasets. It supports various algorithms for classification, regression, clustering, and collaborative filtering.

D. Scikit-learn:

It is Machine Learning Library. Scikit-learn is a popular Python library for machine learning and data mining. While not specifically designed for Big Data, it is often used in combination with distributed computing frameworks like Apache Spark for scalable machine learning.

E. H2O.ai:

It is Automated Machine Learning Platform.H2O.ai offers a suite of AI and machine learning products, including H2O-3, an open-source platform that supports distributed and scalable machine learning. It is designed to work seamlessly with big datasets.

F. IBM Watson Studio:

Watson Studio is an integrated environment by IBM that provides tools for data scientists, developers, and business analysts. It supports machine learning, deep learning, and data analytics, facilitating collaboration and experimentation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

G. Data bricks:
It is one of Unified Analytics Platform Data bricks provides a unified platform built on top of Apache Spark. It enables collaborative and interactive Big Data analytics, combining data engineering, machine learning, and business intelligence capabilities.

H. Amazon Sage Maker:
Amazon own Machine Learning Service. Part of Amazon Web Services (AWS), Sage Maker is a fully managed service that enables the building, training, and deployment of machine learning models at scale. It integrates with various AWS tools for seamless Big Data processing.

I. Microsoft Azure Machine Learning:
Microsoft has own Machine Learning Service. Azure Machine Learning is a cloud-based service by Microsoft that facilitates the creation, training, and deployment of machine learning models. It integrates with Azure's Big Data and analytics services.

J. Rapid Miner:
Its a Data Science Platform. Rapid Miner is a data science platform that provides an integrated environment for data preparation, machine learning, and predictive analytics. It supports both traditional and big data analytics.

IX. BIG DATA AND AI STRATEGIES

Data Zone notes that some ways that AI is applied to Big Data Analytics include :

A. Detecting Anomalies:
AI can analyze Big Data to detect anomalies (unusual occurrences) in the data set. This can be applied to networks of sensors and parameters that have a predefined appropriate range. Any node of the network that is outside of the range is identified as a potential problem that needs attention.

B. Probabilities of Future Outcomes:
AI can analyze Big Data using Bayes theorem. The likelihood of an event occurring can be determined using known conditions that have a certain probability of influencing the future outcome.

C. Recognizing Patterns:
AI can analyze Big Data to look for patterns that might otherwise remain undetected by human supervision.

D. Data Bars and Graphs:
AI can analyze Big Data to look for patterns in bars and graphs that are made from the underlying data set.

X. CONCLUSION

Artificial Intelligence and Big Data have a deep-rooted relationship that is inseparable. Both of these go hand-in-hand, and together these tools have disrupted how the world and we operate daily. Big data empowered with AI capabilities are shaping the sectors' future across industries and providing businesses with valuable data to make informed decisions. Big data and AI have enormous potential to realize highly effective learning and teaching. They stimulate new research questions and designs, exploit innovative technologies and tools in data collection and analysis, and ultimately become a mainstream research paradigm. The adoption of big data and AI in learning and teaching is still in its infancy and limited by technological and mindset challenges for now; however, the convergence of developments in psychology, data science, and computer science shows great promise in revolutionizing educational research, practice, and industry. We hope that the latest achievements and future directions presented in this paper will advance our shared goal of helping learners and teachers pursue sustainable development.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, April 2024

REFERENCES

[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].Grimes, "Big Data: Avoid 'Wanna V' Confusion," InformationWeek. Retrieved 5 January 2016.
[5]. R. Shankarmani, M. Vijayalakshmi, "Big Data Analytics," 2ed Paperback – 1 January 2016, 2nd edition, Wiley.
[6]. Q.McKinsey, "Are you ready for the era of 'big data'?," October 1, 2011.
[7]. "IBM What is big data? – Bringing big data to the enterprise,". ibm.com. Retrieved 26 August 2013.
[8]. Anwar, S., Bascou, N. A., Menekse, M., and Kardgar, A. (2019). A systematic review of studies on educational robotics.