

Managing Big Data Databases using AI– Review

Deepika Rooprai, Srushti Raut and Annukumari Sharma

Students, Department of MCA

Late Bahusaheb Hiray College S. S. Trust's Institute of Computer Application, Mumbai, Maharashtra, India

Abstract: *This research paper examines how Big Data affects the use of AI methods and techniques. Big Data refers to the challenge of handling large amounts of data that can't be effectively processed by traditional applications. It involves capturing, storing, transferring, querying, and processing such data quickly. The paper explores how AI, Machine Learning, and Deep Learning are used for analyzing Big Data. It discusses the impact and potential benefits of combining Big Data and AI. The research highlights the evolving field of Big Data and AI, which has the potential to revolutionize various industries. The paper emphasizes the significance of AI in supporting different business models and the global impact of Big Data research.*

Keywords: AI, Big Data, Analytics, Storage

I. INTRODUCTION

In today's data-driven world, managing large databases of information, known as Big Data, has become a significant challenge. This research paper provides an overview of how Artificial Intelligence (AI) can be applied to manage Big Data databases effectively. The integration of AI and Big Data Management allows organizations to fully utilize their data assets, resulting in improved decision-making, operational efficiency, and the discovery of valuable insights. This approach finds applications in various industries, including finance, healthcare, retail, manufacturing, telecommunications, scientific research, and social media, among others.

In the era of digital transformation, organizations are increasingly adopting AI techniques like machine learning, natural language processing, deep learning, and predictive analytics to extract valuable knowledge from vast datasets. By leveraging AI, organizations can uncover hidden patterns, correlations, and trends within their data, enabling data-driven decision-making, process optimization, personalized customer experiences, fraud detection, improved healthcare outcomes, and more. However, managing Big Data databases using AI comes with challenges. Organizations must address concerns related to data privacy, ethics, data governance, and potential biases in AI algorithms to ensure responsible and transparent use of AI in big data management.

In summary, managing Big Data databases using AI is a transformative field that aims to harness AI techniques to handle the complexities of large datasets. By combining AI with robust big data management practices, organizations can unlock the value of their data assets and gain a competitive advantage in the data-driven world we live in.

II. PROBLEM DEFINITION

A Big Data database depicts the procedure of gathering data using both sequential and non-sequential techniques. Artificial intelligence and machine learning models are used in database administration to better data mapping and classification and to speed up processing to improve analytics. Since 95% of organizations regard the need to handle unstructured data as a challenge for their venture, managing large volumes of information without AI in big data is particularly challenging for enterprises.

III. LITERATURE REVIEW

How big data is managed using AI?

Algorithms based on artificial intelligence (AI) improve the capabilities of large data analytics. Raw (untouched and unstructured) data, meta (information about data), and transformed (value-added) data are the three different types of data. Each of these data categories will benefit from the usage of AI for identification, categorization, and decision-

making. AI has the capacity to transform raw data into information that is relevant and practical for use in decision-making when combined with sophisticated big data analytics.

Managing Big Data Storage-

Big Data requires a significant amount of data storage. One strategy is to collect and handle the localized data before sending the storage to a larger, cloud-based storage system. Utilizing a "virtualized" data system, which produces a virtual layer of the data, is another strategy. The location of the data on the network is known to this virtual layer. Only the data required for that particular calculation is accessed when calculations utilizing the AI algorithm are performed in a virtual system. Without the need to copy data files, the original data storage is still present and undamaged. A network-wide data management protocol is used in this method. It also lessens the requirement for memory for data storage.

Big data management techniques using AI -

- Finding Anomalies: AI can examine Big Data to find anomalies (abnormal occurrences) in the data collection. This is applicable to sensor networks and parameters with a predetermined suitable range, Any network node outside the range is recognized as a potential issue that requires addressing.
- Probabilities of Future Results: Bayes theorem can be used by AI to analyze Big Data. The probability of an event happening can be calculated utilizing existing conditions that have a chance of influencing the final result.
- Spotting Patterns: AI can examine Big Data to search for patterns that could otherwise go unnoticed by human oversight.
- Data bars and graphs: AI can hunt for patterns in Big Data by analyzing data sets.

IV. ANALYSIS & FINDING

Many research papers focus on improving the performance of big data databases using AI techniques. They explore methods such as query optimization, workload management, data partitioning, indexing, and caching to enhance query response times and overall system efficiency. AI algorithms are often employed to automate various aspects of data management, such as data ingestion, data integration, data cleaning, and data transformation. Techniques like machine learning, natural language processing, and computer vision to facilitate efficient and automated data management processes.

Research papers often delve into the application of AI algorithms for data analytics and mining in big data databases. They explore how machine learning, deep learning, and statistical modeling techniques can uncover patterns, correlations, and insights from large volumes of data. Big data databases require scalable and distributed computing architectures.

The analysis show how AI algorithms can be leveraged to distribute and parallelize data processing tasks, enabling efficient handling of massive datasets across multiple nodes or clusters. Managing the security and privacy of big data databases is a critical concern and investigate AI-based approaches for data encryption, access control, anomaly detection, and privacy-preserving techniques to safeguard sensitive information stored in big data databases. The ability to process and analyse streaming data in real-time is vital for many applications.

Then are some illustrations of companies that have successfully enforced AI for big data database operation

- AtScale- AtScale uses AI in data medication, data wisdom, and business intelligence to give precious perception without mortal input. AtScale can use " AI" to produce aggregate tables to accelerate performance grounded on end- stoner query actions."
- IBM- IBM has enforced AI in its Db2 database operation system to help automate tasks like database tuning, backup and recovery, and data security.
- Oracle- Oracle has enforced AI in its Autonomous Database to automate tasks like database tuning, doctoring, and elevation.

- Amazon Web Services(AWS)- AWS has enforced AI in its Amazon Aurora database to automate tasks like database scaling, doctoring, and backup and recovery.
- Microsoft- Microsoft has enforced AI in its Azure SQL Database to automate tasks like database tuning, performance optimization, and security.

These companies have successfully enforced AI in their big data database operation systems to automate tasks, enhance performance, and give precious perception without mortal input.

V. CONCLUSION

On the conclusion, the future of managing big data databases using AI have the huge potential for organizations. Advancements in machine learning techniques, particularly in deep learning, transfer learning, will enable more Refined analysis and prediction capabilities from big data. Natural Language Processing (NLP) and text analytics will play a main role in managing unstructured data, allowing organizations to gain valuable insights from textual data sources.

Even smaller companies are embracing AI and big data by leveraging IT resources from data centers and utilizing cloud-based AI tools for analysing the big data they collect. This accessibility is enabling companies of all sizes to harness the power of AI and big data to drive innovation and improve decision-making.

In conclusion, the demand for AI will continue to rise in the foreseeable future, as it is intricately linked with big data. Their combined usage offers immense potential and will play a crucial role in various industries, driven by the value they bring to data analysis and the increasing availability of big data.

VI. FUTURE SCOPE

The future of big data database management using AI holds great potential for businesses. Here are some key areas of growth and development in simpler and more accurate terms:

Real-time analysis: In the future, big data analytics will prioritize real-time analysis, enabling businesses to process and analyse data as it arrives. As a result, decisions and insights can be made more quickly.

AI integration: Integrating AI into databases will make big data more useful for businesses. AI can automate administrative tasks like managing records, conducting market research, and interacting with customers, saving time and improving efficiency.

Advancements in machine learning: Machine learning is evolving rapidly and will have a significant impact on big data. It enables businesses to make more accurate predictions and decisions based on insights derived from data.

AutoML and Automated Data Management: AutoML will continue to automate the development and optimization of machine learning models. AI will also automate various data management tasks, such as data cleansing and integration, improving efficiency in managing big data databases.

In summary, the future scope of managing big data databases using AI involves advancements in machine learning, NLP, real-time analytics, automated data management, privacy-preserving techniques, data governance, and quantum computing. These advancements will enable organizations to extract insights, enhance data security, and revolutionize the way big data is managed and utilized.

REFERENCES

- [1] https://www.researchgate.net/publication/316679062_Big_Data_Analytics_and_Artificial_Intelligence
- [2] <https://www.xane.ai/managing-big-data-databases-using-ai/>
- [3] <https://american-cse.org/sites/csci2020proc/pdfs/CSCI2020-6SccvdzjqC7bKupZxFmCoA/762400b395/762400b395.pdf>
- [4] <https://topsoftwarecompanies.co/big-data/managing-big-data-databases-using-ai>
- [5] <https://sloanreview.mit.edu/article/how-ai-is-improving-data-management/>