EMERGING TRENDS OF BIG DATA APPLICATIONS IN LIBRARIES: AN OVERVIEW

Dr. Sonali Abhay Deshmukh
Librarian
Shri Shivaji Agriculture College, Amravati
E-mail id -ssaclibrary@gmail.com

Abstract:

In a constantly changing environment, educational institutions, libraries, and books all have their own record-keeping and administration methods, making it challenging to obtain data from a single, central source. Large volumes of big data have been produced as a result of library digitization, but this abundance has made it more difficult for academics, researchers, educators, and policymakers to increase the caliber and effectiveness of their work. Delivering books and articles that suit customers' preferences has consequently become extremely difficult. In order to lessen the time it takes to find pertinent reading material and actually access it, this study suggests a system that gathers and combines data from multiple sources and organizations in real time. Reducing the time between finding and consuming content is the main goal. Bridging this gap is crucial since information access can be expensive, especially for those with restricted internet availability. The goal of this research is to create a technique that drastically cuts down on the amount of time needed to find relevant reading material, in line with the principles of contemporary recommendation systems. The technology efficiently examines book descriptions and metadata to find content that matches users' interests. It gathers vast amounts of data from academics, researchers, educators, and policymakers throughout time. The algorithms that can partially automate certain jobs are then trained using this big data. Therefore, information access can be significantly improved by the insights gained from analyzing this large, integrated library data. This development not only makes it easier and less expensive for researchers, academics, and decision-makers to acquire the resources they need, but it also increases the accessibility of information for marginalized and underprivileged populations.

Keywords : Big data, Digitization, Dynamic, Analyzing, Aggregated Libraries, Recommendation System

---

Introduction :

Due to its significance in promoting data-informed decisions and educational assessments, large-scale data collecting have grown in value in recent years. Large datasets are challenging to store, interpret, and visualize for additional processing or analysis because of their size, complexity, and diversity. Large dataset analytics is the process of gleaning valuable insights and revealing obscure patterns from these enormous data stores.

Definition :

N. Deepa (2022) states that large-scale data processing has recently attracted a lot of attention from various scientific and engineering fields. Big data sets have many benefits and uses, but they also present many issues that must be resolved to increase the capacity of services. These difficulties include the management, analysis, security, and safety of large datasets. Big data services and applications can be significantly improved by integrating blockchain technology, which has decentralized and secure features.

Ali Burak Can, (2023) Trendy sequence appearances within ordered transactional datasets, such as market basket datasets, are extracted by sequential pattern mining algorithms. Research using big data processing methods to find recurring patterns in massive datasets is lacking. Optimized sequential pattern mining algorithms that operate on ordered one-dimensional sequences are also required.

Rejitha Ravikumar,(2022) Every day, higher education institutions—which are regarded as knowledge centers—generate enormous volumes of data that, when properly analyzed using computational techniques and technology, can yield insights that enhance organizational performance and the academic experience of students. Knowledge management is widely recognized as being essential to an organization's survival and growth.

Objectives of the Study:

The primary goals of this study are to increase big data awareness among library workers and look at its applications and analytic techniques.

The following are this study's main goals :

- To analyze big data as it relates to libraries;
- To organize the fundamental actions needed to implement big data in libraries;
- To investigate the potential for Big Data services in libraries;
- To inquire into the Big Data analytics technologies that libraries utilize; additionally
- To highlight global efforts utilizing big data in different library projects.

Review of Literature :

Alhassan Mumuni (2024) shows that by simplifying all intermediate processing stages, modern fully automated data processing frameworks that use automated machine learning (AutoML) techniques may transform raw information into useful characteristics for large-scale data projects. The requirement to effectively manage and use enormous volumes of complex and varied data for machine learning and big data applications is what motivates the push to automate these data handling procedures.

Muhammad Naeem (2021) emphasizes that information is being created constantly, second by second, in the current era of huge data development. The richness and complexity

of data, which is frequently unstructured or semi-structured, provide a number of obstacles for modern companies. These challenges include those related to indexing, sorting, searching, processing, and visualizing this data. The five main characteristics of big data are volume, velocity, veracity, variety, and value. Almost all big data models are built on these "5 Vs." Even though volume and velocity have been the subject of much research, there is currently no complete and efficient solution for handling data variety in the market. The majority of classic DBMS solutions use multidimensional data types. However, these outdated techniques are incompatible with many new data types. Big Data is a vast topic that affects a variety of fields, including science, business, economics, and social security. Analyzing big data sets to find trends and learn new things is known as big data analytics.

Yosra Hajjaji (2021) states that the Internet of Things (IoT) generates enormous amounts of data at the moment. The term "big data" here refers to large datasets that need cutting-edge frameworks and technologies for efficient gathering, processing, and management in order to extract insightful information that aids in better understanding and decision-making. High volume, velocity, variety, and authenticity are characteristics of big data. IoT was predicted to link more than 50 billion devices by 2020, greatly contributing to this enormous data explosion, due to the quick growth of smart gadgets and the rising demand for their services.

Component of Big Data :

Big Data is made up of enormous volumes of information that are too big for conventional CPUs or conventional data storage devices to handle well. Big Data technologies are used by numerous multinational corporations to handle the information and operations of different companies. The predicted daily data flow volume before duplication was more than 150 exabytes.

There are five V's components :

- Volume
- Veracity
- Variety
- Value
- Velocity

Volume :

Big Data is a term used to describe vast volumes of data. It symbolizes enormous 'volumes' of data produced daily from numerous sources, including social media, industrial processes, communication networks, machines, human activity, and many more.

Facebook processes around 4.5 billion "Likes," creates about one billion messages every day, and sees over 350 million new postings every day. Big Data technology effectively handle such enormous amounts of data.

Variety :

Large amounts of semi-structured, unstructured, and structured data are being collected from a variety of sources. Data was previously solely collected from sheets and databases. But these days, a wide range of formatssuch as PDFs, emails, audio files, posts on social media, images, and videos—are used to gather data.

The Data is arranged as follows:

Structured Data :

The structured schema defines all of the necessary fields, and it is arranged in tabular form. Generally, relational database management systems (RDBMS) are used to store structured data.

Semi Structured :

There is no rigid schema for semi-structured data formats such as emails, JSON, XML, CSV, and TSV. Systems that use online transaction processing (OLTP) can handle these kinds of data. Usually, tables or relational structures are used to store this data.

Unstructured Data :

Unstructured data comprises a variety of file kinds, including log files, audio recordings, photographs, and documents without a set format. Despite the fact that many businesses have enormous amounts of this type of data, they frequently find it difficult to extract valuable insights from it since it is unstructured and raw.

Quasi-structured Data :

The text in the data was created over time and with significant effort using a variety of tools, but the formats are still irregular.

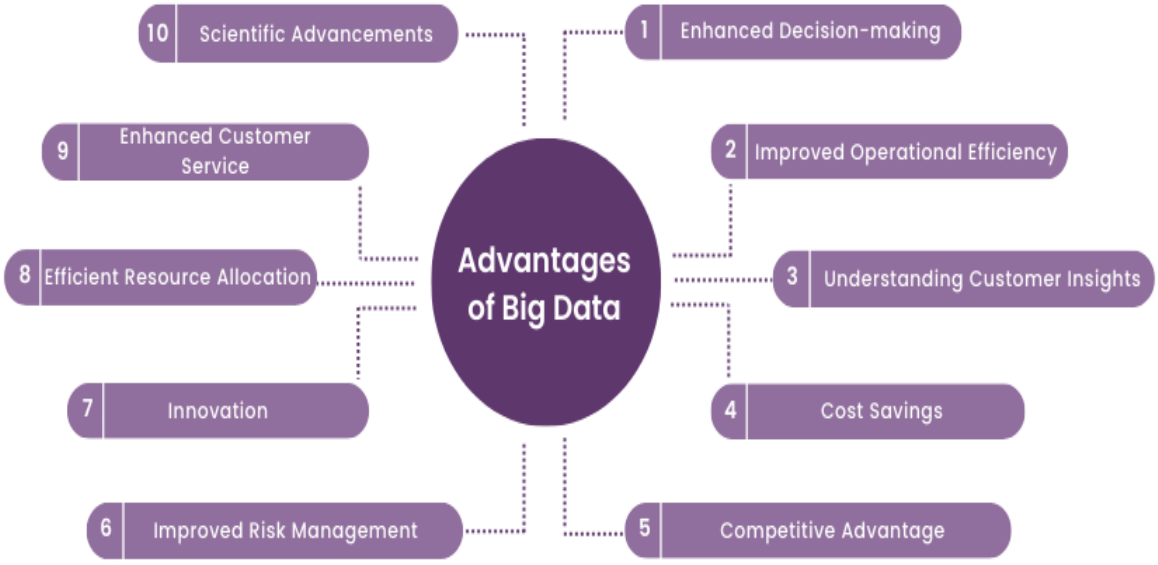
Veracity :

The reliability or trustworthiness of data is referred to as veracity. It entails using a variety of filtering or transformation techniques to guarantee the consistency and accuracy of the data. A crucial component of truthfulness is controlling and preserving data quality. Big Data is also essential for promoting the expansion and advancement of businesses.

Value :

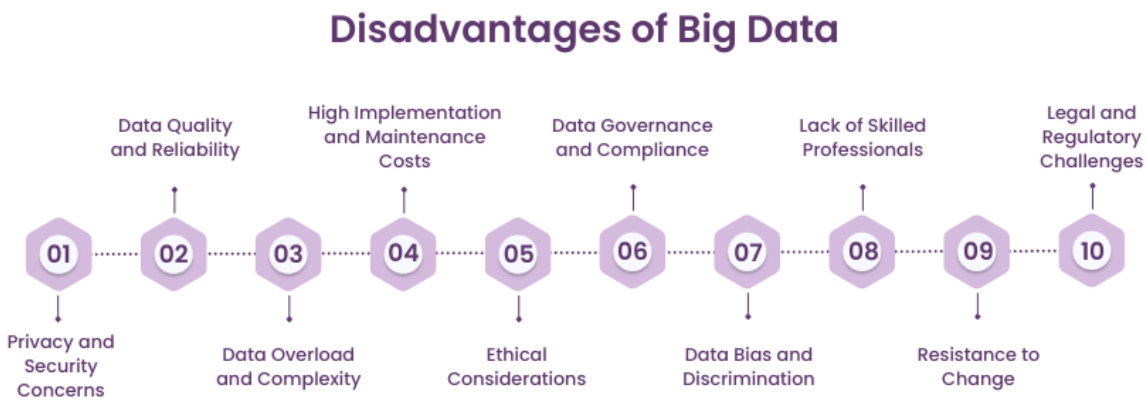
One of Big Data's core characteristics is value. It involves handling data that is significant, dependable, and helpful for analysis and decision-making, not just storing or processing any data.

Advantage & Disadvantage of Big Data :



- Big Data gives businesses access to enormous volumes of data from many sources, empowering them to make well-informed decisions.
- Big Data analytics enables businesses to optimize their operations by discovering inefficiencies, bottlenecks, and areas for development.
- Big Data gives businesses the ability to understand their clients more thoroughly.
- Big Data analytics aids in risk identification and mitigation for organizations.

The advancement of scientific research and the social sciences depends heavily on big data.



- Data privacy and security issues arise from the collection and processing of enormous volumes of data.
- Big data raises dependability and quality issues with data.

- Managing massive data sets might be intimidating.
- Businesses may incur high costs when implementing big data solutions.
- The use of big data raises ethical questions around permission, privacy, and the misuse of personal information.

Future of Big data in Libraries :

Big Data has emerged as a cutting-edge technical development for libraries and information hubs. Data volume is expected to grow dramatically, with IDC projecting that it would reach over a total of 160 zettabytes by 2025. The first—and maybe the only is size. component that becomes apparent when Big Data is spoken. Due to the amount, pace, diversity, variability, and validity of the data that needs to be proces sed in order to provide value and improved understanding, big data is classified as informatio n overload.visualization.Big Data holds the answers to a number of important questions. pertaining to user behavior trends, patterns, and relationships. It has a significant function in assisting libraries in comprehending the evolving needs of users, reorganize and restructure their processes and services appropriately.

Application of Big Data in Libraries :

Languages like French and English can facilitate user-computer communication as well as provide data for system processors to process. Natural Language Processing (NLP) can be used in libraries to create sophisticated expert reference tools and information retrieval systems that let users communicate with the system in conversational, natural language. After processing the input and deciphering the natural language, the computer responds with pertinent information.

Artificial intelligence in libraries also has a useful use in subject indexing. This method uses the technical know-how and intellectual discernment of the librarian or indexer to find, evaluate, and recommend suitable terms while enabling users to interact with the system directly using natural language.Science and Technology Encyclopedia, McGraw-Hill, 2017).

Libraries are using more and more sophisticated information retrieval techniques, including web search engines, electronic databases, Online Public Access Catalogues (OPAC), and robotic systems made especially for book distribution and retrieval. Users can obtain information quickly and creatively thanks to these technologies. Furthermore, the majority of contemporary web search engines, such as Google, now have built-in speech recognition features.

Conclusion:

Libraries can use artificial intelligence as a useful tool to update their offerings and reconsider their approaches in order to stay competitive in the rapidly changing information economy. It is anticipated that AI-driven solutions will significantly improve library

operations in fields including technical services, reference, circulation, resource management, and information retrieval and dissemination. There is conjecture, though, that new technology may lessen the demand for human librarians. Nevertheless, AI is expected to greatly improve service delivery and general library operations, as well as raise the significance of libraries in the rapidly evolving digital ecosystem.

References :

- N. Deepa . (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. Volume 131, June 2022, 209-226 https://doi.org/10.1016/j.future.2022.01.017
- Ali Burak Can,(2023). On the big data processing algorithms for finding frequent sequences. Volume35, Issue24, November 2023, https://doi.org/10.1002/cpe.7660
- Rejitha Ravikumar,(2022). The Impact of Big Data Quality Analytics on Knowledge Management in Healthcare Institutions: Lessons Learned from Big Data's Application within The Healthcare Sector, Special Volume No. 5,2022, https://doi.org/10.56801/seejph.vi.309
- Alhassan Mumuni (2024) Automated data processing and feature engineering for deep learning and big data applications: A survey ,Janauray 2024, https://doi.org/10.1016/j.jiixd.2024.01.002
- Muhammad Naeem (2021) Trends and Future Perspective Challenges in Big Data 26 November 2021, https://link.springer.com/chapter/10.1007/978-981-16-5036-9_30
- Yosra Hajjaji (2021) Big data and IoT-based applications in smart environments: A systematic review, Volume 39, February 2021, https://doi.org/10.1016/j.cosrev.2020.100318
- Chiu, T. K. F., Meng, H., Chai, C. S., King, I., Wong, S., & Yeung, Y. (2022). Creation and evaluation of a pre-tertiary Artificial Intelligence (AI) curriculum. IEEE Transactions on Education, 65(1), 30–39. https://doi.org/10.1109/TE.2021.3085878.
- Ex Libris. (2019). How AI can enhance the value of research libraries. Retrieved October 4, 2019 from: www.libraryjournal.com/?detailStory=how-ai-can-enhancethe-value-of-research-libraries
- Leo,S.Lo.https://americanlibrariesmagazine.org/2024/03/01/reading-between-the-bots/
- American Library Association (2023) Artificial Intelligence. Retrieved June 24. 2023 from https://americanlibrariesmagazine.org/blogs/the-scoop/information-literacy-chatgpt/
- John McCarthy(2023),from https://www.adservio.fr/post/artificial-intelligence-meaning-types.
- Lo (2024), from https://acrl.ala.org/IS/tips-and-trends-ai-developments-and-resources-for-academic-librarians/