RESEARCH PAPER

on

Sales Insight

Submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

Session 2022-23 in

CSE-Data Science

By:-

Kamakshi Agarwal(200321540033)

Arjit Shandilya (200321540015)

Under the guidance of:

Mr. Tarun Kumar Gautam

Assistant Professor

DEPARTMENT OF CSE-DS

ABES ENGINEERING COLLEGE, GHAZIABAD

Estd. 2000

Grade A

NAAC Accredited

AFFILIATED TO

DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW

(Formerly UPTU)

ABSTRACT

Decision-makers now have access to vast volumes of data thanks to the digital age. Big data refers to datasets that are not only large in size but also have a high level of diversity and velocity, making it challenging to manage them with conventional tools and procedures. Solutions must be researched and offered in order to handle and extract value and information from large datasets due to the fast rise of such data. Decision-makers also need to be able to make informed decisions based on such a wide range of fast evolving data, including information from social networks, everyday transactions, and consumer contacts. Massive data analytics, or the use of sophisticated analytics techniques on big data, can deliver this value. This essay seeks to examine.

INTRODUCTION

Imagine a world without data storage, a world where all information about individuals or organisations, all transactions, and all observable aspects are immediately erased after usage. As a result, businesses would be unable to gather important data and expertise, carry out in-depth studies, or provide fresh opportunities and benefits. Everyday continuity now depends on information such client names and addresses, product availability, purchases made, personnel employed, etc. Any organization's success depends on its ability to use data effectively.

This work contributes by offering an examination of the big data analytics literature that is currently accessible. As a result, a selection of the many big data tools, processes, and technologies that may be used are reviewed, along with the applications and opportunities they offer across a number of decision domains.

LITERATURE REVIEW

Big Data Analytics

Recently, the phrase "Big Data" has been used to refer to datasets that get so big that they are difficult to handle using conventional database management systems. They are data sets that are too large to be captured, stored, managed, and processed within a reasonable amount of time using routinely employed software tools and storage systems. A single big data collection can now include anything from a few dozen terabytes (TB) to several petabytes (PB) of data due to the ongoing growth in big data sizes. Because of this, gathering, storing, searching, sharing, analysing, and displaying massive data can be challenging. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before.

Therefore, big data analytics refers to the application of advanced analytical techniques to large data collections. Large data sample analytics reveals and takes advantage of business transformation. However, managing a greater collection of data becomes more challenging [17]. We shall begin this part by going over the qualities and significance of big data. Naturally, analysing larger and more complicated data sets that requires real-time or near-real-time capabilities may frequently result in commercial benefits; nevertheless, this creates a need for new data structures, analytical techniques, and tools. Consequently, the next section will elaborate on big data analytics tools and methods, specifically, starting with big data storage and administration, then going on to the huge.

Characteristics of Big Data

Big data refers to data that, due to its size, dispersion, diversity, and/or timeliness, necessitates the use of novel technological architectures, analytics, and tools in order to provide insights that open up fresh sources of economic value. The three V's, or volume, variety, and velocity, are the three key characteristics of big data. The magnitude and scope of the data are revealed by its volume, according to N. Elgendy and A. Elragal. The term "velocity" describes how quickly or how frequently data is generated or changed. In addition to the many applications and methods of data analysis, diversity also encompasses the various forms and types of data.

Big Data Analytics Tools and Methods

There is a growing demand for quicker and more effective methods of evaluating this data due to the development of technology and the vast amounts of data that are now coming into and out of enterprises every day. Making effective judgements at the appropriate moment requires more than just having a tonne of data available. Traditional data management and analysis methods and infrastructures are no longer suitable for the quick processing of such data sets. As a result, new big data analytics-specific tools and techniques as well as the necessary architectures for storing and handling such data are needed. As a result, everything from the data itself and its collection through its processing to the final derived conclusions is impacted by the rise of big data.

Big Data Storage and Management

Where and how this data will be stored once it has been obtained is one of the first issues that businesses must deal with when working with big data. Relational databases, data marts, and data warehouses are examples of classic techniques for storing and retrieving structured data. Using tools called Extract, Convert, Load (ETL) or Extract, Load, Transform (ELT), which extract data from external sources, transform it to meet operational needs, and then load the data into the database or data warehouse, the data is uploaded to the storage from operational data stores. As a result, the data is prepared for data mining and online analytical operations by being cleansed, processed, and catalogued.

A replication technique is also used to secure the data among the nodes, guaranteeing availability and dependability despite any node failures [3]. The Data Nodes and the Name Nodes are the two different categories of HDFS nodes. The Name Node serves as a mediator between the client and the Data Node, pointing the client to the specific Data Node that has the requested data. Data is stored in duplicated file blocks among the many Data Nodes.

Big Data Analytic Processing

The processing of the huge data occurs after storage. According to, large data processing has four essential criteria. The quickest data loading is the initial need. It is vital to speed up data loading because disc and network traffic interferes with query executions while the data is being loaded. Fast query processing is the second need. Many queries are response-time essential due to the demands of high workloads and real-time requests. As a result, as the number of inquiries grows fast, the data placement structure must be able to maintain high query processing rates. The third prerequisite for large data processing is the extremely effective use of storage space.

Map Reduce is a parallel programming approach that is appropriate for huge data processing and is modelled by the "Map" and "Re-duce" of functional languages. It serves as the foundation of Hadoop and handles data processing and analytics. EMC claims that the MapReduce concept is built on scaling out rather than scaling up, or on adding additional computers or resources as opposed to increasing the power or storage capacity of a single machine. The core concept of MapReduce is to divide a work into stages and carry out those steps concurrently in order to reduce the amount of time needed to accomplish the activity.

Big Data Analytics

In addition to some of the most popular advanced data analytics techniques, such as association rules, clustering, classification and decision trees, regression, and others, several other analyses are now often used with big data. Social media is now crucial for social networking and content exchange, for instance. However, the volume of material produced by social media platforms is immense and generally untapped. Social media analytics, on the other hand, may be used to evaluate such data and derive insightful knowledge and forecasts. To gather, monitor, synthesise, analyse, and visualise social media data, informatics frameworks and tools are developed and evaluated as part of social media analytics.

Additionally, as online opinion data sources like blogs, product reviews, forums, and social data from social media sites like Twitter and Facebook expand significantly, sentiment analysis, also known as opinion mining, is becoming more and more crucial. Text mining enables sentiment analysis, which is focused on examining and comprehending emotions from subjective text patterns. It helps categorise perspectives as positive or negative by identifying the beliefs and attitudes of individuals toward particular issues. Sentiment analysis finds words that are indicative of a sentiment as well as associations between words so that sentiments may be accurately detected. Sentiment analysis employs natural language processing and text analytics to identify and extract information.

Big Data Analytics and Decision Making

Big data is important from the standpoint of the decision maker since it may offer valuable information and knowledge that can be used as a foundation for decisions. The process of making managerial decisions has received significant attention and study throughout the years. Big data is a resource that decision-makers are finding to be increasingly valuable. Scanners, mobile phones, loyalty cards, the web, and social media platforms are just a few examples of the many sources of large amounts of highly specific data that have opportunities to offer companies enormous advantages. This is only possible if the data is properly analysed to yield insightful knowledge that enables decision-makers to take advantage of the opportunities that result from the wealth of historical and real-time data generated through supply chains, production processes, customer data, and other sources.

Additionally, businesses are used to examining internal data like sales, shipments, and inventories. Big data may, however, offer cumulative value and information when used to analyse external data, such as customer marketplaces and supply networks. Making better educated judgements based on meaningful inferences from the data is important as a result of the growing volumes and varieties of unstructured data available.

Customer Intelligence

Big data analytics has a great deal of promise for consumer intelligence and may be very advantageous to sectors like telecommunications, finance, and retail. Big data may increase transparency and give stakeholders faster, easier access to pertinent data. Big data analytics may give businesses the opportunity to profile and segment their consumer bases depending on various socioeconomic factors, as well as raise customer retention and satisfaction levels. This can help them recognise sales and marketing possibilities, advertise to different segments depending on their preferences, and make more educated marketing decisions.

Additionally, monitoring consumer attitudes about companies and identifying key influencers using SNAs may help businesses respond to trends and carry out direct marketing. Big data analytics can make it possible to create predictive models of consumer behaviour and purchasing trends, increasing total profitability [4]. Even businesses that have long employed segmentation are starting to embrace more advanced big data approaches, such real-time micro-segmenting of clients, to target promotions and advertising. Therefore, big data analytics may help firms by allowing better targeted social influencer marketing, identifying and anticipating trends from market emotions, as well as examining and comprehending churn and other consumer behaviours.

Supply Chain and Performance Management

Big data analytics may be used in supply chain management to foresee changes in demand and then match their supply. Manufacturing, retail, as well as the transportation and logistics sectors, may all gain from this. Organizations may automate replenishment choices, which can save lead times, costs, and delays, as well as process disruptions, by assessing stock utilisation and geolocation data on deliveries. Additionally, by examining supplier data to track performance, choices on changing suppliers based on quality or price competitiveness may be made. Additionally, instantaneous execution of alternative price scenarios might result in lower inventory and higher profit margins.

Performance management is another area where big data analytics may be beneficial, and the public sector and healthcare sector in particular can gain much from this. Staff performance data may be tracked and projected using predictive analytics solutions, which can help with the growing demand to increase productivity. As a result, departments may be able to more effectively relate their strategic goals to the results of their services or users' experiences. Predictive KPIs, balanced scorecards, and dashboards can be used within the organisation to introduce operational benefits by enabling the monitoring of performance as well as improving transparency, objectives setting, planning, and management tasks. This is due to the availability of big data and performance information as well as its accessibility to operations managers.

Quality Management and Improvement

Big data may be utilised for quality management, particularly in the manufacturing, energy and utilities, and telecommunications sectors. By enhancing the quality of the products and services offered, this can boost profitability and save expenses. Predictive analytics on big data, for instance, may be used to reduce performance variability and prevent quality problems by sending early warning signals during the production process.

Due to the ability to prevent costly production disruptions by recognising them in advance, this can lower scrap rates and shorten the time to market. Big data analytics may also contribute to improvements in manufacturing lead times. Additionally, managers may be able to make choices for quality control more quickly thanks to real-time data analysis and machine log monitoring. Big data analytics can also make it possible to estimate bandwidth in reaction to client behaviour and monitor network demand in real-time.

Additionally, transportation may be changed and enhanced by evaluating data from distributed sensors on portable devices, roadways, and automobiles, which give real-time traffic information. Drivers may travel more safely and with less disturbance to the flow of traffic if they can anticipate and avert traffic bottlenecks. With "intelligent" linked automobiles, this new kind of traffic ecology has the potential to transform both transportation and how roads are used. Big data applications can therefore offer intelligent routing based on current traffic statistics derived from individual location data. Such applications may also alert users in real-time of accidents, planned roadwork, and congested regions, as well as instantly contact for assistance when problem is identified by the sensors.

Big data may also be used to understand changes in the frequency, location, and severity of weather and climate. This can help people and businesses who depend on the weather, such farmers, as well as the travel and transportation industries. Furthermore, with the development of new sensors and analytic methods, it is now possible to foresee weather-related natural disasters and implement preventative or adaptable actions in advance.

Risk Management and Fraud Detection

Big data analytics may be useful for risk management in sectors like investment or retail banking as well as insurance. Big data analytics may assist in the selection of investments by comparing the possibility of returns against the chance of losses since the assessment and acceptance of risk is a crucial component for the financial services sector. Big data from both within and outside the company may be used to fully and dynamically assess risk exposures. It is possible to foresee weather-related natural disasters and implement preventative or adaptable actions in advance thanks to climate models and more accurate weather predictions.

Big data analytics may be used to detect and stop fraud, particularly in the government, banking, and insurance industries. Automated fraud detection already frequently uses analytics, but businesses and industries are striving to leverage big data's potential to enhance their systems. They may be able to match electronic data from many sources, including both public and private ones, and carry out analyses more quickly thanks to big data.

By accurately highlighting outlier events, customer intelligence may be utilised to predict typical customer behaviour and identify suspicious or diverging activity. Big data on prevalent fraud patterns may also help systems learn about new fraud kinds and take appropriate action when criminals become used to the previous techniques that were meant to catch them. Additionally, SNAs may be used to locate networks of cooperating fraudsters and find proof of false insurance or benefit claims, which will prevent more fraudulent behaviour from being undetected [4]. Therefore, big data technologies, methodologies, and governance procedures may significantly speed up the discovery and detection of compliance trends within all accessible data sets, which can boost the prevention and recovery of fraudulent transactions.

Conclusion

In this study, we looked at the cutting-edge subject of big data, which has recently attracted a lot of interest due to its alleged unheard-of prospects and benefits. We are now living in the information age, and a vast variety of high-velocity data are being created every day. These data include intrinsic intricacies and patterns of hidden knowledge that should be collected from them and put to use. Therefore, by using cutting-edge analytical techniques on large data and uncovering hidden insights and valuable knowledge, big data analytics may be employed to leverage business transformation and better decision making.

Big data may be subjected to these analytics to extract useful information that can be used to improve decision-making and support well-informed judgements. As a result, a few of the several domains where big data analytics may assist and support decision making were looked at. Big data analytics was discovered to open up a wide range of possibilities in a variety of uses and fields, including consumer intelligence, fraud detection, and supply chain management.

Additionally, its advantages can assist several fields and businesses, including manufacturing, retail, communications, and healthcare.

And last, if used properly, any new technology, including big data, which is a great sector with a promising future, may bring about a number of possible advantages and improvements. Big data, however, is extremely challenging to manage. It has to be properly managed, integrated, federated, cleaned, processed, analysed, etc. Due to the increased quantities, speeds, and variety of data and sources that must be managed, big data exponentially worsens all the issues associated with traditional data management. Future research might thus concentrate on developing a large data management framework or roadmap that can address the aforementioned challenges.

References

1. Adams, M.N.: Perspectives on Data Mining. International Journal of Market Research 52(1), 11–19 (2010)

2. Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499 (2010)

3. Bakshi, K.: Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7 (2012)

4. Cebr: Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44 (2012)

5. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C.: MAD Skills: New Analy-sis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492 (2009)

6. Cuzzocrea, A., Song, I., Davis, K.C.: Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011)

7. Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1–24 (2012)

8. Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. “Teachers and Student Achievement in the Chicago Public High Schools.” Journal of Labor Economics 25 (1): 95–135. https://doi.org/10.1086/508733.

9. Abadie, Alberto. 2021. “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.” Journal of Economic Literature 59 (2): 391–425. https://doi.org/10.1257/jel.20191450.

10. Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings.” Econometrica 70 (1): 91–117. https://doi.org/10.1111/1468-0262.00270

11. Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2017. “When Should You Adjust Standard Errors for Clustering?” November. https://doi.org/10.3386/w24003.

12. Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” Journal of the American Statistical Association 105 (490): 493–5. https://doi.org/10.1198/jasa.2009.ap08746.