Educational Administration: Theory and Practice
2024, 30(1), 932 - 938
ISSN: 2148-2403
https://kuey.net/
Research Article

"Big Data Analytics: A Review"

Dr. Nirmal Kaur[1*], Mr. Amarjeet Singh[2]

[1*]Csa (Computer Science And Applications) Sbbsu (Sant Baba Bhag Singh University) Jalandhar, India Nparharsbbs@Gmail.Com
[2]Csa (Computer Science And Applications) Sbbsu (Sant Baba Bhag Singh University) Jalandhar, India Psamar2003@Gmail.Com

Citation: Dr. Nirmal Kaur, Mr. Amarjeet Singh (2023), "Big Data Analytics: A Review" , Educational Administration: Theory and Practice, 30(1), 932 - 938
Doi: 10.53555/kuey.v30i1.5816

ARTICLE INFO

ABSTRACT

The rapid expansion of social networks has propelled a surge in the examination of social data,, garnering significant interest in identifying communication patterns among users and comprehending their behaviors. The imperative task lies in extracting actionable insights from the variety, vast volume, and speed at which data are acquired, tailored to meet specific business requirements. This necessitates the evolution of specialized tools and methodologies for large-scale data analytics in combination with robust architectures for data management and processing. The convergence of social media and big data presents an unprecedented opportunity to elevate management practices to new heights. This paper delves into the integration of big data within social media realms. It delineates the two primary categories of big data analytic approaches: content-oriented and network-oriented methodologies. Furthermore, it elucidates the core concepts, tools, evaluation techniques, applications, and advantages of huge-scale data analytics within this context.

Keywords: SNNs, SNA, big data, social media, approaches.

I. INTRODUCTION

Large data sets are rapidly emerging as a critical area for businesses across various sectors, including education. Put simply, it involves amalgamating data from diverse sources and discerning patterns within it, which may leveraged for numerous purposes such as enhancing market intelligence and educational research. Businesses, irrespective of size, are either implementing or planning to implement big data strategies. Beyond market intelligence, its application spans across a plethora of sectors including healthcare, scientific research, complex manufacturing industries like aviation and heavy machinery, public utilities, traffic management, oil and gas exploration, telecommunications, retail, banking, insurance, defense, and security.

In the education sector, applications highlighted in this discourse encompass the amalgamation of various data sources concerning students, including test records, behavior patterns, and teacher observations over time, to facilitate more precise and timely interventions.

Current developments in technology have resulted in a substantial proliferation of data generated in everyday activities such as shopping, travel, banking, manufacturing, trading, public utilities, governance, sports, entertainment, science, education, and health. Consequently, commercial entities, research institutions, and governments are increasingly recognizing the significance of harnessing this data for their advancement. Consequently, The 9 study of large-scale data has attracted a lot of interest from Academics from diverse fields of study, in addition to stimulating interest in non-academic spheres.

Big data encompasses the collection, aggregation, processing, and utilization of information gathered from diverse sources. Specifically, it refers to considerable databases necessitating intricate processing and visualization beyond the capabilities of traditional data processing software. The McKinsey Global Institute claims that, big data denotes datasets exceeding the capacity of typical database software instruments for capture, storage, management, and analysis. Gartner Inc. popularized the 3V's model of big data, defining it as high volume, velocity, and/or a range of information resources demanding novel processing methods to in order to facilitate improved decision- making, insight discovery, and process optimization. The term "volume" pertains to dataset complexity rather than merely its size, "variety" denotes diverse structured or unstructured data types such as text, numeric, video, audio, and log files, while "velocity" refers to the rapid availability of data for analysis. Additionally, other V's like "Veracity" (ensuring data integrity and organizational confidence in its use) and "Value" (determining if new data enhances organizational value) are

also underscored.
The fig. 1 illustrates the units which have become used to store information within large data. The below fig. represents that the highest unit for storing information in massive data is yottabyte and the lowest unit is kilobyte.

Metric prefixes (multiples of bytes)				
Decimal		*Binary*		
Value	*Metric*	*Value*	*JEDEC*[1]	*IEC*[2]
1000	kB Kilobyte	1024	KB Kilobyte	KiB kikibyte
1000^2	MB megabyte	10242	MB Megabyte	MiB mebibyte
1000^3	GB gigabyte	10243	GB Gigabyte	GiB gibibyte
1000^4	TB terabyte	10244	TB Terabyte	TiB tebibyte
1000^5	PB petabyte	10245		PiB pepibyte
1000^6	EB Exabyte	10246		EiB exbibyte
1000^7	ZB zettabyte	10247		ZiB zebibyte
1000^8	YB yottabyte	10248		YiB yobibyte

Fig. 1: Information Storage Units in Big Data

Digital media and huge amounts of data

Understanding customers is vital for business success, and monitoring their online behavior has become increasingly crucial. Organizations are investing in gathering analytics, utilizing big data as a pivotal component, particularly in monitoring social media activity on platforms like Twitter, Facebook, and LinkedIn. Social media analytics provide insights into internet users' behavior, encompassing data on web browsing, online shopping, customer feedback, and marketing research on social networks. This enables organizations to gain timely and comprehensive insights into consumer behavior, guiding market intelligence strategies across various objectives such as advertising, product launches, brand management, customer loyalty, personalized services, market trends, competitor analysis, risk mitigation, cost- saving, and business expansion.
The integration Social networks and large amounts of data has spurred the emergence of sentiment analysis as a burgeoning field of study. Its objective is to comprehend people's opinions and shares in their daily lives. Businesses leverage this information to understand their customers better and refine their operations accordingly. Educational institutions stand to get further advantages by "listening" to students and gaining deeper insights into their perceptions. Utilizing sentiment analysis on students' social media activity offers a useful resource for comprehending their online behavior and feedback on various aspects of the educational system, including university admissions, qualification features, exams, and aspirations. Organizations can utilize this information to inform their marketing strategies, such as targeting regions with lower-than-expected online activity, monitoring exam experiences through online forums, understanding their brand perception among students, and gathering feedback on new products.

II. WORKING OF MASSIVE DATA IN SOCIAL MEDIA

Social media stands as the richest wellspring of massive data upon which organizations rely for projecting their business needs and advancement. Establishing pages on platforms like Twitter, Facebook, and Google+ initiates the tallying of likes, tweets, shares, and comments, constituting the raw material of big data. Even if these numbers might impressive, they often lack granularity concerning sales, marketing campaigns, new clientele, or revenue generation. The overwhelming amount and speed of data streaming in incessantly render traditional database tools inadequate for organizing and storing big data. Hence, devising strategies to select pertinent data and employing appropriate tools for processing are imperative to yield desired outcomes.
Enterprising businesses harness this data to discern trends, detect patterns, and extract other invaluable insights through huge data evaluation. Analytics for Big data tools, encompassing software products supporting predictive and prescriptive analytics applications, operate on big data computing platforms.

These platforms typically entail parallel processing systems comprising clusters of servers, scalable distributed storage, and technologies such as NoSQL and Hadoop databases. These tools are engineered to facilitate swift analysis of vast datasets, often in real-time scenarios. Effective analytical processes and skilled personnel are pivotal in unveiling essential information. Data scientists, acting as big data analysts endowed with strong business acumen, navigate through the sea of data to identify patterns. Their role entails exploration, inquiry, conducting "what if" analyses, challenging existing assumptions and processes. Big data analytics leverage inductive statistics and nonlinear system identification concepts to derive laws of regressions, nonlinear relationships, and causal effects from extensive datasets, facilitating insights into relationships, dependencies, and predictive behaviors.

Armed with data and analytical insights, data scientists draw informed conclusions and offer recommendations to organizations. Big data technology furnishes companies with insights into consumer decision-making processes and significantly influences how organizations tackle business challenges.

In essence, big data revolves around extracting fresh insights from the vast data pool and integrating them into business operations, data warehouses, processes, and applications. Challenges in harnessing big data from social media include data inconsistency and complexity, deciphering data patterns, time constraints for data consumption, logical output design, security concerns, and familiarity with data extraction and handling tools and techniques. Analyzing big data empowers analysts, researchers, and business users to make more informed and prompt decisions. Using two cutting-edge analytics methods, like text analytics, data mining analytics, machine learning, statistics, predictive analytics, and processing of natural language enables businesses to explore previously untapped data sources, thereby garnering new insights and facilitating significantly improved decision-making processes. "Social Media Demographics to Inform a Better Segmentation Strategy" offers comprehensive statistics on various facets of social media to aid in crafting winning marketing strategies.

III.TOOLS USED IN BIG DATA ANALYSIS

In the past, analysis relied on database software, which employed JOIN mechanisms to merge tables and aggregate fields before passing a subset of data to reporting software. However, executing complex JOIN commands could often lock up the database for extended periods, halting other operations. With big data, the complexity increases due to the vast scale of information, typically distributed across multiple servers, necessitating coordinated data compilation among them.

i. Hadoop: a widely adopted tool for organizing servers into racks, complemented by NoSQL databases for storing data on these racks. This amalgamation surpasses the capabilities of traditional single-machine setups. The terms "Hadoop" and "big data" are often used interchangeably. Hadoop, an Apache distributed data processing software library, facilitates distributed processing of large datasets across computer clusters via straightforward programming models. Developed in Java, it is engineered to scale seamlessly from individual servers to expansive clusters comprising thousands of machines, each offering local computation and storage. The Apache Foundation sponsors several related projects that extend Hadoop's capabilities, compatible with various platforms and operating systems. Moreover, numerous vendors provide supported versions of Hadoop and its associated technologies.

ii. NoSQL: NoSQL also known as "Not Only SQL," encompasses a framework of databases tailored for high-performance, rapid processing of massive- scale information. It serves as a database infrastructure adept at meeting the rigorous demands of big data. NoSQL databases, characterized by their unstructured nature, prioritize agility and speed over strict consistency requirements. They embrace the concept of distributed databases, enabling storage of unstructured data across multiple processing nodes or servers.

Originating from Google, MapReduce revolutionized data processing with its programming model and software framework, facilitating parallel processing of vast datasets across large compute node clusters. For instance, a data procedure that might consume 20 hours on a centralized relational database system could be completed in just 3 minutes using MapReduce. GridGain, HPCC (High-Performance Computing Cluster), and Storm are tools augmenting Hadoop's capabilities. GridGain, serving as an alternative to MapReduce, is appropriate for use with Hadoop Distributed File System, offering in-memory processing for swift analysis of real-time data. Storm, owned by Twitter, earns the moniker "Hadoop of real-time" due to its support for distributed real-time computation.

Cassandra, a NoSQL database initially developed by Facebook and now managed by the Apache Foundation, finds extensive usage in organizations with large, dynamic datasets, including Netflix, Urban Airship, Twitter, Constant Contact, Cisco, Reddit, and Digg. HBase serves as the non- relational data store for Hadoop. Other noteworthy databases include Neo4j, MongoDB, Hibari, CouchDB, Terrastore, OrientDB, and Hypertable, each offering distinctive features tailored to diverse needs.

iii. Analytics Sweet Spots: In the pursuit of effective decision-making, ample opportunities arise from analyzing data gleaned from social networks to understand market dynamics. The proliferation of social media, internet users, and mobile devices has provided organizations with a means to comprehend external factors influencing their decision-making processes. Today, the challenge lies not in the selection of analytical tools but in pinpointing the Sweet Spot of information that empowers organizations to make

timely, informed decisions, gaining a competitive edge.
Consider, for instance, the analogy of interest rates occupying a sweet spot when they manage inflationary pressures without detriment to the overall market. Similarly, optimal levels of employment stimulating economic growth without inducing inflation through wage pressures constitute another sweet spot. However, a strategic sweet spot is transient, evolving with shifts in markets, technologies, competitors, organizational dynamics, and available skills. Consequently, businesses must continually reassess and redefine their strategic sweet spots.
As Paul O'Dea elucidates in his article "Hitting the Sweet Spot," a business's strategic sweet spot aligns customer needs with its unique offerings. For instance, in the Pareto principle where "20% of customers deliver 80% of the profit," this 20% represents the sweet spot. Focusing energies on identifying and capitalizing on the sweet spot fosters clarity, focus, and alignment, propelling the business in a positive trajectory.
However, rushing to execute a business strategy before identifying the right sweet spot wastes valuable resources on the wrong customers, hindering growth. Customers falling outside the sweet spot impede progress and should be deemed off-strategy. Great companies intimately understand their sweet spot customers and consistently deliver superior value, tailoring their business around this niche.
Identifying the sweet spot involves asking the fundamental questions related to business.
Answers to the questions, framed as a questionnaire, yield a rich repository for analysis. The business's strategic sweet spot emerges at the intersection of these elements, offering clarity, opportunity, and optimized outcomes. This enables more efficient resource utilization, sharper focus, and enhanced cost-effectiveness, ultimately resulting in increased profitability, greater return on investment, and the cultivation of potential customers.
Businesses adept at identifying their sweet spot enjoy superior positioning and optimal outcomes compared to their counterparts. Sweet spots can be leveraged to:
Identify and target the most lucrative market segments

- Innovate new products and services
- Assess partnerships and alliances to maximize opportunities
- Concentrate efforts for higher return on investment (ROI)

iv. Analytics Quagmires: Among the foremost challenges in big data analytics lies the sharing of social media data, constituting a significant quagmire. A primary deterrent to the widespread adoption of big data analytics is the scarcity of real- time data, inadequacies in data collection specifications, and the absence of seamless integration across datasets. Information often becomes skewed due to the lack of standardized procedures for data integration and communication of findings, exacerbated by security concerns surrounding data sharing.
Achieving interoperability among information systems to cohesively organize data necessitates the establishment of standards. The four axioms of interoperability dictate that data should be available for multiple purposes, reusable, and preferably auto- collectible. Cooperation among stakeholders is imperative to ensure application consistency across domains, adhering to legal, ethical, and societal requirements, including security, privacy, and confidentiality. Moreover, a universal set of data elements with precise, unambiguous definitions must serve as the foundation for all systems, fostering interoperability among diverse information systems. Another prevalent issue in big data predictive analytics is the existence of siloed data. What sets predictive analytics for social media apart is the immediacy of information. Companies equipped with real-time insights can proffer timely offers tailored to a customer's current interests, thus increasing the likelihood of immediate action. However, leveraging present insights to predict the future poses challenges, as companies grapple with uncertainty regarding the accuracy of gathered insights and translating them into actionable plans. Additionally, the presence of siloed departments hampers data aggregation and collaborative strategy formulation. Social listening platforms play a pivotal role in scanning media for specific text, images, keywords, and sentiments, which must be correlated with individual characteristics and timeframes. For instance, discerning whether a person expressing negative sentiments about a product is a longstanding customer adds context and value to sentiments when integrated with the company's Customer- Relationship-Management (CRM) system. However, if the CRM operates as a siloed department disconnected from social media listening platforms, comprehensive prediction becomes unattainable. Consequently, analysis employing dual parameters for effective inference becomes unfeasible.
Businesses traditionally solicit customers' addresses, mobile numbers, and email addresses, yet it is imperative to collect Twitter handles and other social media identifiers. Technical limitations stemming from mismanagement and commercial turf battles further impede the seamless implementation of big data analytics.
v. Analytics Maturity: As per a study conducted by the IBM Institute of Business Value (IBV), in 2014, 63% of organizations experienced a positive return on their analytics investments within a year. Furthermore, the study highlighted that 74% of respondents anticipate a continued acceleration in the speed at which executives expect new data-driven insights in the years ahead. This success stems from the recognition that the big data and analytics maturity model not only encompasses technological aspects but also considers crucial business factors. The maturity model can be delineated into six key areas:

- Information: Foundational to this model is the utilization of data to manage business operations. However, highly mature organizations perceive data as a valuable business asset, sourcing it not only from transactional records but also from personal and external sources. These organizations facilitate governed access to data across the organization, providing it with meaning and context.
- Analytics: Mature use of analytics optimizes business outcomes. While organizations may possess reports showcasing financial performance and regulatory compliance, analytics is indispensable for understanding causality and predicting future trends. Insights derived from analytics aid in enhancing customer engagement and operational efficiency.
- Architecture: A coherent infrastructure and system are imperative for establishing enduring organizational capabilities. It ensures ease of access for end-users, agility in addressing business needs, and a managed approach to data access. A mature architecture accommodates the four characteristics of big data—volume, variety, velocity, and veracity—through the creation and systematic reuse of architectural patterns, assets, and standards.
- Implementation: Merely accessing data and deriving insights through analytics does not generate business value unless implemented effectively with the collaboration of all stakeholders. A mature organization fosters a culture of visualization, sharing, and feedback, facilitating continuous learning and improvement. It implements architecture provided by rich data and analytics services, evolving a mature business model optimized in all aspects.
- Business Strategy: While technology is instrumental in acquiring data and executing analytics, business expertise is essential for deriving meaningful insights and leveraging them to drive differentiation. Mature organizations possess the capability to explore data for new opportunities and construct quantified business cases, enabling innovation and the creation of new business models.
- Governance: Information governance is pivotal for the success of big data projects. Policies must be established and enforced to ensure confidence in information and that resulting insights inform decision-making efforts. Governance policies encompass provenance, currency, data quality, master data and metadata, lifecycle management, security, privacy, and ethical use.

By adhering to these six areas of the analytics maturity model, organizations can enhance their capabilities, derive deeper insights, and drive strategic outcomes effectively in the realm of big data analytics.

IV. APPLICATIONS OF BIG DATA ANALYTICS

Big data analytics has myriad applications across various industries. Here are some prominent ones:

i. Healthcare: Big data analytics can be used to analyze large volumes of medical data to identify patterns, predict disease outbreaks, personalize treatment plans, and improve patient outcomes. Clinical research and drug discovery are also done with the aid of big-data-analytics.

ii. Finance: In finance, big data analytics is used for fraud detection, risk management, algorithmic trading, customer segmentation, and personalized financial services.

iii. Retail: Retailers use big data analytics for customer segmentation, demand forecasting, inventory management, pricing optimization, and personalized marketing campaigns.

iv. Telecommunications: Telecom companies utilize big data analytics to analyze network performance, predict customer churn, optimize bandwidth allocation, and improve customer service.

v. Manufacturing: Big data analytics can improve efficiency in manufacturing processes through predictive maintenance, quality control, supply chain optimization, and real-time monitoring of equipment and assets.

vi. Transportation and Logistics: Big data analytics is used for route optimization, fleet management, predictive maintenance of vehicles, demand forecasting, and supply chain visibility.

vii. Energy: In the energy sector, big data analytics can optimize energy production and distribution, predict equipment failures, monitor energy consumption patterns, and enable smart grid management.

viii. Marketing and Advertising: Marketers use big data analytics for targeted advertising, customer sentiment analysis, social media monitoring, and campaign performance measurement.

ix. Government and Public Sector: Governments use big data analytics for policy planning, law enforcement, traffic management, urban planning, healthcare management, and social welfare programs.

x. Education: Big data analytics is increasingly being used in education for personalized learning, student performance analysis, adaptive learning platforms, and predictive analytics to identify at-risk students.

These are just a few examples, but the potential applications of big data analytics are vast and continue to expand as technology advances and data sources proliferate.

V. ADVANTAGES OF BIG DATA ANALYTICS IN SOCIAL MEDIA

Big data analytics has several advantages in the context of social media:

i. Understanding Customer Behavior: Big data analytics can analyze vast amounts of social media data to gain insights into customer behavior, preferences, and sentiment. This understanding helps businesses tailor their products, services, and marketing strategies to better meet customer needs.

ii. Targeted Advertising: By analyzing user data and behavior on social media platforms, businesses can create highly targeted advertising campaigns. Big data analytics helps identify specific demographics, interests, and behaviors, allowing for more effective ad targeting and higher ROI.

iii. Customer Engagement and Feedback: Social media analytics enables businesses to monitor and engage with customers in real-time. By analyzing social media conversations and feedback, businesses can respond promptly to customer inquiries, address concerns, and build stronger relationships with their audience.

iv. Competitive Analysis: Big data analytics can be used to monitor competitors' activities and performance on social media. By analyzing competitor data, businesses can identify trends, benchmark their own performance, and gain insights to inform their own social media strategy.

v. Content Optimization: Social media analytics can help businesses understand which types of content resonate most with their audience. By analyzing engagement metrics such as likes, shares, and comments, businesses can optimize their content strategy to create more engaging and shareable content.

vi. Influencer Identification: Big data analytics can identify influential users and opinion leaders within a given industry or niche. By analyzing social media data, businesses can identify potential influencers to collaborate with, amplify their brand message, and reach a larger audience.

vii. Crisis Management: Social media analytics enables businesses to monitor and respond to crises or negative publicity in real-time. By analyzing social media conversations and sentiment, businesses can identify emerging issues, mitigate reputational damage, and proactively address customer concerns.

viii. Product Development: Social media analytics can provide valuable insights into customer preferences and unmet needs. By analyzing social media conversations and feedback, businesses can gather feedback on existing products or services and identify opportunities for innovation and product development.

Overall, big data analytics empowers businesses to leverage social media as a powerful tool for customer engagement, brand building, market research, and competitive advantage.

VI. CONCLUSION

Big-Data-Analytics in market of social media are still in their nascent stages, particularly in developing countries where organizational responsibilities remain unclear. Despite the potent impact of social media, progress is hindered by deficiencies in technological infrastructure and a scarcity of personnel possessing data management and analysis skills. Nonetheless, organizations can leverage the information collected to their advantage by fostering the appropriate company mindset, devising strategic approaches, and implementing suitable technology solutions. Effectively measuring the business value of social activities empowers organizations to glean crucial insights, thereby enhancing and promoting their products and services.

REFERENCES

1. Sepideh Bazzaz Abkenar, Mostafa Haghi Kashani, Ebrahim Mahdipour, Seyed Mahdi Jameii, “Big data analytics meets social media: A systematic review of techniques, open issues, and future directions”, Elsevier: Telematics and Informatics, 57 (2021) 101517.
2. Jitimon Angskun, Suda Tipprasert and Thara Angskun, “Big data analytics on social networks for real-time depression detection “. Angskun et al. Journal of Big Data (2022) 9:69 https://doi.org/10.1186/s40537-022-00622-2.Available at:
3. http://www.cambridgeassessment.org.uk/research-matters/
4. Einav, L. & Levin, J.D. (2013). The data revolution and economic analysis (NBER Working Paper no. 19035). Retrieved from http://www.nber.org/papers/ w19035.
5. Lohr, S. (2012, February 11). The age of big data. The New York Times. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impactin-the-world.html?pagewa nted =all&_r=0.
6. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity (McKinsey Global Institute report). Retrieved from http://www.mckinsey.com/insights/business_technology/big
7. _data_the_next_frontier_for_innovation.
8. Swoyer, S. (2012). Big data – why the 3Vs just don't make sense.Retrieved from http://tdwi.org/articles /2012/07/24/big-data-4th-v.aspx.
9. Loshin, D., “Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph”. Elsevier, 2013.Bowden, J., “Reasons to Explore Big data with Social Media Analytics”, social media Today, 2015, http://www.socialmediatoday.com.
10. Patterson, M., “Social Media Demographics to Inform a Better Segmentation Strategy” Social Media Demographics for Marketers, 2015, http://sproutsocial.com/insights.
11. C. Harvey, 50 Top Open Source Tools for Big data, Datamation, IT Business Edge, 2012, http://www.datamation.com/data-cente.
12. R. Schefren, Find Your Strategic Sweet Spot– Its What Makes a Business Different: Great Companies Always Know Their Sweet Spot, BizShifts Trends, from http://bizshifts- trends.com/2015/.

13. Roski, J., & McClellan, M., Measuring health care performance now, not tomorrow: essential steps to support effective health reform. Health Affairs, 30(4), 2011, 682-689.
14. Hammond, W. E., Bailey, C., Boucher, P., Spohr, M., & Whitaker, P., Connecting information to improve health. Health Affairs, 29(2), 2010, 284-288.
15. DeLoatch, P., Predictive social analytics must clear data silo hurdles, Content Management, TechTarget, 2014, http://www techtarget.com.
16. Palmer,B., Realizing a return on big data investments— within a year, IBM Big data and Analytics Hub, 2014, from http://www.ibmbigdatahub.com.
17. Nott,C., A maturity model for big data and analytics, IBM Big data and Analytics Hub, 2015, from http://www.ibmbigdatahub.com