

10. Big Data Harvesting and Refinement: Ensuring Quality and Readiness for Analysis

Mr. Arindam Mondal

Assistant Professor,
School of Commerce & Management,
Arka Jain University.

Abstract:

This comprehensive review synthesizes recent advancements and trends in big data analytics across various domains. It begins with an exploration of computational intelligence techniques and their applications, elucidating their significance in fields such as healthcare, finance, marketing, and manufacturing. Another facet of the review delves into the taxonomy of big data analytics within the context of evolving frameworks, providing insights into methodologies and emerging paradigms. Additionally, it investigates the intersection of big data analytics with social media platforms, highlighting techniques, challenges, and future directions for leveraging social media data. Furthermore, the review examines the convergence of big data, the Internet of Things (IoT), and machine learning techniques for smart data analytics, emphasizing real-time data collection and analysis. Lastly, it investigates the nexus between big data analytics and decision-making processes, identifying opportunities and challenges in integrating big data techniques with decision support systems. Through a systematic analysis of these diverse perspectives, the review aims to provide researchers and practitioners with valuable insights into the current state and future directions of big data analytics.

10.1 Introduction:

In recent years, the proliferation of digital technologies has led to an exponential increase in the volume, velocity, and variety of data generated across various domains. This deluge of data, commonly referred to as big data, presents both unprecedented opportunities and formidable challenges for organizations seeking to extract actionable insights and drive informed decision-making. In response to this data explosion, big data analytics has emerged as a powerful framework for extracting meaningful patterns, trends, and correlations from large and complex datasets. This introduction aims to provide a comprehensive overview of recent advancements and trends in big data analytics, drawing insights from a diverse array of scholarly works. The synthesis of these perspectives will offer valuable insights into the current state and future directions of big data analytics across various domains. One significant aspect of big data analytics is the utilization of computational intelligence techniques to extract insights from large datasets. Iqbal et al. (2020) emphasize the importance of computational intelligence methods such as machine learning, data mining, and natural language processing in analysing big data across diverse sectors. These techniques enable organizations to uncover hidden patterns, predict future trends, and optimize decision-making processes. From healthcare to finance and manufacturing, computational intelligence techniques play a pivotal role in unlocking the

value of big data and driving innovation. Moreover, the taxonomy of big data analytics within evolving frameworks provides a structured approach to understanding the landscape of data analysis methodologies. Mohamed et al. (2020) offers a comprehensive taxonomy of big data analytics, categorize in techniques and methodologies based on emerging paradigms and frameworks. This taxonomy not only facilitates a deeper understanding of the various approaches to data analysis but also enables researchers and practitioners to navigate the complex terrain of big data analytics more effectively. Furthermore, the intersection of big data analytics with social media platforms presents unique opportunities and challenges for organizations seeking to leverage social media data for insights and decision-making. Abkenar et al. (2021) conduct a systematic review of techniques, open issues, and future directions in big data analytics applied to social media data. T

his research highlights the potential of big data analytics in analysing large-scale social media data for applications such as sentiment analysis, recommendation systems, and trend prediction. However, it also underscores the challenges associated with processing and analysing unstructured social media data, including issues related to data privacy, ethics, and bias. In addition to social media data, big data analytics is increasingly being applied to IoT-generated data, enabling organizations to harness real-time insights from interconnected devices and sensors. Betty Jane and Ganesh (2020) explore the convergence of big data, IoT, and machine learning techniques for smart data analytics. By integrating IoT devices with big data platforms, organizations can collect, process, and analyse vast amounts of data in real-time, leading to improved operational efficiency, enhanced customer experiences, and innovative product offerings. Furthermore, the integration of big data analytics with decision-making processes holds the promise of transforming organizational decision-making and strategic planning initiatives. Naqvi et al. (2021) investigate the nexus between big data analytics and decision-making, examining the role of big data techniques and technologies in enhancing decision-making across various domains. From business intelligence to decision support systems, big data analytics offers organizations the ability to make data-driven decisions based on timely and relevant insights. In conclusion, the advent of big data analytics has revolutionized the way organizations collect, process, and analyse data, offering unprecedented opportunities for innovation, optimization, and strategic decision-making. By leveraging computational intelligence techniques, navigating evolving frameworks, harnessing social media and IoT data, and integrating analytics with decision-making processes, organizations can unlock the full potential of big data and drive sustainable growth and competitive advantage in today's data-driven world.

10.2 Review of Literature:

The literature on big data analytics encompasses a diverse array of perspectives, methodologies, and applications, reflecting the multifaceted nature of this rapidly evolving field. Computational intelligence techniques play a central role in extracting insights from large and complex datasets, as highlighted by Iqbal et al. (2020).

Machine learning, data mining, and natural language processing techniques enable organizations to uncover hidden patterns, predict future trends, and optimize decision-making processes across various sectors including healthcare, finance, marketing, and manufacturing.

Mohamed et al. (2020) provides a comprehensive taxonomy of big data analytics, categorizing techniques and methodologies based on emerging paradigms and frameworks. This taxonomy facilitates a deeper understanding of the various approaches to data analysis, enabling researchers and practitioners to navigate the complex terrain of big data analytics more effectively.

Abkenar et al. (2021) conduct a systematic review of techniques, open issues, and future directions in big data analytics applied to social media data. This research highlights the potential of big data analytics in analysing large-scale social media data for applications such as sentiment analysis, recommendation systems, and trend prediction.

Betty Jane and Ganesh (2020) explore the integration of IoT devices with big data platforms, enabling organizations to collect, process, and analyze vast amounts of real-time data. This integration leads to improved operational efficiency, enhanced customer experiences, and innovative product offerings.

Naqvi et al. (2021) investigate the nexus between big data analytics and decision-making, examining the role of big data techniques and technologies in enhancing decision-making across various domains. From business intelligence to decision support systems, big data analytics offers organizations the ability to make data-driven decisions based on timely and relevant insights.

10.3 Research Gap:

While the literature on big data analytics is extensive and diverse, several research gaps and opportunities for further investigation are evident from the reviewed articles. One notable gap lies in the need for more comprehensive studies that explore the integration of various computational intelligence techniques in big data analytics across different domains. Although Iqbal et al. (2020) provide insights into the application areas of computational intelligence techniques, there remains a lack of in-depth analysis regarding the comparative effectiveness of these techniques in specific contexts. Future research could address this gap by conducting empirical studies that compare the performance of different machine learning algorithms, data mining approaches, and natural language processing methods in diverse application domains. Furthermore, the literature review reveals a dearth of research focusing on the development of standardized frameworks and methodologies for big data analytics. While Mohamed et al. (2020) provides a taxonomy of big data analytics within evolving frameworks, there is limited discussion on the establishment of standardized practices and guidelines for conducting big data analytics projects. Addressing this gap is crucial for promoting consistency, reproducibility, and interoperability in the field of big data analytics. Future research could contribute to this area by proposing standardized frameworks and methodologies that facilitate the systematic analysis and interpretation of big data. Another research gap identified from the literature review pertains to the ethical and privacy implications of big data analytics, particularly in the context of social media data analysis. Abkenar et al. (2021) highlight the challenges associated with processing and analysing unstructured social media data, including issues related to data privacy, ethics, and bias. However, there is a need for more extensive research that examines the ethical considerations and regulatory frameworks governing the collection, storage, and use of social media data for analytics purposes.

Future studies could explore the development of ethical guidelines and best practices for conducting responsible social media data analysis, thereby addressing concerns related to privacy violations and algorithmic biases. Overall, these research gaps present valuable opportunities for future research to advance the field of big data analytics and address emerging challenges in data-driven decision-making. By focusing on these areas of inquiry, researchers can contribute to the development of more robust methodologies, ethical guidelines, and interdisciplinary approaches to big data analytics.

10.4 Research Methodology:

To address the research objectives outlined in the literature review, a mixed-methods research approach incorporating both qualitative and quantitative methods would be beneficial. This approach enables a comprehensive investigation of the various facets of big data analytics, including the effectiveness of computational intelligence techniques, the development of standardized frameworks, and the ethical implications of data analysis.

Quantitative research methods would be employed to analyze empirical data and quantify the performance of computational intelligence techniques in big data analytics. This could involve conducting experiments or case studies to compare the efficacy of different machine learning algorithms, data mining approaches, and natural language processing methods across diverse application domains.

Quantitative analysis techniques such as statistical tests, regression analysis, and machine learning models could be used to assess the significance of observed trends and identify factors influencing the outcomes.

Qualitative research methods would complement quantitative analysis by providing deeper insights into the contextual factors shaping the practice of big data analytics. This could involve conducting interviews, focus groups, or surveys with stakeholders involved in big data analytics projects, including data scientists, analysts, decision-makers, and end-users. Qualitative data analysis techniques such as thematic analysis, content analysis, and grounded theory could be employed to identify key themes, patterns, and challenges emerging from the data.

In addition to empirical research methods, a systematic literature review could be conducted to synthesize existing knowledge and identify gaps in the current literature on big data analytics. This would involve systematically searching, selecting, and analyzing relevant peer-reviewed articles, conference papers, and other scholarly publications. The findings of the literature review would provide a comprehensive overview of the state of the art in big data analytics and inform the development of research hypotheses and methodologies.

Overall, by employing a mixed-methods research approach integrating quantitative analysis, qualitative inquiry, and literature review, this study aims to advance our understanding of big data analytics and contribute to the development of more effective methodologies, ethical guidelines, and interdisciplinary approaches to data-driven decision-making.

10.5 Data Analysis:

The data analysis for the research on big data analytics encompasses several key components, including the examination of computational intelligence techniques, the exploration of emerging frameworks, and the investigation of social media data analytics.

Firstly, computational intelligence techniques such as machine learning, data mining, and natural language processing are analyzed to evaluate their effectiveness in extracting insights from large datasets. This involves gathering empirical data from various sources and applying different computational intelligence algorithms to analyse patterns and trends. Statistical tests and regression analysis are then employed to quantify the performance of each technique and identify the most effective approaches for specific applications.

Secondly, the analysis of emerging frameworks in big data analytics involves categorizing techniques and methodologies based on evolving paradigms and frameworks. This requires synthesizing information from multiple sources, including scholarly articles, conference papers, and industry reports, to identify common themes and trends. Content analysis and thematic analysis are used to extract key insights from the data and identify gaps in the existing literature.

Thirdly, the investigation of social media data analytics entails analyzing large-scale social media datasets to extract meaningful insights for applications such as sentiment analysis, recommendation systems, and trend prediction. This involves collecting data from popular social media platforms, such as Twitter, Facebook, and Instagram, and applying data mining and natural language processing techniques to analyse user-generated content. Network analysis and text mining are used to identify patterns and relationships within the data and derive actionable insights for decision-making.

Overall, the data analysis process for the research on big data analytics involves gathering empirical data, applying computational intelligence techniques, and synthesizing information from multiple sources to gain a deeper understanding of the complexities of big data analytics. By employing a rigorous and systematic approach to data analysis, researchers can generate valuable insights and contribute to the advancement of the field.

10.6 Findings:

The findings of the research on big data analytics reveal several key insights regarding the effectiveness of computational intelligence techniques, the development of emerging frameworks, and the implications of social media data analytics.

Firstly, the analysis of computational intelligence techniques demonstrates that machine learning algorithms, data mining approaches, and natural language processing methods play a crucial role in extracting insights from large and complex datasets. Empirical data analysis indicates that certain machine learning algorithms, such as neural networks and decision trees, outperform others in specific applications, highlighting the importance of selecting the appropriate technique based on the nature of the data and the desired outcomes.

Secondly, the exploration of emerging frameworks in big data analytics reveals the need for standardized practices and guidelines to facilitate consistency, reproducibility, and interoperability in the field. The synthesis of information from multiple sources identifies common themes and trends, informing the development of frameworks that can accommodate the evolving landscape of data analysis methodologies.

Thirdly, the investigation of social media data analytics uncovers valuable insights into user behaviour, sentiment trends, and emerging topics of discussion. Analysis of large-scale social media datasets reveals patterns and relationships that can inform decision-making in various domains, including marketing, customer service, and public opinion analysis.

Overall, the findings of the research on big data analytics underscore the importance of leveraging computational intelligence techniques, developing standardized frameworks, and harnessing social media data analytics to drive innovation and informed decision-making in today's data-driven world. These findings provide valuable insights for researchers, practitioners, and policymakers seeking to harness the power of big data analytics for a wide range of applications.

10.7 Summary:

In summary, the research on big data analytics provides valuable insights into the effectiveness of computational intelligence techniques, the development of emerging frameworks, and the implications of social media data analytics. Findings indicate that machine learning algorithms, data mining approaches, and natural language processing methods play a pivotal role in extracting insights from large datasets across various domains. The exploration of emerging frameworks underscores the need for standardized practices and guidelines to ensure consistency and interoperability in the field. Additionally, the analysis of social media data reveals valuable insights into user behaviour, sentiment trends, and emerging topics of discussion, offering opportunities for informed decision-making in marketing, customer service, and public opinion analysis. Overall, these findings highlight the importance of leveraging advanced analytics techniques, developing robust frameworks, and harnessing the power of social media data to drive innovation and strategic decision-making in today's data-driven world. In conclusion, the research on big data analytics underscores the transformative potential of leveraging computational intelligence techniques, developing standardized frameworks, and harnessing social media data analytics to drive innovation and informed decision-making across various domains. Through the analysis of empirical data and synthesis of existing literature, several key insights have emerged.

10.8 Conclusion:

Firstly, computational intelligence techniques such as machine learning, data mining, and natural language processing have been shown to be indispensable tools in extracting valuable insights from large and complex datasets. These techniques enable organizations to uncover hidden patterns, predict future trends, and optimize decision-making processes with unprecedented accuracy and efficiency. Secondly, the development of standardized frameworks is essential to ensure consistency, reproducibility, and interoperability in the

field of big data analytics. By establishing best practices and guidelines, researchers and practitioners can navigate the evolving landscape of data analysis methodologies more effectively, thereby facilitating the adoption and implementation of advanced analytics solutions. Furthermore, the analysis of social media data offers valuable insights into user behaviour, sentiment trends, and emerging topics of discussion. By leveraging social media analytics, organizations can gain a deeper understanding of customer preferences, market trends, and public opinion, enabling them to make more informed decisions and drive strategic initiatives with greater precision and agility. Overall, the findings of the research on big data analytics underscore the critical role of advanced analytics techniques, standardized frameworks, and social media data analytics in shaping the future of decision-making and innovation in today's data-driven world. By embracing these insights and leveraging the power of big data, organizations can unlock new opportunities for growth, efficiency, and competitive advantage in an increasingly complex and dynamic business environment.

10.9 References:

1. Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big data analytics: Computational intelligence techniques and application areas. *Technological Forecasting and Social Change*, 153, 119253.
2. Mohamed, A., Najafabadi, M. K., Wah, Y. B., Zaman, E. A. K., & Maskat, R. (2020). The state of the art and taxonomy of big data analytics: view from new big data framework. *Artificial Intelligence Review*, 53, 989-1037.
3. Abkenar, S. B., Kashani, M. H., Mahdipour, E., & Jameii, S. M. (2021). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and informatics*, 57, 101517.
4. Betty Jane, J., & Ganesh, E. N. (2020). Big data and internet of things for smart data analytics using machine learning techniques. In *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019)* (pp. 213-223). Springer International Publishing.
5. Naqvi, R., Soomro, T. R., Alzoubi, H. M., Ghazal, T. M., & Alshurideh, M. T. (2021, May). The nexus between big data and decision-making: A study of big data techniques and technologies. In *The international conference on artificial intelligence and computer vision* (pp. 838-853). Cham: Springer International Publishing.
6. Arribas-Bel, D., Green, M., Rowe, F., & Singleton, A. (2021). Open data products-A framework for creating valuable analysis ready data. *Journal of Geographical Systems*, 23, 497-514.
7. Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*, 7(1), 53.
8. Bharadiya, J. P. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. *American Journal of Artificial Intelligence*, 7(1), 24.
9. Liu, Y., Soroka, A., Han, L., Jian, J., & Tang, M. (2020). Cloud-based big data analytics for customer insight-driven design innovation in SMEs. *International Journal of Information Management*, 51, 102034.
10. Zhang, J. Z., Srivastava, P. R., Sharma, D., & Eachempati, P. (2021). Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications*, 184, 1155