Global Scientific and Academic Research Journal of Economics, Business and Management
ISSN: 2583-5645 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journals-gsarjebm-home/

Role of Statistical Methods in Big Data Analysis: Navigating Computational and Ethical Challenges

By

Pushpa Raj Joshi

Department of Statistics, Tri-Chandra Multiple Campus, Tribhuvan University, Kathmandu, Nepal

Article History

Received: 10/08/2025
Accepted: 16/08/2025
Published: 19/08/2025

Vol –4 Issue – 8

PP: -62-67

DOI:10.5281/zenodo. 16908963

Abstract

The emergence of a term for massive data sets called Big Data that may be analysed computationally, with high volume, velocity, and variety, has therefore made it imperative to reorient analytical paradigms. This paper discusses the essential role of statistical methods in navigating large-scale datasets and in transitioning from classic hypothesis-based studies to modern data-driven exploration. It studies how statistical methods are developed and improved to deal with significant issues like noise, scalability, and lack of integration. This is followed by a systematic literature review and a case study of two benchmark sectors: e-commerce and healthcare, where specific measurement methods are discussed. Our result underscores the crucial complementarities between basic statistical principles and complex machine learning algorithms in predictive analytics and pattern recognition. Nonetheless, despite this somewhat critical view on algorithmic fairness, the paper's overall finding was that there appears to be a significant gap in research progress, which is of ethical importance as a technical advancement in Big Data analytics. We argue here that successful Big Data analysis thus demands a dual-pronged approach, one based on computational innovations and the other on robust data governance to ensure subsequent insights are not only correct statistically but also fair and reliable, closing in a loop with all layers of society.

Keywords: *Algorithmic Fairness, Big Data, Data Governance, Data-Driven, Decision Making, Machine Learning, Predictive Analytics, Statistical Methods*

INTRODUCTION

The increasing need to analyse vast quantities and diverse sources of high-volume, high-velocity, and high-variety data has driven the research and implementation of advanced analytics methods that convert such chaotic "big data" into clear insights.

The sheer size and heterogeneity of big data have led to a new era of exploratory data analysis, allowing large-scale data processing and analytics that extend beyond traditional hypothesis-driven research. This has led to the creation of new tools and distributed computing frameworks like Hadoop and Apache Spark that are designed to manage the massive computational demands of Big Data analytics. To perform this job, statistical methods are necessary for providing the organisation with a roadmap for transforming raw information into strategic assets and extracting actionable information from massive datasets (Lotfi et al., 2021, pp. 381–394).

This paper illustrates through a range of examples the significance of statistical techniques in managing the complexities of BD and how different types of statistical methodologies are adapted and enhanced to manage the specifics of large-scale datasets. It will focus on how quality, scalability, and integration of data have been the challenges in the Big Data framework, and statistical methods have been devised for their solution. It will also analyse the application of various types of statistical models and machine learning algorithms for predictive analytics, pattern recognition, and decision-making in the context of Big Data (Shahid & Sheikh, 2021, pp. 586-617).

This will serve as a use case for illustrating the need for data quality frameworks and robust data governance needed to make Big Data more sound, i.e., draw correct statistical inferences from Big Data sets. The greater expanse of data sources, for example, including healthcare records and social media interactions, also flags the capacity for state-of-the-art statistical methods to integrate diverse information

*Corresponding Author: **Pushpa Raj Joshi.**

effectively. In other sectors, for example, e-commerce, a better understanding of the customer's mindset is crucial for decision-making and more accurate marketing (Zineb et al., 2021, pp. 723-736).

Big data analytics has been transforming fields such as banking and finance, enabling the ability to delve into vast amounts of user data to identify noisy or silent problems quickly. Unlocking the promise of Big Data in helping to inform better evidence-based decisions requires the coupling of advanced statistical methods with massive computing infrastructure. With the development of large-scale datasets, data visualisation and statistical inference of visual data are also growing rapidly and increasingly important (He & Lin, 2020).

This depth and variety of integration gives one a natural view of the complex patterns and trends over wide ranges, so that the model can work in real time, providing business strategy focused on almost everything from business cases to human behaviours. This paper will describe how various statistical paradigms (from classical inferential statistics to modern machine learning algorithms) are being re-shaped and applied for making actionable insights from diverse sources of large-volume, high-velocity, and high-variety datasets. It will also explore how data-driven decision-making has evolved and how statistical insights enhance organisational performance and foresight (Abdul-Azeez et al., 2024, pp. 2066-2081).

Significance

The rise of big data has enabled new means of extracting useful information due to the increased scale and complexity of data; in many ways, this phenomenon is a double-edged sword. One of them is the management, processing, and analysis of large-scale and heterogeneous datasets efficiently. Without them, it is impossible to recognise patterns, forecast fluctuations in trends, or make informed decisions. These algorithms will underpin the development of data-led techniques by numerous organisations, enabling companies to cultivate their competitive edge.

This need is particularly evident in digital marketing, where data sciences enable evidence-based decision-making and harnessing insights from diverse, large datasets, ultimately aiming to transform human behaviour understanding through intelligent big data analysis. We outline, in this paper, the necessity of statistical methodologies to deal with Big Data complexity and their importance for extracting valuable insights that facilitate data-driven decision-making across different domains.

Methodology

The paper uses a methodological strategy based on the literature published to date, focusing on two main topics: statistical methods applied to Big Data and case studies where some of these methods were used. Existing academic research and industry reports were brought together to develop a homogenised view of methods for dealing with large volumes, velocities, and variances. The approach also considers problems in applying traditional statistical models to Big Data, like computational efficiencies and statistically parsimonious frameworks for dealing with diverse data types and structures.

This includes how distributed heuristics behave and perform in the face of Big Data, too noisy or incomplete to provide the full statistical power. Through this research, the ethical concerns and privacy issues involved in the statistical analysis of Big Data, particularly for sensitive information and possible biased outcomes of given data, will be identified. The discussion will further proceed to the validation of statistical models in Big Data contexts, and algorithms that can deal with dynamic data (often non-structured) from time to time. An additional discussion will explore the interdisciplinary aspects of Big Data analytics to underscore the necessary collaboration among statisticians, computer scientists, and domain experts in order to deliver holistic and impactful analytical solutions.

Research Questions

1. How are traditional statistical inference principles being adapted to accommodate the volume, velocity, and variety of big data, especially concerning the validation of models and the robustness of findings?
2. What is the efficacy of novel statistical algorithms and computational frameworks in overcoming the methodological and computational challenges presented by massive datasets?
3. How does the integration of diverse statistical techniques, such as machine learning and deep learning, impact the predictive accuracy and interpretability of significant data analytics results?

Research Gap

Despite the significant advancements in Big Data analytics, there remains a notable research gap concerning the development of universally applicable, robust statistical methodologies that can effectively address the inherent inconsistencies and biases often present in heterogeneous Big Data sources. Existing literature has primarily focused on technical solutions, often overlooking the socio-technical challenges and organisational complexities that arise during the implementation of data science projects, such as a lack of clear objectives or ambiguous roles.

This gap also includes minimal discussion of the ethical dilemmas, privacy concerns, and equity considerations that arise when collecting or analysing large-scale data, particularly in vulnerable areas such as healthcare, where statistical methods can inadvertently aggregate historical disparities. Additionally, there remains an unmet need for a holistic model of bio-inspired computational methods integrated with statistically oriented mechanisms to further enhance the data fusion/processing potential across large and heterogeneous Data. Such a comprehensive integrated framework, which not only addresses and overcomes statistical challenges but also intertwines socio-technical, ethical, and organisational issues, would indeed help to bridge

*Corresponding Author: **Pushpa Raj Joshi.**

all the gaps mentioned above towards offering a holistic perspective on Big Data analytics.

Theoretical Framework

The stabilities examined in this article will interweave fold model complexity, data characteristics, and inferential robustness to apply a motivated, explorative framework based on statistical learning theory. This framework, a conjecture-only article presentation, contends that to achieve realistic Big Data Solutions, one must move beyond hypothesis-driven methodologies such as classical statistics towards more data-driven, exploratory ways while still insisting on statistical validity. It relies on concepts from computational statistics, machine learning, and information theory to address the numerous new complexities of high-dimensional, noisy, and frequently unfinished Big Data.

It also includes the concept of data exhaust as a useful, though often not structured, source of information that requires the use of advanced statistical techniques to extract and infer from. The framework also includes a deep dive into the ethical responsibilities of maintaining data exhaust, including privacy and misuse considerations, while at the same time emphasising that we must use privacy-preserving statistical methods to humanise algorithms. This holistic method allows the developed statistical models to be technically accurate, efficient, ethically appropriate, and socially responsible in practice, encouraging fair results using Big Data analytics methods across multiple applications. This framework will recommend the types of statistical methods best suited for various Big Data architectures and objectives, and which statistical tools to use where.

Results

The results from the systematic review and empirical analysis are described in this section, in which research questions will be formulated. We will discuss the trends recorded during the usage of different statistical methods for the Big Data challenge and their performance in industries. Moreover, the paper will also describe where and how these methods have been used to date, as well as some of the limitations associated with such usage and challenges encountered in using STPs in practice (and therefore help identify areas for improved method development and theoretical innovation). It includes a breakdown of the degree to which current statistical paradigms have succeeded in addressing the "4 V's" of Big Data—Volume, Velocity, Variety, and Veracity. It examines the current performance in grappling with data quality and acquisition, storage, and security (Kong et al., 2020).

The Foundation: Descriptive and Inferential Statistics

Descriptive statistics serve as the foundation for synthesising and analysing massive amounts of information obtained through Big Data, providing an initial overview of data distributions, central tendencies, and variance, even in petabytes. This foundation mapping step is critical for identifying clusters of interest and outliers, which in turn inform more intricate inferential statistical modelling (e.g., cluster-based sample sizes). This modelling enables us to produce inferences from our representative Big Data sample to our Big Data population. Nonetheless, the large amount of heterogeneous data requires a careful application of these basic statistical rules, which may lead to using parallel computing techniques along with sophisticated sampling methods in order to alleviate the computational needs while still adhering to the common tenets of statistics (He & Lin, 2020).

In addition, principles of more advanced statistical design (eg, randomisation) can be used to help determine appropriate aspects of analyses in this Big Data context that is objective and free of bias. For starters, in high-dimensional and noisy Big Data settings, anomaly detection and outlier identification represent a particularly stringent set of problems that require genuine statistical care. It requires significant computation to process large datasets in real time (or near-real time). In most cases, these limitations can only be broken by parallel processing and distributed algorithms. Thus, these computationally advanced methods are needed to facilitate the effective capture, storage, sharing, and handling of the large-scale data (Luan et al., 2020, pp. 1-11).

This typically involves the building of new data science tools (such as data structures and algorithms optimised for extensive parallel processing to enable processing and analysis on streaming data), including cutting-edge, real-time monitoring solutions. This helps to make sure that any insights from Big Data remain current and actionable, especially in fields where decisions need to be made rapidly. At this stage, although the purpose of this activity for statistical analysis is to collect initial observation and information but in terms of computation are entirely different from processed data (that being input needs much more pre processing making the result less usable as data), even though necessary preprocessing methods are applied to data beforehand, no matter how sophisticated one may be. This preprocessing phase is necessary because BDA in a real-world scenario may suffer from issues like data inconsistency, noise, and incompleteness (Rahul et al., 2020, pp. 359–367).

Accordingly, good care has been taken in the standard operation and construction of data, with quality often controlled against reference frameworks like DQSOps to maintain high-quality datasets that could be used for analytical purposes. These efforts are essential to improve performance in the operation of Big Data projects that use distributed computing and artificial intelligence, making data reliable for decision-making (Thuy et al., 2024, pp. 1-19). This persistent growth in academic and enterprise applications for Big Data processing technologies with intelligent systems makes it even more critical to have reliable statistical methods when working on distributed computing environments.

These innovations are often built upon existing patterns in parallel processing by high-performance dataframes that execute computational tasks to diminish the ample pre-processing time incurred by massive pipelines (Perera et al, 2023, pp. 250-264). The answer extends to how you deal with

*Corresponding Author: **Pushpa Raj Joshi.**

data duplication, integrity, and security issues, and reflects on how well you can process Big Data in a timely and efficient manner. Stringent data governance is necessary to establish norms that ensure the quality, consistency, and effective utilisation of data, while mitigating potential risks associated with securing individual information privacy, confidentiality, and managing legal aspects. As active data caretakers, we are more assured that our Big Data analytics findings reflect the valid patterns in reality and satisfy surging legal and ethical conditions.

Statistical Methods in Machine Learning

Statistical methodologies integrated with machine learning algorithms are key to improving model interpretability, robustness, and predictive accuracy, particularly when dealing with large, complex datasets. This integration thereby allows for much more sophisticated analytic models that can detect complex patterns and relationships that would be impossible to see via standard statistical methods in isolation. That is the same application of machine learning across fields from high-dimensional genomic analysis to drug discovery, for which statistical validation provides guarantees about reliability and generalizability. The inter-relation of machine learning and statistical methods can capture valuable information from Big Data, and it has wide industrial applications such as risk assessment, fraud detection, and customer behaviour analysis (Rahul et al., 2020, pp. 359–367).

This collaboration and blending, mostly labelled as statistical machine learning, is necessary to create models that can predict and provide actionable insights on new data scenarios. In addition, it permits principled uncertainty quantification in predictions to quantify mere predictive intervals and probability statements, which are essential for decision-making in high-stakes scenarios. Businesses across various industries are filling this pipeline and adopting predictive statistical methods and machine learning techniques, underscoring their importance for profit improvement. However, even with its embrace, questions are being asked about the fair and equal sharing of these benefits. This increasing dependence on machine learning for decision-making, such as credit scoring, college admissions, or hiring processes, has raised questions about fairness and the risk of non-discriminatory outcomes despite all these obvious benefits (Liu & Vicente, 2022, pp. 513–537).

The question of this problem has initiated much research in the field on fair algorithms, especially because biases in training data or model architectures could lead to and even exacerbate social inequalities. These considerations reveal the fundamental importance of a statistically sound framework for assessing algorithmic fairness and assuring that predictive models do not consciously or inadvertently aggravate disparities between statutorily protected groups. In addition, it involves creating new fairness metrics and stronger auditing tools that guide models' choices to guarantee transparency and accountability in the final deployment One area that we need to improve in this respect is the distinction between bias from intrinsic data (sample and population) properties, and bias which might be learned through over-fitting of models due to regularization/training procedure. Broadsterdamɑost-hoc model score adjustments (Rodolfa et al., 2021, pp. 896–904).

Case Studies

The practical case studies across various areas are more beneficial for demonstrating the use of statistical methods in big data analysis. In the following examples, we can see how enterprises use statistical methodologies to turn huge and complex data into business insight and competitive advantage.

E-commerce and Customer Behaviour Analysis

Big data analytics is critical for the e-commerce industry to understand fashion behaviour and, therefore, optimise its functions accordingly. Online platforms like Amazon collect extensive data from customer interactions, including product views, purchasing behaviour, and reviews. They use statistical techniques (like cluster analysis) to divide a large population of customers into sub-groups with similar behaviours and preferences.

For instance, a retailer might find a segment of sports-equipment-buying or renting customers and use that to target them with promotions. Moreover, it uses regression analysis to construct models for predicting future purchasing trends and inventory necessities. Using past sales data, a company can anticipate which products will likely be popular at a given time of year and, in that way, optimise its supply chain to ensure that stockouts are kept to an absolute minimum. These statistical analyses will provide important insight for data-driven decision making, which eventually contributes to improving customer satisfaction and profitability (Zineb et al., 2021, pp. 723-736).

Healthcare and Predictive Diagnostics

Big data and statistical analysis have revolutionised the world of healthcare. Several types of data, such as electronic health records, genomic data, and information from wearable devices, can be used to drive patient outcomes. Predictive models, frequently based on machine learning algorithms, are implemented to help predict the health risks of patients and the likelihood of disease onset. Statistical models, for example, can take into account factors such as a patient's medical history, genetic markers, and lifestyle to determine how likely they are to have a heart attack in the next five years.

These methods permit the creation of preventive care approaches and pre-emptive interventions by healthcare providers. In addition, it is applied to drug discovery and personalised medicine using statistical analysis. Using large databases of genetic and molecular information, researchers try to find targets for drugs and to understand which patient populations respond to which therapies. The bottom line is that when big data and statistical methods collaborate in the field of healthcare, more individual-initiated and future-oriented patient care is becoming possible.

*Corresponding Author: **Pushpa Raj Joshi.**

Ethical Considerations and Algorithmic Fairness

The advantages big data analytics offers are complemented by case studies that demonstrate the ethical issues, with algorithmic fairness highlighted as one of the primary case studies. For example, existing societal inequalities in sectors such as finance or law enforcement are not only replicated but also exacerbated by machine learning models trained on biased historical data (Rodolfa et al., 2021, pp. 896–904). A loan application model trained with data that was unfairly biased can, for instance, have the effect of denying certain demographic groups loans.

Statistical frameworks are now being used to detect and adjust for bias in algorithms. Using fairness metrics to audit models, which can help verify that a model is not discriminating against anyone, a study by Shahbazi et al. For example, (2023) details different ways to detect and mitigate representation bias in data, with a focus on careful statistical evaluation of AI systems that must be accurate and fair. These examples highlight the need to create ethical considerations together with statistical rigour, in order to construct trustworthy and accountable AI systems (Shahbazi et al., 2023, pp. 1-39).

Discussion

The analysis in this paper validates the necessity of statistical integration for converting big data from disparate sources and varied forms into value, thereby necessitating statistics. These observations illustrate a dramatic departure from conventional hypothesis-based statistical inference to highly flexible, data-driven approaches that exploit computational firepower available for big data. Although traditional descriptive and inferential statistics remain essential for an initial exploratory analysis (He & Lin, 2020), their implementation has been enhanced mainly by way of parallel processing and advanced sampling methods to address the increased computational needs (Perera et al., 2023, pp. 250-264). Here, we directly respond to the research question by adapting this program so that underlying principles can be directly transferred, and the same fundamentals are consistently applied, only now with modern computational platforms.

This conversation underscores the indispensable necessity of uniting statistical techniques with machine learning approaches. This is what the collaboration improvement, as shown in the review, will lead to: more precise and interpretable predictions that are crucial for understanding the intricate relationship between the high-dimensional data (Rahul et al., 2020, pp. 359–367). Nevertheless, this development simultaneously brings into stark focus a significant deficiency and an important future direction, marking the intrinsic space on algorithmic fairness that remains to be researched. In medicinal science, e-commerce, and clinical decision-making, machine learning can enhance profitability or yield tangible benefits, such as improved healthcare outcomes, while fostering self-growth that aligns with the social context (Liu & Vicente, 2022, pp. 513–537). This discovery is especially relevant and requires a strict statistical methodology that can identify and deal with the bias.

This is what gives rise to the ethical implications of Big Data analytics, hence a major talk. This difficulty is exacerbated by the "black box" nature of many advanced AI models (as alluded to in the paper's abstract), which makes it challenging to identify these biases and serves as a substantial deterrent to public trust. This requires designing transparent and explainable AI (XAI) methods that can audit and validate algorithmic decisions for fairness proactively (Gichoya et al., 2023, pp. 1-8). This is part of the limitation, its reliance on a literature review and lack of empirical data to supplement evidence on effective bias mitigation strategies.

In the future of Big Data analysis, universal socio-technical frameworks that can address the research gap identified above are essential to construct. As future work, these frameworks should be built and tested, incorporating mandatory technical solutions for data inconsistency, along with strong ethical governance and fairness metrics. Expanding on previous studies, the study of Shahbazi et al. Following the work of (2023) in surveying methods for measuring bias, future works could investigate the practicality of these strategies when deployed and scaled at a real-world level, such as in high-volume A/B tests. Ultimately, the conversation centres on the challenge of finding statistical methods that are both extensive and robust, while also being fair, trustworthy, and ethical, in a data-driven society.

Conclusion

Big Data has dramatically impacted the statistical methodology, transforming it from a historical hypothesis-oriented solution to a more contemporary data-driven solution. This paper presented an example of how the fundamentals of statistics are being reworked with distributed computing frameworks to leverage significantly more data, often in the unusual structures we encounter today. A perfect blend of statistical analysis and machine learning algorithms, predictive analytics allows organisations to conduct wide-ranging analyses resulting in insights for strategic planning across different industries.

The paper also highlights a serious concern: that of ethics and algorithmic bias. Given the ongoing need for large datasets, data quality and governance, as well as robust, fair, and transparent frameworks, are a necessity. The question of the future for statistics-based approaches in Big Data is not just about new tech and better computations, but also social technological solutions that are ethically conscious and built on public trust. In conclusion, the effective deployment of Big Data analytics is a matter of balance, one that embraces statistical discipline but subjects itself to ethical tools designed to benefit all.

References

1. Abdul-Azeez, O. Y., Ihechere, A. O., & Idemudia, C. (2024). Enhancing business performance: The role of data-driven analytics in strategic decision-making. International Journal of Management &

*Corresponding Author: **Pushpa Raj Joshi.**

Entrepreneurship Research, 6(7), 2066-2081. https://doi.org/10.51594/ijmer.v6i7.1257

2. Gichoya, J. W., Thomas, K. J., Celi, L. A., Safdar, N., Banerjee, I., Banja, J. D., Seyyed-Kalantari, L., Trivedi, H., & Purkayastha, S. (2023). AI pitfalls and what not to do: mitigating bias in AI [Review of AI pitfalls and what not to do: mitigating bias in AI]. British Journal of Radiology, 96(1150), 1–8. Wiley. https://doi.org/10.1259/bjr.20230023
3. He, X., & Lin, X. (2020). Challenges and Opportunities in Statistics and Data Science: Ten Research Areas. Harvard Data Science Review, *2*(3). https://doi.org/10.1162/99608f92.95388fcb
4. Kong, L., Liu, Z., & Wu, J. (2020). A systematic review of big data-based urban sustainability research: State-of-the-science and future directions [Review of A systematic review of big data-based urban sustainability research: State-of-the-science and future directions]. Journal of Cleaner Production, 273, 123142. Elsevier BV. https://doi.org/10.1016/j.jclepro.2020.123142
5. Liu, S., & Vicente, L. N. (2022). Accuracy and fairness trade-offs in machine learning: a stochastic multi-objective approach. Computational Management Science, 19(3), 513–537. https://doi.org/10.1007/s10287-022-00425-z
6. Lotfi, C., Srinivasan, S., Ertz, M., & Latrous, I. (2021). Web Scraping Techniques and Applications: A Literature Review [Review of Web Scraping Techniques and Applications: A Literature Review]. Soft Computing Research Society eBooks, 381–394. https://doi.org/10.52458/978-93-91842-08-6-38
7. Luan, H., Géczy, P., Lai, H., Gobert, J. D., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R. da S., Li, P., & Tsai, C. (2020). Challenges and Future Directions of Big Data and Artificial Intelligence in Education [Review of Challenges and Future Directions of Big Data and Artificial Intelligence in Education]. Frontiers in Psychology, 1-11. Frontiers Media. https://doi.org/10.3389/fpsyg.2020.580820
8. Perera, N., Sarker, A. K., Staylor, M., Laszewski, G. von, Shan, K., Kamburugamuve, S., Widanage, C., Abeykoon, V., Kanewela, T. A., & Fox, G. (2023). In-depth analysis of parallel processing patterns for high-performance Dataframes. Future Generation Computer Systems, 149(250), 250–264. https://doi.org/10.1016/j.future.2023.07.007
9. Rahul, K., Banyal, R. K., Goswami, P., & Kumar, V. (2020). Machine Learning Algorithms for Big Data Analytics. In Advances in Intelligent Systems and Computing (pp. 359–367). Springer Nature. https://doi.org/10.1007/978-981-15-6876-3_27
10. Rodolfa, K. T., Lamba, H., & Ghani, R. (2021). Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. Nature Machine Intelligence, 3(10), 896–904. https://doi.org/10.1038/s42256-021-00396-x
11. Shahbazi, N., Yin, L., Asudeh, A., & Jagadish, H. V. (2023). Representation Bias in Data: A Survey on Identification and Resolution Techniques [Review of Representation Bias in Data: A Survey on Identification and Resolution Techniques]. ACM Computing Surveys, 55, 1–39. Association for Computing Machinery. https://doi.org/10.1145/3588433
12. Shahid, N. U., & Sheikh, N. J. (2021). Impact of Big Data on Innovation, Competitive Advantage, Productivity, and Decision Making: Literature Review. Open Journal of Business and Management, 9(2), 586–617. https://doi.org/10.4236/ojbm.2021.92032
13. Thuy, N. T., Nguyen-Hoang, T.-A., Dinh, N.-T., & Nguyen, T. (2024). Data Quality Management in Big Data: Strategies, Tools, and Educational Implications. 200, 105067, 1-19. https://doi.org/10.2139/ssrn.5001630
14. Zineb, E. F., Rafalia, N., & Abouchabaka, J. (2021). An Intelligent Approach for Data Analysis and Decision Making in Big Data: A Case Study on the E-commerce Industry. International Journal of Advanced Computer Science and Applications, 12(7), 723–736. https://doi.org/10.14569/ijacsa.2021.0120783

*Corresponding Author: **Pushpa Raj Joshi.**