International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5978~5985
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5978-5985

On big data predictive analytics-trends, perspectives, and challenges

Yassine Benlachmi, Abdelaziz El Yazidi, Abdallah Rhattoy, Moulay Lahcen Hasnaoui
ISIC-TEAM, L2ISEI-Laboratory-ESTM, Moulay Ismail University of Meknes, Meknes, Morocco

Article Info

Article history:

Received Jun 10, 2023
Revised Jul 25, 2025
Accepted Sep 16, 2025

Keywords:

Big data analytics
Data mining
Educational big data
Predictive analytics
Predictive models

ABSTRACT

The world is experiencing explosive growth in numerous sectors such as healthcare, engineering, scientific studies, business, social networking. This growth is causing an immense surge in data generation too. And with the emergence of technologies like internet of things (IoT), Mobile, and cloud computing, the volume of data being produced is skyrocketing. However, making sense of this colossal amount of data is a daunting challenge. Enter big data computing, a new paradigm that blends large datasets with advanced analytical techniques. Big data is characterized by the three V's: Volume, velocity, and variety, and refers to massive datasets. By processing this data, we can uncover new opportunities and gain valuable insights into market trends. Traditional techniques are simply not equipped to handle the scale of Big Data. The purpose of this article is to gather reviews of various predictive analytics applications related to big data and the advantages of using big data analytics across various decision-making domains.

Corresponding Author:

Yassine Benlachmu
ISIC-TEAM, L2ISEI-Laboratory-ESTM, Moulay Ismail University of Meknes
Meknes, Morocco
Email: yassin040@gmail.com

1. INTRODUCTION

Today's digital world creates massive data—some easy to find (like weather or stock info), some needing active collection (like social media). Data science helps reveal hidden insights. Despite progress, big challenges remain. This study uses bibliometric methods to explore how big data and predictive analytics can work across industries, offering four key contributions.

In today's digital world, huge amounts of data are generated—some easily accessible like sports or weather data, others like social media need collection. Data science helps uncover hidden insights. Despite existing research, major challenges remain. This study uses bibliometric methods to explore cross-sector integration of big data and predictive analytics, offering four key contributions:

a. Examine big data's fundamental principles and complexities, encompassing its core components, defining characteristics, and the latest technological advancements that enable its existence.
b. Explore real-world examples of the successful implementation of big data predictive analysis in various industries such as healthcare, finance, and marketing.
c. Deepen our understanding about potential challenges and suggest ways to overcome them.

We based this study entirely on existing materials like books, academic journals, and Google Scholar articles. Our goal was to better understand the topic and provide useful insights for future research. The paper walks through our approach in section 2 and introduces a fresh way to think about big data—the 'sunflower model' as shown in Figure 1—designed to reflect today's tech reality.

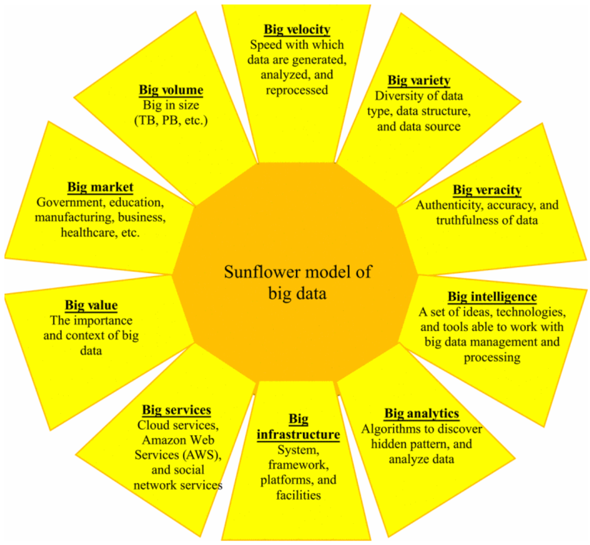

Figure 1. Utilizing the sunflower model as an analogy to explain the idea of big data [1]

2. METHOD

To ensure trustworthy sources, we combined two review approaches: systematic literature review (SLR) and narrative review. We focused on topics like "big data," "predictive analytics," and related areas such as analytics performance, knowledge management, and IoT. Our process involved thoroughly reviewing relevant studies in a clear and structured way: i) Investigate various combinations of relevant keywords; ii) Identify and differentiate articles that contain pertinent keyword phrases in both the title and body of the document; iii) Exclude articles that may contain relevant keywords but are not substantially related to the field of big data or predictive analytics; and iv) Aggregate relevant research papers. Our pursuit for relevant literature was extensive and exhaustive, encompassing a thorough exploration of several digital databases, such as: ScienceDirect (*https://www.sciencedirect.com/*), ResearchGate (*https://www.researchgate.net/*), and IEEEXplore (*https://ieeexplore.ieee.org/*). We carefully filtered the collected articles, removing any that were not relevant—like editorials, book reviews, or introductory papers that did not focus on big data or predictive analytics. We used the PRISMA framework, a well-known method for systematic reviews, to identify the most suitable studies for our research. Figure 2 illustrates the PRISMA model in action.

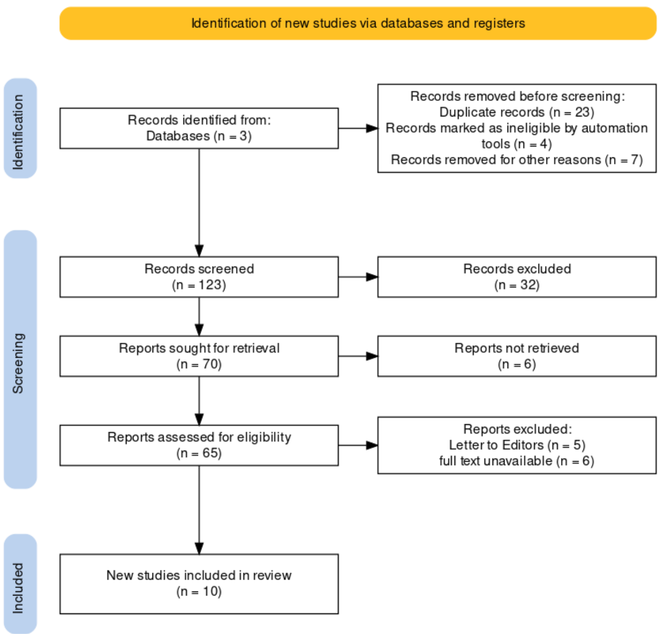

Figure 2. PRISMA model

3. BIG DATA AND PREDICTIVE ANALYSIS – CORE CONCEPTS

3.1. From data to big data

Data comes in two main types: structured (like census stats, financial records, or surveys) and unstructured (like web content or XML files). Only about 5% of data is structured [2], while the rest is messy and hard to process—what we now call big data. Big data is known for its 3Vs: volume, velocity, and variety—first described by Laney in 2001 [3], [4]. Later, veracity and value were added to form the 5V model [4], [5]. Some experts are now exploring ideas beyond the 5Vs [6]. The Bigs framework expands the 5Vs of big data by adding technology (intelligence, analytics, infrastructure) and socio-economic aspects (service, value, market). It captures data scale, speed, diversity, trust, and highlights big data's impact on innovation and growth.

3.2. Big data analytics

Big data analytics involves curating and analyzing large, complex data sets to discover patterns and insights [7]. It helps organizations make better decisions by focusing on the most relevant data, driving impactful outcomes. Table 1 compares conventional data analysis with big data analysis.

Table 1. Conventional data analytics vs big data analytics

Character	Conventional approach	Big data approach
Analysis method	Hypothesis-based	Machine learning
Primary goal	Performance management and internal decision support	data-driven products and business processes driver
Data type	Structured and defined (formatted in columns & rows)	Unstructured, semi structured
Volume	10's of TB or less	100 terabytes to petabytes

3.3. Predictive analytics

Predictive analytics forecasts future outcomes and uncovers opportunities and risks using techniques like data mining, modeling, and machine learning. It helps organizations gain a competitive edge across industries [8]. Unlike other traditional BI methodologies that focus on analyzing past events to make sense of the present, predictive analytics is future-facing and provides organizations with the ability to make proactive decisions, gain valuable insights, and stay ahead of the competition as shown in Figure 3.

Figure 3. Of all the business intelligence disciplines, prediction offers the greatest potential for delivering business value but is also the most intricate to execute. Furthermore, each discipline builds upon the preceding one, culminating in a comprehensive, additive approach to business intelligence, rather than an exclusive one

4. TECHNICAL APPLICATIONS FOR THE BIG DATA PREDICTIVE ANALYTICS ECOSYSTEM

The potential application areas of big data predictive analytics are vast and varied, ranging from medical and financial industries to climate and agriculture. In the medical field, AI-based big data analytics models can help to predict and diagnose diseases more accurately and efficiently. For instance, the use of big data analytics can aid in predicting the onset of diseases and epidemics, as evidenced by the application of big data and artificial intelligence in epidemic surveillance and containment. In the finance sector, big data predictive analytics can be employed for cryptocurrency portfolio allocation using hybrid and predictive big data decision support systems. Additionally, big data analytics can aid in predicting climate factors, which can help to inform agricultural disaster management strategies. Big data predictive analytics can also be utilized in the fashion and apparel industry through social media user behavior analysis. Educational institutions can leverage big data analytics to make informed predictions and applications for student success. Finally, big data analytics has promising applications in power systems by optimizing energy consumption and reducing costs.

4.1. Predictive analytics in medical big data

Big data research is vital in healthcare due to the massive data generated annually, now at 0.115 billion terabytes with 2 exabytes growth per year [9]. This enables large-scale population health analysis to improve decision-making. One study analyzed 356,507 patients' data from 1982-2010, revealing consistent patterns useful for predictive modeling [10]. Wearable smartwatches collect physiological data like pulse and oxygen levels, which are analyzed for medical diagnosis and warnings [11]. Big data and AI have also played key roles in managing coronavirus disease (COVID-19) by providing insights and guiding strategies [12]. Additionally, big data aids child and adolescent mental health research [13] and helps predict health risks from pesticide exposure and air quality in China [14], [15].

4.2. Predictive analytics in educational big data

Big data is widely viewed as useful for aiding educational reforms at various levels of teaching and administration, and even research management [16]. It is split into two types of analytical orientations: one is data-based analysis which focuses on resources such as enrollment and curriculum data, while the other is demand-based analysis which ties students to the educational quality indicators as illustrated in Figure 4.

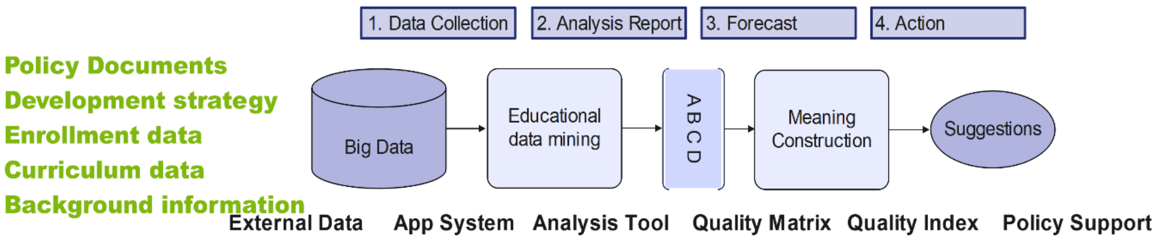

Figure 4. Data-based analysis orientation

Some researchers are increasingly focused on big data in the context of higher education and its impact on the development, planning, efficiency of instruction, and the overall quality of learning. Important tools such as learning analytics and educational data mining serve for forecasting, classification, and clustering [17]. The MASTER tool, aimed at predicting students' adaptation issues, has implemented SMOTE and priority forest resource allocation techniques to streamline resource allocation [18]. Other studies look at the role of big data in online education, especially during the COVID-19 pandemic, focusing on the innovation and data-driven approaches to teaching [19].

4.3. Predictive analytics in social media big data

Social media platforms generate vast unstructured data valuable for predictive analysis [20], [21]. Studies use this data to improve marketing in fashion [22], predict healthcare outcomes [23], and forecast tourism with machine learning [24]. The BD-SMAB model aids competitor analysis via social feedback [25], while further research explores social media's business and academic uses [26]–[28].

4.4. Predictive analytics in economic big data

Big data boosts resource allocation, production efficiency, and sustainable development, with AI, cloud computing, IoT, and blockchain playing key roles [29]. The DHM-BDA model improves disaster management via better prediction and risk reduction [30]. Big data also transforms digital economics and guides macroeconomic strategies [31], while enhancing sales management through data-driven control and predictive models to foster business growth [32].

5. OPEN CHALLENGES TO THE METAVERSE IMPLEMENTATION AT A LARGE SCALE

Big data predictive analysis can revolutionize healthcare, finance, education, and more by offering better insights and predictions. However, challenges like data quality, privacy, and the need for skilled experts must be addressed [33]. These challenges fall into three categories: ensuring valid data through processing, aggregating diverse data sources, and overcoming data availability issues. Figure 5 summarizes these challenges in big data predictive analysis.

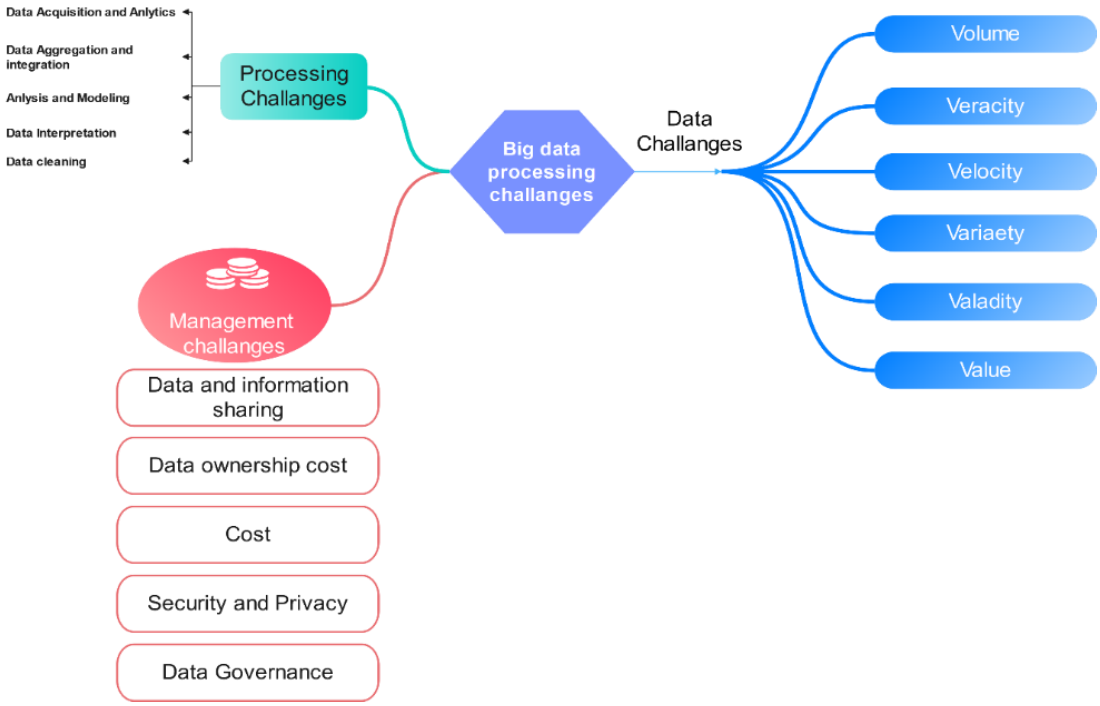

Figure 5. Difficulties encountered in the analysis of large datasets

5.1. Analyzing un-structured data

About 95% of big data is made up of unstructured data. Semi-structured data is a type of unstructured data that does not follow strict formatting rules. A good example of semi-structured data is XML, a language used to exchange data over the web. XML documents contain tags that can be read by machines and are defined by the user. So to address the unstructured data is a big issue in big data predictive analysis.

5.2. Maintaining data quality

The quality of the data used for analysis is crucial to the success of big data predictive analysis [34]. Inaccurate or incomplete data can lead to incorrect predictions and can ultimately impact business decisions. It is important to ensure that the data is relevant, accurate, and reliable. Additionally, data cleaning and preprocessing techniques should be employed to eliminate any irrelevant or incomplete data.

5.3. Data protection and privacy

Big data incorporates various types of data, including personal information, which may lead to privacy concerns even when anonymized [35]. Therefore, privacy protection is necessary for organizations, individuals, and governments, requiring a comprehensive framework [36]. It is crucial to involve the public and other stakeholders in the conversation surrounding data security and privacy to address concerns and avoid excessively restrictive regulations.

5.4. Data integration

One of the primary challenges in data integration is the issue of data heterogeneity, which refers to the differences in data formats, structures, and semantics. The integration of such diverse data sources may lead to data inconsistency, loss of information, and inaccuracies in analysis results. Therefore, data integration requires sophisticated tools and techniques, including data preprocessing, data mapping, and data transformation. To address the challenges of data integration in big data analysis, several studies have proposed different approaches and frameworks. For instance Hammad *et al.* [37] proposed a framework that utilizes semantic technologies to integrate heterogeneous data from multiple sources.

5.5. Data salability

Scalability is a crucial challenge in big data predictive analysis. As the volume, velocity, and variety of data increase, the computational resources needed to process and analyze it also increase. The traditional computing resources may not be sufficient to handle the scale of data, which can lead to performance issues, such as slow processing times, delays, and system crashes. Therefore, it is essential to design and implement a scalable architecture that can handle the increasing data volume while maintaining optimal performance.

To address the scalability challenge in big data analytics, researchers and industry professionals are exploring various techniques and technologies, such as distributed computing, cloud computing, and parallel processing [38]. These techniques enable data processing and analysis to be performed simultaneously on multiple machines, thereby significantly reducing processing time and improving system performance. However, implementing these techniques requires careful consideration of the system's design, data distribution, and processing algorithms.

6. CONCLUSION

The potential for utilizing big data predictive analysis in governance is extensive, encompassing a range of applications from solving traffic problems to improving healthcare, managing supply chains, protecting the environment, customizing education, and enhancing security. As technology continues to evolve, governments have the opportunity to transform the way they operate, including citizens in decision-making processes and improving overall efficiency. Predictive analysis can help public workers be more effective with limited resources, but it should not replace intuition, local knowledge, and expertise. The most effective use of this system is to complement existing practices. Predictive analysis can also help build trust in change and encourage experimentation. However, there are challenges to be considered when implementing big data in the public sector, and more research is needed to find solutions.

REFERENCES

[1] M. S. Rahman and H. Reza, “A systematic review towards big data analytics in social media,” *Big Data Mining and Analytics*, vol. 5, no. 3, pp. 228–244, 2022, doi: 10.26599/BDMA.2022.9020009.

[2] M. M. Yousef, “Big data analytics in health care: a review paper,” *International Journal of Computer Science and Information Technology*, vol. 13, pp. 17–28, 2021.

[3] D. Laney, “3D data management: controlling data volume, velocity and variety,” *META Group Research Note*, 2001.

[4] B. Sena, A. P. Allian, and E. Y. Nakagawa, “Characterizing big data software architectures: a systematic mapping study,” *Proceedings of the 11th Brazilian Symposium on Software Components, Architectures, and Reuse, SBCARS '17*, Brazil, 2017, doi: 10.1145/3132498.3132510.

[5] M. S. Rahman and H. Reza, “Systematic mapping study of non-functional requirements in big data system,” in *2020 IEEE International Conference on Electro Information Technology (EIT)*, 2020, pp. 25–31, doi: 10.1109/EIT48999.2020.9208288.

[6] Z. Sun, K. Strang, and R. Li, “Big data with ten big characteristics,” in *Proceedings of the 2nd International Conference on Big Data Research*, 2018, pp. 56–61, doi: 10.1145/3291801.329182.

[7] N. M. Adams, “Perspectives on data mining,” *International Journal of Market Research*, vol. 52, no. 1, pp. 11–19, 2010, doi: 10.2501/S147078531020103X.

[8] K. Vassakis, E. Petrakis, and I. Kopanakis, “Big data analytics: applications, prospects and challenges,” in *Mobile Big Data: A Roadmap from Models to Technologies*, G. Skourletopoulos, G. Mastorakis, C. X. Mavromoustakis, C. Dobre, and E. Pallis, Eds. Cham: Springer International Publishing, 2018, pp. 3–20.

[9] C. Mahanty and B. K. Mishra, “Medical data analysis in eHealth care for industry perspectives: applications,” in *An Industrial IoT Approach for Pharmaceutical Industry Growth*, V. E. Balas, V. K. Solanki, and R. Kumar, Eds. Academic Press, 2020, pp. 305–335.

[10] R. Chauhan and E. Yafi, “Big data analytics for prediction modelling in healthcare databases,” in *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2021, pp. 1–5, doi: 10.1109/IMCOM51814.2021.9377403.

[11] C. Xiao-Yong, Y. Bo-Xiong, Z. Shuai, D. Jie, S. U. N. Peng, and G. A. N. Lin, “Intelligent health management based on analysis of big data collected by wearable smart watch,” *Cognitive Robotics*, vol. 3, pp. 1–7, 2023, doi: https://doi.org/10.1016/j.cogr.2022.12.003.

[12] Z. Jiao, H. Ji, J. Yan, and X. Qi, “Application of big data and artificial intelligence in epidemic surveillance and containment,” *Intelligent Medicine*, vol. 3, no. 1, pp. 36–43, 2023, doi: 10.1016/j.imed.2022.10.003.

[13] A. Wickersham and J. Downs, “Clinical applications of big data to child and adolescent mental health care,” in *Shaping the Future of Child and Adolescent Mental Health*, M. Hodes and P. J. De Vries, Eds. Academic Press, 2023, pp. 59–79.

[14] X. Wang, D. Yu, L. Ma, X. Lu, J. Song, and M. Lei, “Using big data searching and machine learning to predict human health risk probability from pesticide site soils in China,” *Journal of Environmental Management*, vol. 320, p. 115798, 2022, doi: 10.1016/j.jenvman.2022.115798.

[15] H.-C. Hsu, C.-Y. Pan, I. C. Wu, C.-C. Liu, and Z.-Y. Zhuang, “Using the big data analysis and basic information from lecture halls to predict air change rate,” *Journal of Building Engineering*, vol. 66, p. 105817, 2023, doi: 10.1016/j.jobe.2022.105817.

[16] X. Bai and Others, “Educational big data: predictions, applications and challenges,” *Big Data Research*, vol. 26, p. 100270, 2021, doi: 10.1016/j.bdr.2021.100270.

[17] M. I. Baig, L. Shuib, and E. Yadegaridehkordi, “Big data in education: a state of the art, limitations, and future research directions,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 44, 2020, doi: 10.1186/s41239-020-00223-0.

[18] T. Guo, X. Bai, S. Zhen, S. Abid, and F. Xia, “Lost at starting line: Predicting maladaptation of university freshmen based on educational big data,” *Journal of the Association for Information Science and Technology*, 2023, doi: 10.1002/asi.24718.

[19] Y. Cui *et al.*, “A survey on big data-enabled innovative online education systems during the COVID-19 pandemic,” *Journal of Innovation & Knowledge*, vol. 8, no. 1, p. 100295, Jan. 2023, doi: 10.1016/j.jik.2022.100295.

[20] C. J. Aivalis, K. Gatziolis, and A. C. Boucouvalas, “Evolving analytics for e-commerce applications: Utilizing big data and social media extensions,” in *2016 International Conference on Telecommunications and Multimedia (TEMU)*, Jul. 2016, pp. 1–6, doi: 10.1109/TEMU.2016.7551938.

[21] N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, “Social media big data analytics: A survey,” *Computers in Human Behavior*, vol. 101, pp. 417–428, Dec. 2019, doi: 10.1016/j.chb.2018.08.039.

[22] Z. Xue, Q. Li, and X. Zeng, “Social media user behavior analysis applied to the fashion and apparel industry in the big data era,” *Journal of Retailing and Consumer Services*, vol. 72, p. 103299, May 2023, doi: 10.1016/j.jretconser.2023.103299.

[23] E. Cano-Marin, M. Mora-Cantallops, and S. Sanchez-Alonso, “The power of big data analytics over fake news: a scientometric review of Twitter as a predictive system in healthcare,” *Technological Forecasting and Social Change*, vol. 190, p. 122386, May 2023, doi: 10.1016/j.techfore.2023.122386.

[24] Y. Li, Z. Lin, and S. Xiao, “Using social media big data for tourist demand forecasting: a new machine learning analytical approach,” *Journal of Digital Economy*, vol. 1, no. 1, pp. 32–43, Jun. 2022, doi: 10.1016/j.jdec.2022.08.006.

[25] H. Zhang, Z. Zang, H. Zhu, M. I. Uddin, and M. A. Amin, “Big data-assisted social media analytics for business model for business decision making system competitive analysis,” *Information Processing & Management*, vol. 59, no. 1, p. 102762, Jan. 2022, doi: 10.1016/j.ipm.2021.102762.

[26] S. Bazzaz Abkenar, M. Haghi Kashani, E. Mahdipour, and S. M. Jameii, “Big data analytics meets social media: A systematic review of techniques, open issues, and future directions,” *Telematics and Informatics*, vol. 57, p. 101517, Mar. 2021, doi: 10.1016/j.tele.2020.101517.

[27] Y.-J. Chen and Y.-M. Chen, “Forecasting corporate credit ratings using big data from social media,” *Expert Systems with Applications*, vol. 207, p. 118042, Nov. 2022, doi: 10.1016/j.eswa.2022.118042.

[28] C. Tandon, S. Revankar, H. Palivela, and S. S. Parihar, “How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics,” *International Journal of Information Management Data Insights*, vol. 1, no. 2, p. 100035, Nov. 2021, doi: 10.1016/j.jjimei.2021.100035.

[29] L. Wang, Y. Wu, Z. Huang, and Y. Wang, “How big data drives green economic development: evidence from China,” *Frontiers in Environmental Science*, vol. 10, p. 1055162, Nov. 2022, doi: 10.3389/fenvs.2022.1055162.

[30] M. M. Jaber *et al.*, “Predicting climate factors based on big data analytics based agricultural disaster management,” *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 128, p. 103243, Dec. 2022, doi: 10.1016/j.pce.2022.103243.

[31] G. Dai, “Research on digital economy information system through cloud computing and big data technology,” in *2021 IEEE International Conference on Data Science and Computer Application (ICDSCA)*, Oct. 2021, pp. 734–737, doi: 10.1109/ICDSCA53499.2021.9650173.

[32] W. Shan, “Research on refined sales management, data analysis and forecasting under big data,” in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Oct. 2020, pp. 305–308, doi: 10.1109/MLBDBI51377.2020.00065.

[33] C. Komalavalli and C. Laroiya, “Challenges in big data analytics techniques: a survey,” in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2019, pp. 223–228, doi: 10.1109/CONFLUENCE.2019.8776932.

[34] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, “Critical analysis of big data challenges and analytical methods,” *Journal of Business Research*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.001.

[35] S. Rawat, R. Tyagi, and P. Kumar, “An investigative study on challenges of live migration,” in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Sep. 2016, pp. 129–134, doi: 10.1109/ICRITO.2016.7784939.

[36] P. Pant, P. Kumar, I. Alam, and S. Rawat, “Analytical planning and implementation of big data technology working at enterprise level,” in *Advances in Intelligent Systems and Computing (AISC,volume 672)*, Springer, Singapore, 2018, pp. 1031–1043.

[37] R. Hammad, M. Barhoush, and B. H. Abed-alguni, “A semantic-based approach for managing healthcare big data: a survey,” *Journal of Healthcare Engineering*, vol. 2020, pp. 1–12, Nov. 2020, doi: 10.1155/2020/8865808.

[38] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, “Spark: cluster computing with working sets,” in *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, 2010, p. 10.

BIOGRAPHIES OF AUTHORS

Yassine Benlachmi master's degree in systems networks, he's currently a Ph.D. student in the big data predictive analysis since 2018 at ISIC-TEAM, L2ISEI-Laboratory-ESTM Moulay Ismail University of Meknes. He can be contacted at email: yassin040@mail.com.

Abdelaziz El Yazidi specializing in networks and telecommunications, doctoral student in big data analytics since 2019 at ISIC-TEAM, L2ISEI-Laboratory-ESTM Moulay Ismail University of Meknes. He can be contacted at email: elyazidi.abdelaziz@gmail.com and abd.elyazidi@edu.umi.ac.ma.

Abdallah Rhattoy Department of Computer Science, High School of Technology, ISIC Team, Moulay-Ismail University, Meknes, Morocco. He can be contacted at email: rhattoy@gmail.com.

Moulay Lahcen Hasnaoui received his Ph.D. Microelectronics and Solid State Electronics at Paris-Sud University, France (1995). He worked as research associate Department of Engineering; Polytechnic School, Montreal, Canada (1996-1998). He earned his bachelor's degree in computer science from University of Montreal, Canada (1998-2002). Self-employed as a software developer (2002-2004). He worked as research assistant professor at Computer Sciences Department at the Faculty of Sciences Dhar Al Mahraz, Fez (2011-2016). Since 1016 to present he is working in as a research professor at Moulay Ismail University of Meknès. The actual area of his research is semiconductor device full cell, WSN, VANET, IoT and big data analysis. He can be contacted at email: myhasnaoui@gmail.com.