

# Big Data Mining

Subjects: [Computer Science, Theory & Methods](#) | [Computer Science, Information Systems](#) | [Business](#)

Contributor: Bartłomiej Hadasik

**Big data mining** (BDM) is an approach that uses the cumulative data mining or extraction techniques on large datasets / volumes of data. It is mainly focused on retrieving relevant and demanded information (or patterns) and thus extracting value hidden in data of an immense volume. BDM draws from the conventional data mining notation but also combines the aspects of big data, i.e. it enables to acquire useful information from databases or data streams that are huge in terms of “big data V’s”, like volume, velocity, and variety.

big data

big data mining

customer insights

## 1. Big data as an overriding notation

In a world overflowing with information, which is particularly evident in information societies with access to the Internet, the term “big data” is indispensable and interconnected. Nowadays, data are being sent to the global network not only by people who do it consciously and manually (e.g., via social networks or e-mails), but also by all kinds of sensors and with the use of cloud computing. It is a Web 3.0 domain of which big data is one of the main pillars.<sup>[1]</sup>

Big data, as a nascent concept, has a rather turbulent history of trying to define it, and an attempt to organize the definitions was made by Gandomi and Haider.<sup>[2]</sup> The common axis of all definitions is the perception that big data, which can be viewed as a ‘new era’ of the data-driven paradigm, has opened up new possibilities for the improved decision support.<sup>[3]</sup> Big data is not only vast and dynamic, but it also necessitates the use of cutting-edge technologies to analyze and process.<sup>[4]</sup> Big data is distinguished from the traditional “data” notation, because big data, due to its stupendous size, cannot be processed and managed with conventional data mining tools.<sup>[5][6]</sup>

## 2. Big data mining (BDM)

The concept of big data mining (BDM) is intimately associated to conventional data mining notation. These two concepts differ mainly in the methods of obtaining data, and not in the idea itself. BDM enables to obtain useful information from databases or data streams that are huge in terms of “big data V’s”, like volume, velocity, and variety.<sup>[7]</sup> The main functions of data mining in general are descriptive functions (such as clustering, association, and pattern mining) and predictive functions (such as classification, time series analysis, etc.). These functions (enablers) differ only slightly from each other, mainly in terms of the temporal reference, i.e., descriptive functions mainly concern the present and settled dependencies, and the predictive functions mainly refer to the study of the future and tentative dependencies. Having collected (big) data, it is necessary to analyze it to extract information

hidden in it. In this case, it is crucial to use big data analytical tools. BDA was created in response to the need to analyze vast volumes of quickly collected complex data. As a result, data acquisition and processing occur at a high pace, which is impossible to achieve with calcareous computational methods.<sup>[8]</sup> Big data analytics, as being a big data derivative, can also be described with big data “V” characteristics. The ultimate and the pivotal step of the whole BDA process is “action on insight”, as Akhtar *et al.* <sup>[9]</sup> claim. The use and implementation of BDA, as one of the most principal factors for engendering meaningful insights for decision-making,<sup>[10][11]</sup> is crucial to extract value from the multitude of data being obtained. An organizational capability to handle BDA has recently become mainstream to create value.<sup>[12]</sup> It should be noted, however, that the blossoming of BDA potential in organizations can be withheld due to the lack of IT infrastructure, data storage facilities and organization strategy.<sup>[13]</sup>

## 2.1. Big data mining for customer insights

In order to extract the hidden value of customer insights, big data (along with derived approaches, such as big data mining or wider: big data analytics) come in handy. Having presented the historical and theoretical background regarding the values of customer insights, it is possible to explore this topic in an even more extensive way, paying attention e.g., to the practical use of big data for customer insights in organizations.

Big data-enabled societies – particularly based on the foundations of the digital economy – are capable of opening new perspectives for organizations striving to get to know their customers better.<sup>[14]</sup> The enormous volumes of data deriving from a variety of sources allow to analyze greater number of dimensions depicting customers, than before. This is mostly due to the new sources of data origin, as e.g., social media. Thanks to the characteristics of big data, it offers gargantuan possibilities for gaining new insights.<sup>[15][16]</sup> These new insights are not narrowed to customer-centric decision-making processes but affect the whole operating space of an organization. In simplification, the insights from big data – if properly used – may contribute to value generation, as well as to innovations, and to the competitive advantage.<sup>[17][18]</sup> Specifically, big data analytics is considered together with data mining issues.<sup>[19]</sup> For example, Xindong Wu *et al.* <sup>[20]</sup> propose a big data processing model from the data mining perspective. They point out that mining big data is data-driven and demand-driven. This context of data mining is present in many big data analytics definitions. In the paper of Mohsenian-Rad *et al.* <sup>[21]</sup> this type of analytics is described as the process of uncovering hidden patterns, unknown correlations, irregularities, and other data-driven intelligence. Data mining related to big data analytics’ tasks also encompasses text mining (for sentiment analysis) and social media analytics (for community detection or social influence analysis).<sup>[2]</sup> Especially the latter may be of paramount importance in modern dynamic operational environments, due to empowerment of organizations to perform the so-called situational data analytics instead of – or at least together with – classic static data analytics of transactional or enterprise data.<sup>[22]</sup>

## 2.2. Big data challenges in the context of BDM

Generating insights from big data is a process consisting of two main activities – data management and data analytics. The former encompasses data acquisition, extraction, cleaning, integration, and representation. The latter consists of data modeling, analysis, and interpretation.<sup>[2]</sup> Data mining techniques are among the most used

ones in big data analytics. Hence it can be assumed that challenges in BDA concern also big data mining, even if not expressed explicitly. In a very general way, the first and foremost challenge of big data analytics is to generate business value.<sup>[23]</sup> It is also one of the ultimate goals of BDA and big data mining. The other ones are also provision of competitive advantage,<sup>[24][25]</sup> and generation of new business ideas from big data insights. Obviously, the quality of insights results from a proper orchestration of big data-related resources, that is data, technology, processes, and people within the framework of organization.<sup>[26]</sup> Human skills and organizational culture are as important as the technological dimensions of BDA in providing valuable results for the success of organization.<sup>[17]</sup><sup>[27]</sup> The overall success of big data analytics is therefore dependent on such factors as top management support, organizational change, technical infrastructure, the data science skillset, data availability and quality, data security and privacy.<sup>[28]</sup> Conceptually, the overall big data challenges can be summarized as presented in Figure 1.

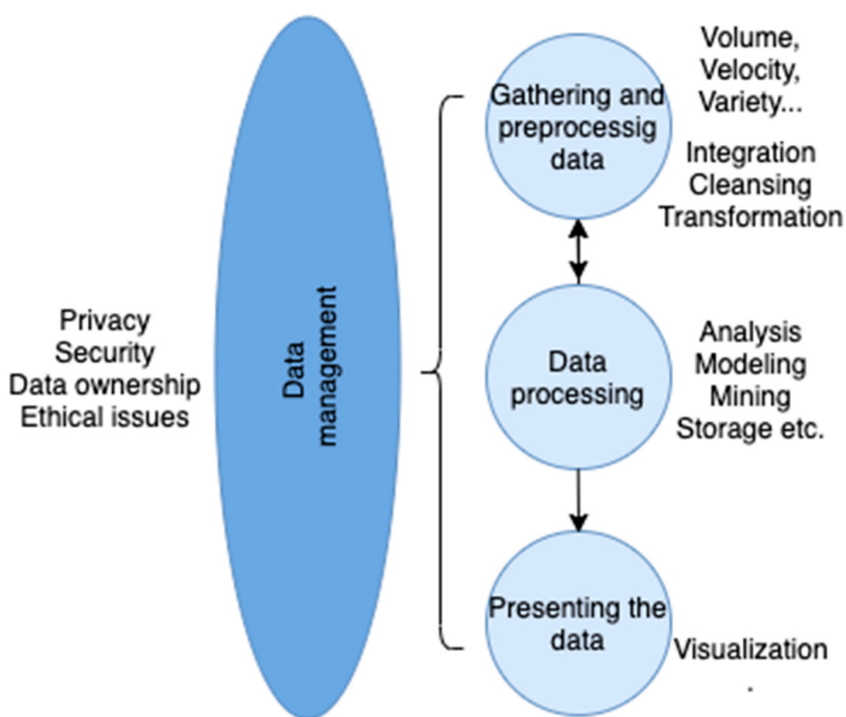


Figure 1. Conceptual classification of big data

challenges. Source: based on <sup>[29]</sup>.

In the big data lifecycle, the first group of challenges concerns the characteristics of big data itself (the “Vs”) which in turn affect the issues of big data preprocessing, e.g., integration, cleansing, and transformation. At the data processing stage, the typical tasks of analysis, modeling, mining etc. must be adjusted to properly address the challenges of the first stage. Also, at the stage of presenting the results, the graphical methods must be able to cope with visualizing a huge amount of big data analysis results. The big data management stage extends over all other stages and is associated with challenges such as privacy, security, data ownership and other ethical issues.<sup>[30][31][32][33][34]</sup> Among the features determining data, information, and insights quality the completeness, accuracy, and currency are mentioned to be the most significant.<sup>[35]</sup> Especially the last feature is a challenge when applied to big data analysis and mining. Not only the data comes as a stream or flow (the velocity dimension of big data) but also it must be analyzed/mined in real-time manner to provide value to organizations. We will cover this temporal challenge later.

With data mining (DM) as one of the most important elements of big data analytics, it is not surprising that the DM software is one of the most appreciated tools among various analytical tools used for BDA.<sup>[36]</sup> However, even the best software will not produce valuable results from garbage data. Hence, among big data mining challenges the first group includes data inconsistency and incompleteness, scalability, timeliness, and data security. Challenges also concern data capture, storage, searching, sharing, analysis, and visualization.<sup>[37]</sup> There is a common agreement that before mining the data it is mandatory to consider such issues as validity and reliability of data. The bigger the data quantity is, the bigger the challenge. With the amount of data, discovering dependencies and valuable patterns becomes extremely difficult.<sup>[15][38]</sup> The next group of challenges is associated with the mining process itself. The algorithms and techniques used for “classic” data mining in e.g., data warehouses sometimes are not suited to be used with huge amounts of constantly incoming big data. This is so because traditional data mining approaches start with a centralized data repository, able to store and process data. With the prodigious size and variety characterizing big data such centralized approach may not be used. There is a strong need of more distributed approaches capable of mining huge amounts of unstructured data.<sup>[39]</sup> Some other challenges include e.g., lack of large-scale data representation (for mining purposes), lack of effective and efficient on-line large-scale machine learning techniques, lack of data confidentiality mechanism.<sup>[40]</sup> Challenges concern mining algorithms which must deal with sparse, uncertain, incomplete, complex, and dynamic data.<sup>[20]</sup> Also, the constant inflow of data to be mined can be recognized a momentous challenge, as many mining algorithms do not provide proper sequences or patterns.<sup>[41]</sup> Some of the proposals to overcome this obstacle include e.g., incremental pattern mining and cluster analysis, when the discovered patterns and clusters are incrementally augmented with updated information,<sup>[42]</sup> post-processing enhancements of mined patterns<sup>[43]</sup> and special spatio-temporal representations of data for further mining.<sup>[40]</sup> Therefore, these stages, like data cleaning, integration, ranking and querying, are often considered as the sources of “algorithmic bias”.<sup>[44][45]</sup> Reasoning about them as well as attenuating inequity upstream from the final data analysis phase is potentially more impactful.<sup>[45]</sup>

### 2.3. Temporal big data mining for gaining customer insights

However, it becomes obvious that for valuable insights from big data mining it is essential to consider temporal-related issues. As Xindong Wu *et al.*<sup>[20]</sup> point out, in a dynamic world data and information representing interesting features from the environment of an enterprise enlargement. Hence while mining useful patterns from big data, it is indispensable to consider these evolving changes. However, it seems symptomatic that a miniature number of big data analytics definitions even mention the question of dynamics. For example, Mikalef *et al.*<sup>[24]</sup> while presenting sample definitions of BDA consider only two ones addressing the dynamic dimension of BDA. A challenge in big data mining hence arises – how to deal with dynamic/temporal aspects of the realm described by big data. One of the ways to do it is to implement agile big data analytics. BDA is seen as a ‘bridging’ instrument in development of software applications using agile methods.<sup>[46]</sup> Agility is achieved by creating a data infrastructure enabling identification and evaluation of various big data sources.<sup>[25]</sup> Afterwards, there are approaches focused on big data stream processing enabling a flexible mining solution.<sup>[39]</sup> However, these solutions are insufficient when it comes to real-time data processing.<sup>[47]</sup> The real-time big data analytics presents another challenge related to big data mining which must be considered.<sup>[30]</sup> Many of the phenomena of interest to the organization are represented as time series,<sup>[48]</sup> this applies to e.g., sentiment analysis or user’s website activities. But many other phenomena are too

intricate to be represented this way. Knowledge coming from organization's environment evolves very quickly because of a constant inflow of data and information. The big data mining in real time may ameliorate decision-making processes in organizations because it would enable dealing with real-time uncertainties.<sup>[16][49]</sup> The time dimension of big data is reflected in the speed of their inflow. This causes big data to be transient which implies the need to mine them as and when they are generated.<sup>[50]</sup> The timeliness of data analysis and mining is the succeeding challenge, tightly linked with the challenge of dealing with temporal dimension of big data. This timeliness challenge is discussed in <sup>[37]</sup> in more detail.

The most intuitive way to deal with temporal aspects of big data mining is to treat the data inflow as a set of events. This is quite natural because events are the building blocks of surroundings of organizations, hence they need to be represented and mined during big data analytics. The big data mining process should be therefore focused on events implied from the massive volume of data. It is thus clear that big data mining is closely related to events.<sup>[51]</sup> The consecutive big data mining challenges may be formulated as: the challenge of event capturing and representation for further analytics, the challenge of constructing temporal big data mining algorithms, and the challenge of representing temporal features of the mined knowledge. The events and temporal information in big data should be identified, the temporal relations among events should be found and represented, event-based information retrieval and analytics should be done. Wang *et al.* <sup>[40]</sup> proposed a big data temporal analytics solution but only for texts, while leaving apart many other forms of data leading to the big data variety feature.

Another approach to temporal big data has been proposed in the work of Singh *et al.* <sup>[42]</sup> where the frequent patterns mining, and cluster analysis model are used on constantly incoming data. The model encompasses a progressive and incremental update of mined patterns and clusters with new information, and newly discovered patterns and clusters are incrementally added to the existing ones. However, events are not addressed in this approach and there is lack of temporal representation. In fact, the model is concentrated on time series instead of event sequences. An answer to the challenge of temporal BD mining has been proposed in.<sup>[40][52]</sup> Both approaches consider Complex Event Processing (CEP) systems as a solution. CEP systems are particularly useful for real-time analytics and stream reasoning. These solutions differ in time representation. Some of them are based on point structures, while others are based on intervals (cf. <sup>[53]</sup>). The CEP systems also differ in their complexity and in orientation: computation-oriented vs. detection-oriented ones.<sup>[40]</sup> A variation of event processing systems has been proposed by in <sup>[39]</sup>, namely Semantic Complex Event Processing augmented with an agent that dynamically builds an ontology which can then be queried temporally. However, even the mining systems based on event processing are yet not capable of mining causality relations <sup>[41]</sup> which would contain a lot of useful information on complex phenomena in organization's environment.

Another group of approaches to analyzing streaming and/or temporal big data is built upon the so-called ontology-based data access (OBDA). OBDA origins from the Semantic Web analytics and its core feature lies in separating conceptual and database levels of data.<sup>[22]</sup> Unfortunately, OBDA itself does not adapt to changes in data sources. The W3C standardized an ontology and a query language for the ODBA: OWL2QL and SPARQL,<sup>[54]</sup> but these solutions do not handle essentially temporal big data. Incorporating complex temporal information into OBDA

together with the ability to process heterogeneous data poses a serious challenge.<sup>[55]</sup> A temporal OBDA is then requisite. There are various ways to the development of such a temporal version of OBDA:

- Extending OWL2QL with various temporal operators;<sup>[56]</sup>
- Extending both OWL2QL and SPARQL with the LTL (linear temporal logic) temporal operators;<sup>[57][58]</sup>
- Extending OWL2QL and SPARQL with the MTL (metric temporal logic) temporal operators;<sup>[54]</sup>
- Extending OWL2QL and SPARQL with the interval logic by Halpern and Shoham;<sup>[59]</sup>
- Developing fully temporal OBDA.<sup>[59]</sup>

The advantage of all the above solutions lies in the direct incorporation of time dimension into analytics. On the other hand, the main disadvantage and weakness in the context of big data mining concern the nature of data and analytics. All the above solutions are directed towards relational/structured data and queries, and do not deal with any data mining tasks. Hence, they cannot be considered satisfactory for temporal big data mining. The challenge which then is seen concerns augmenting the existing big data mining models, methods, and algorithms with explicit temporal expressions and with ways to handle them to mine temporal big data.

### 3. Summary

Summing up, we note several challenges for big data mining, especially in the context of customer insights. These are:

1. completeness, accuracy, and currency of discovered insights/patterns,
2. quality of data to be mined,
3. issues concerning big data storing/processing,
4. modification of mining algorithms and techniques to deal with abundant, heterogeneous, and streaming data,
5. dealing with evolving changes,
6. dealing with dynamics/velocity of big data,
7. flexibility of mining algorithms and techniques,
8. representing and processing big data as events and event sequences.

All these challenges constitute important and promising research areas, but as shown, the most important and challenging issue concerns incorporating explicit time notion into representation and mining procedures of big data. There is a strong need to express temporal dimension of big data and in big data itself, using more complex temporal representations than event calculus. There is a need to represent causality of phenomena, of discovering changes in phenomena depicted by big data, and of mining useful temporal patterns to get deep and profound insights on the way the world around organizations evolves.

We have focused on the challenges associated with big data mining which is a specific subarea of BDA. Obviously, the broader BDA field also faces several challenges. These are primarily challenges with big data's volume characteristics. The large sample size may result in several biases as e.g., sampling error, measurement error, aggregation error etc.<sup>[60]</sup> Especially the sampling error may result in highly biased data. Researchers have shown examples of such biased data collected from social networks. While gathering this data, it may be erroneously

assumed that social media users are representative of the population <sup>[61]</sup> while there are many social groups excluded from using the SM. E.g., people digitally excluded – due to age, education, low socioeconomic status may not be represented in the retrieved big data sample.<sup>[61][62][63]</sup> A noticeable bias in big data may also result from gender and race issues.<sup>[64][65][66]</sup> All these challenges should be kept in mind while addressing the question of big data analytics, however, they are beyond the scope of this entry.

## References

1. Russell Newman; Victor Chang; Robert John Walters; Gary Wills; Web 2.0—The past and the future. *International Journal of Information Management* **2016**, 36, 591-598, 10.1016/j.ijinfomgt.2016.03.010.
2. Amir Gandomi; Murtaza Haider; Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* **2015**, 35, 137-144, 10.1016/j.ijinfomgt.2014.10.007.
3. Power, D.J. 'Big Data' Decision Making Use Cases. In *Decision Support Systems V – Big Data Analytics for Decision Making*; Springer: Belgrade, Serbia, 2015; pp. 1–9.
4. Jasmine Zakir; Tom Seymour; Kristi Berg; Big Data Analytics. *Issues in Information Systems* **2015**, 16, 81-90, 10.48009/2\_iis\_2015\_81-90.
5. T. Sajana; C. M. Sheela Rani; K. V. Narayana; A Survey on Clustering Techniques for Big Data Mining. *Indian Journal of Science and Technology* **2016**, 9, 1-5, 10.17485/ijst/2016/v9i3/75971.
6. Albert Bifet; Mining Big Data in Real Time. *Informatica* **2013**, 37, 15-20.
7. Jaseena, K.U.; David, J.M. Issues, Challenges and Solutions : Big Data Mining. In *Proceedings of the Computer Science & Information Technology (CS & IT). Sixth International Conference on Networks & Communications*; Academy & Industry Research Collaboration Center (AIRCC), December 2014; pp. 131–140.
8. Ioanna D Constantiou; Jannis Kallinikos; New Games, New Rules: Big Data and the Changing Context of Strategy. *Journal of Information Technology* **2015**, 30, 44-57, 10.1057/jit.2014.17.
9. Pervaiz Akhtar; Jędrzej George Frynas; Kamel Mellahi; Subhan Ullah; Big Data-Savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *British Journal of Management* **2018**, 30, 252-271, 10.1111/1467-8551.12333.
10. Rameshwar Dubey; Angappa Gunasekaran; Stephen J. Childe; Constantin Blome; Thanos Papadopoulos; Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. *British Journal of Management* **2018**, 30, 341-361, 10.1111/1467-8551.12355.

11. Usama Awan; Saqib Shamim; Zaheer Khan; Najam Ul Zia; Syed Muhammad Shariq; Muhammad Naveed Khan; Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. *Technological Forecasting and Social Change* **2021**, *168*, 120766, 10.1016/j.techfore.2021.120766.
12. Barbara H. Wixom; Bruce Yen; Michael Relich; Maximizing Value from Business Analytics. *MIS Quarterly Executive* **2013**, *12*, 111-123.
13. Rakesh Raut; Vaibhav Narwane; Sachin Kumar Mangla; Vinay Surendra Yadav; Balkrishna Eknath Narkhede; Sunil Luthra; Unlocking causal relations of barriers to big data analytics in manufacturing firms. *Industrial Management & Data Systems* **2021**, *ahead-of-print*, ahead-of-print, 10.1108/imds-02-2020-0066.
14. Michel Wedel & P. K. Kannan; Marketing Analytics for Data-Rich Environments. *Journal of Marketing* **2016**, *80*, 97-121, 10.1509/jm.15.0413.
15. David Lazer; Ryan Kennedy; Gary King; Alessandro Vespignani; The Parable of Google Flu: Traps in Big Data Analysis. *Science* **2014**, *343*, 1203-1205, 10.1126/science.1248506.
16. Zhenning Xu; Gary L. Frankwick; Edward Ramirez; Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research* **2016**, *69*, 1562-1566, 10.1016/j.jbusres.2015.10.017.
17. Candice Walls; Brian Barnard; Success Factors of Big Data to Achieve Organisational Performance: Theoretical Perspectives. *Expert Journal of Business and Management* **2020**, *8*, 1-16.
18. Celina M. Olszak; Jozef Zurada; Big Data in Capturing Business Value. *Information Systems Management* **2019**, *37*, 240-254, 10.1080/10580530.2020.1696551.
19. Danyel Fisher; Rob Deline; Mary Czerwinski; Steven Drucker; Interactions with big data analytics. *Interactions* **2012**, *19*, 50-59, 10.1145/2168931.2168943.
20. Xindong Wu; Xingquan Zhu; Gong-Qing Wu; Wei Ding; Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* **2013**, *26*, 97-107, 10.1109/tkde.2013.109.
21. Hamed Mohsenian-Rad; Emma Stewart; Ed Cortez; Distribution Synchronizers: Pairing Big Data with Analytics to Create Actionable Information. *IEEE Power and Energy Magazine* **2018**, *16*, 26-34, 10.1109/mpe.2018.2790818.
22. Sergi Nadal; Oscar Romero; Alberto Abelló; Panos Vassiliadis; Stijn Vansummeren; An integration-oriented ontology to govern evolution in Big Data ecosystems. *Information Systems* **2019**, *79*, 3-19, 10.1016/j.is.2018.01.006.
23. Shahriar Akter; Samuel Fosso Wamba; Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets* **2016**, *26*, 173-194, 10.1007/s12525-016-0219



-0.

24. Patrick Mikalef; Ilias Pappas; John Krogstie; Michail Giannakos; Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management* **2017**, *16*, 547-578, 10.1007/s10257-017-0362-y.
25. Brent Kitchens; David Dobolyi; Jingjing Li; Ahmed Abbasi; Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. *Journal of Management Information Systems* **2018**, *35*, 540-574, 10.1080/07421222.2018.1451957.
26. Patrick Mikalef; Maria Boura; George Lekakos; John Krogstie; Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research* **2019**, *98*, 261-276, 10.1016/j.jbusres.2019.01.044.
27. Saqib Shamim; Jing Zeng; Syed Muhammad Shariq; Zaheer Khan; Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & Management* **2019**, *56*, 103135, 10.1016/j.im.2018.12.003.
28. Mohanad Halaweh; Ahmed E. Massry; Conceptual Model for Successful Implementation of Big Data in Organizations. *Journal of International Technology and Information Management* **2015**, *24*, 2.
29. Uthayasankar Sivarajah; Muhammad Mustafa Kamal; Zahir Irani; Vishanth Weerakkody; Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research* **2017**, *70*, 263-286, 10.1016/j.jbusres.2016.08.001.
30. Vassakis, K.; Petrakis, E.; Kopanakis, I. Big Data Analytics: Applications, Prospects and Challenges. In *Mobile Big Data. Lecture Notes on Data Engineering and Communications Technologies*, vol 10.; Skourletopoulos, G., Mastorakis, G., Mavromoustakis, C., Dobre, C., Pallis, E., Eds.; Springer, Cham: Berlin-Heidelberg, 2018; pp. 3–20 ISBN 978-3-319-67924-2.
31. Che, D.; Safran, M.; Peng, Z. From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. In *Database Systems for Advanced Applications*; Hong, B., Meng, X., Chen, L., Winiwarter, W., Song, W., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp. 1–15 ISBN 978-3-642-40270-8.
32. Desamparados Blazquez; Josep Domenech; Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change* **2018**, *130*, 99-113, 10.1016/j.techfore.2017.07.027.
33. Rong, Y.; Xu, Z.; Yan, R.; Ma, X. Du-Parking: Spatio-Temporal Big Data Tells You Realtime Parking Availability. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*; ACM Press: New York, New York, USA, 2018; pp. 646–654.

34. Syed, L.; Jabeen, S.; Manimala, S.; Elsayed, H.A. Data Science Algorithms and Techniques for Smart Healthcare Using IoT and Big Data Analytics: Towards Smarter Algorithms. In *Studies in Fuzziness and Soft Computing*; Mishra, M.K., Mishra, B.S.P., Patel, Y.S., Misra, R., Eds.; Springer Verlag: Cham, 2019; Vol. 374, pp. 211–241 ISBN 978-3-030-03130-5.
35. Steven Ji-Fan Ren; Samuel Fosso Wamba; Shahriar Akter; Rameshwar Dubey; Stephen J. Childe; Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research* **2016**, *55*, 5011-5026, 10.1080/00207543.2016.1154209.
36. Shridhar Dhamodaran; K.R. Sachin; Rahul Kumar; Big Data Implementation of Natural Disaster Monitoring and Alerting System in Real Time Social Network using Hadoop Technology. *Indian Journal of Science and Technology*, **2015**, *8*, 1.
37. C.L. Philip Chen; Chun-Yang Zhang; Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* **2014**, *275*, 314-347, 10.1016/j.ins.2014.01.015.
38. Elisabetta Raguseo; Claudio Vitari; Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. *International Journal of Production Research* **2018**, *56*, 5206-5221, 10.1080/00207543.2018.1427900.
39. Christian Esposito; Massimo Ficco; Francesco Palmieri; Aniello Castiglione; A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing. *Knowledge-Based Systems* **2015**, *79*, 3-17, 10.1016/j.knosys.2014.05.003.
40. Wang, J.; Zhang, W.; Shi, Y.; Duan, S.; Liu, J. Industrial Big Data Analytics: Challenges, Methodologies, and Applications. arXiv preprint arXiv:1807.01016 2018.
41. B. Thejaswini; Reddi Durga Sree; Karamala Suresh; Study of user's behaviour in Structured E-Commerce Websites. *International Journal of Scientific Research & Engineering Trends* **2018**, *4*, 665-668.
42. Shailendra Singh; Abdulsalam Yassine; Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. *Energies* **2018**, *11*, 452, 10.3390/en11020452.
43. Ray Y. Zhong; George Q. Huang; Shulin Lan; Q.Y. Dai; Xu Chen; T. Zhang; A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics* **2015**, *165*, 260-272, 10.1016/j.ijpe.2015.02.014.
44. Keith Kirkpatrick; It's not the algorithm, it's the data. *Communications of the ACM* **2017**, *60*, 21-23, 10.1145/3022181.
45. Serge Abiteboul; Julia Stoyanovich; Transparency, Fairness, Data Protection, Neutrality. *Journal of Data and Information Quality* **2019**, *11*, 1-9, 10.1145/3310231.

46. Katarzyna Biesialska; Xavier Franch; Victor Muntés-Mulero; Big Data analytics in Agile software development: A systematic mapping study. *Information and Software Technology* **2021**, *132*, 106448, 10.1016/j.infsof.2020.106448.
47. Saldivar, A.A.F.; Goh, C.; Chen, W.; Li, Y. Self-Organizing Tool for Smart Design with Predictive Customer Needs and Wants to Realize Industry 4.0. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC); IEEE: Vancouver, July 2016; pp. 5317–5324.
48. Pouyanfar, S.; Chen, S.-C.; Shyu, M.-L. Deep Spatio-Temporal Representation Learning for Multi-Class Imbalanced Data Classification. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI); Khan, L., Palanisamy, B., Mehedy Masud, M., Bifet, A., Eds.; IEEE: Salt Lake City, July 2018; pp. 386–393.
49. Mochamad Nizar Palefi Ma'ady; Chuan-Kai Yang; Renny Pradina Kusumawardani; Hatma Suryotrisongko; Temporal Exploration in 2D Visualization of Emotions on Twitter Stream. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* **2018**, *16*, 376-384, 10.12928/telkomnika.v16i1.6591.
50. V. Rajaraman; Big data analytics. *Resonance* **2016**, *21*, 695-716, 10.1007/s12045-016-0376-7.
51. Junsheng Zhang; Changqing Yao; Yunchuan Sun; ZengQuan Fang; Building text-based temporally linked event network for scientific big data analytics. *Personal and Ubiquitous Computing* **2016**, *20*, 743-755, 10.1007/s00779-016-0940-x.
52. della Valle, E.; Dell'Aglio, D.; Margara, A. Taming Velocity and Variety Simultaneously in Big Data with Stream Reasoning. In Proceedings of the Proceedings of the 10th ACM International Conference on Distributed and Event-based Systems - DEBS '16; ACM Press: New York, New York, USA, 2016; pp. 394–401.
53. Fisher, M. Temporal Representation and Reasoning. In Foundations of Artificial Intelligence; Van Harmelen, F., Lifschitz, V., Porter, B., Eds.; Elsevier: Amsterdam, 2008; Vol. 3, pp. 513–550 ISBN 9780444522115.
54. Sebastian Brandt; Elem Güzel Kalaycı; Vladislav Ryzhikov; Guohui Xiao; Michael Zakharyashev; Querying Log Data with Metric Temporal Logic. *Journal of Artificial Intelligence Research* **2018**, *62*, 829-877, 10.1613/jair.1.11229.
55. Elem Güzel Kalaycı; Sebastian Brandt; Diego Calvanese; Vladislav Ryzhikov; Guohui Xiao; Michael Zakharyashev; Ontology-based access to temporal data with Ontop: A framework proposal. *International Journal of Applied Mathematics and Computer Science* **2019**, *29*, 17-30, 10.2478/amcs-2019-0002.
56. Kharlamov, E.; Brandt, S.; Jimenez-Ruiz, E.; Kotidis, Y.; Lamparter, S.; Mailis, T.; Neuenstadt, C.; Özçep, Ö.; Pinkel, C.; Svingos, C.; et al. Ontology-Based Integration of Streaming and Static Relational Data with Optique. In Proceedings of the Proceedings of the ACM SIGMOD

- International Conference on Management of Data; Association for Computing Machinery: New York, New York, USA, June 26 2016; Vol. 26-June-20, pp. 2109–2112.
57. Artale, A.; Kontchakov, R.; Kovtunova, A.; Ryzhikov, V.; Wolter, F.; Zakharyashev, M. First-Order Rewritability of Temporal Ontology-Mediated Queries. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence; 2015.
58. Gutiérrez-Basulto, V.; Jung, J.C.; Kontchakov, R. Temporalized EL Ontologies for Accessing Temporal Data: Complexity of Atomic Queries. In Proceedings of the Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence; 2016; pp. 1102–1108.
59. Xiao, G.; Calvanese, D.; Kontchakov, R.; Lembo, D.; Poggi, A.; Rosati, R.; Zakharyashev, M. Ontology-Based Data Access: A Survey. In Proceedings of the Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; International Joint Conferences on Artificial Intelligence Organization: California, July 2018; pp. 5511–5519.
60. Robert M. Kaplan; David A. Chambers; Russell E. Glasgow; Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias. *Clinical and Translational Science* **2014**, 7, 342-346, 10.1111/cts.12178.
61. Eszter Hargittai; Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review* **2018**, 38, 10-24, 10.1177/0894439318788322.
62. Grant Blank; Christoph Lutz; Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist* **2017**, 61, 741-756, 10.1177/0002764217717559.
63. Michael J. Stern; Ipek Bilgen; Colleen McClain; Brian Hunscher; Effective Sampling From Social Media Sites and Search Engines for Web Surveys. *Social Science Computer Review* **2016**, 35, 713-732, 10.1177/0894439316683344.
64. Erkan Saka; Big Data and Gender-Biased Algorithms. *The International Encyclopedia of Gender, Media, and Communication* **2020**, no volume, 1-4, 10.1002/9781119429128.iegmc267.
65. Mike Thelwall; Gender bias in machine learning for sentiment analysis. *Online Information Review* **2018**, 42, 343-354, 10.1108/oir-05-2017-0153.
66. Pierre-André G. Maugis; Big data uncertainties. *Journal of Forensic and Legal Medicine* **2018**, 57, 7-11, 10.1016/j.jflm.2016.09.005.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/30158>