Tensor Decompositions for Large-Scale Data Mining: Methods for Uncovering Latent Patterns in Multidimensional Big Data

Aleksandar Petrović
Computer Science, University of Novi Sad
Jelena Milošević
Data Science, University of Novi Sad

Abstract

With the proliferation of big data across many domains, there is an increasing need for advanced analytical methods that can uncover latent patterns and extract useful knowledge from massive, multidimensional datasets. Tensor decompositions offer a powerful approach for large-scale data mining by representing higher-order data arrays as a multilinear model via decomposition into factor matrices. This allows for dimensionality reduction while preserving the essential structure and relationships within the data. In this paper, we provide a comprehensive overview of tensor decompositions for data mining, including the mathematical foundations, algorithms, applications, and software implementations. We focus on the two most widely used techniques: CANDECOMP/PARAFAC (CP) and Tucker decompositions. Through detailed numerical examples on real-world datasets, we demonstrate how tensor decompositions can be utilized for latent pattern discovery in areas such as social network analysis, neuroimaging analysis, recommender systems, and text mining. We also discuss computational aspects and scalability challenges associated with applying tensor methods to massive datasets. Overall, tensor decompositions provide versatile tools for uncovering hidden signals in big data, with tremendous potential for gaining actionable insights across many domains.

Keywords:
- Tensor decompositions
- Large-scale data mining
- Multidimensional data
- Latent pattern discovery
- Tucker decomposition

Excellence in Peer-Reviewed Publishing:
QuestSquare

Introduction

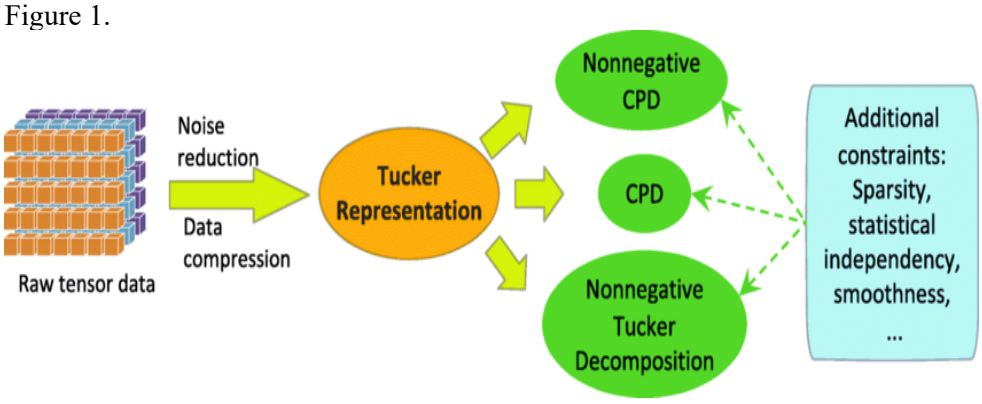
The rapid expansion of data volume, velocity, and variety in the contemporary era of big data presents a spectrum of both opportunities and challenges for knowledge discovery and pattern recognition. The proliferation of data collected from diverse sources offers the potential to unveil previously undiscovered relationships and attain profound insights [1]. However, the sheer magnitude and intricacy of multidimensional big data pose a formidable obstacle to the direct analysis and interpretation of raw data in its entirety. In this context, tensor decompositions emerge as a pivotal unsupervised learning method tailored for large-scale data mining

endeavors. By disassembling higher-order data arrays into lower-order factors, tensor methods facilitate essential functionalities such as dimensionality reduction, latent structure discovery, feature extraction, and subspace clustering within massive, multidimensional datasets [2].

The advent of big data has ushered in an era where traditional data processing techniques often fall short in effectively handling the vast and intricate datasets that characterize contemporary information landscapes [3]. The three V's of big data – volume, velocity, and variety – encapsulate the magnitude, speed, and diversity of data sources, respectively. This surge in data dimensions creates a dual landscape of promise and complexity. On the promising side, the unprecedented scale and diversity of data provide an extensive canvas for exploration, offering the potential to uncover hidden patterns, correlations, and insights that were previously elusive. Conversely, the complex nature of multidimensional big data presents a formidable challenge, necessitating innovative approaches to extract meaningful information and knowledge.

Tensor decompositions stand out as a sophisticated and potent technique in addressing the challenges posed by large-scale, multidimensional datasets. These decompositions, by breaking down high-order data arrays into lower-order components, enable a more manageable representation of the underlying structure within the data. One of the primary advantages lies in the realm of dimensionality reduction [4]. The ability to distill complex data into lower-dimensional forms not only facilitates more efficient storage and processing but also enhances the interpretability of the data. In essence, tensor methods serve as a conduit for transforming unwieldy datasets into a more digestible format, thereby laying the groundwork for subsequent analysis and interpretation. Moreover, tensor decompositions play a crucial role in latent structure discovery within big data. The intricate relationships and hidden structures inherent in massive datasets often elude conventional analysis techniques. Tensor methods, by virtue of their ability to unveil latent factors within the data, contribute significantly to revealing underlying patterns and structures. This latent structure discovery is instrumental in enhancing our understanding of complex systems, identifying key influencers, and refining predictive models. In the context of knowledge discovery, the capability of tensor decompositions to expose latent structures opens avenues for novel insights and a deeper comprehension of the underlying dynamics driving the observed data patterns.

Figure 1.



Feature extraction represents another pivotal application of tensor methods in the realm of big data analytics. As datasets grow in size and complexity, the challenge of identifying and extracting relevant features becomes increasingly daunting. Tensor decomposition offers an elegant solution by isolating essential features embedded within the multidimensional data. This process not only streamlines subsequent analysis but also contributes to the development of more robust and efficient machine learning models. By distilling the salient features from the expansive dataset, tensor methods empower practitioners to focus on the most relevant aspects of the data, fostering more accurate and meaningful outcomes in various applications, from image recognition to natural language processing. Furthermore, tensor decompositions excel in subspace clustering, a critical task in the analysis of multidimensional datasets [5]. The inherent complexity of high-dimensional data often leads to the presence of subspaces lower-dimensional structures within the overall data space. Identifying and clustering these subspaces is fundamental for discerning distinct patterns and groups within the data. Tensor methods, through their ability to capture the underlying structure of data in a lower-dimensional space, prove invaluable in the task of subspace clustering. This not only aids in grouping similar data points but also contributes to the identification of outliers and anomalies, enhancing the overall robustness of the analytical process [6].

Tensors provide a natural way to represent multidimensional data arrays, which frequently arise in domains such as social network analysis, neuroimaging, chemometrics, signal processing, and more. For example, a social network can be characterized by a 3-way tensor with modes corresponding to users, friends, and interactions. An image collection forms a 3-way tensor with pixels, colors, and images as modes. Even ordinary data tables are 2-way tensors. While vectors and matrices (1st and 2nd order tensors) can be analyzed using well-established methods like PCA and SVD, these techniques do not extend easily to higher-order tensors. Tensor decompositions provide the required mathematical framework and computational tools to harness the rich information content in multidimensional datasets [7].

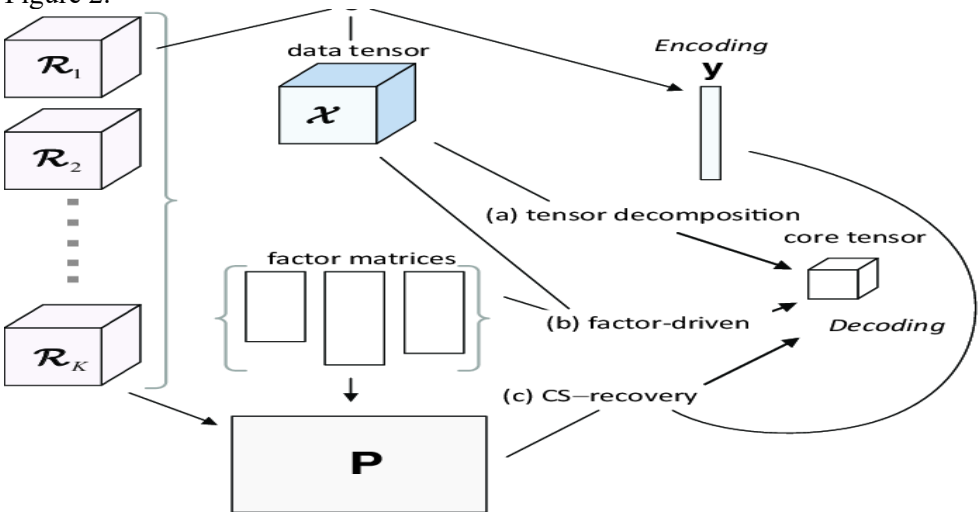
This paper aims to provide a comprehensive overview of tensor decomposition methods for uncovering latent patterns and extracting knowledge from large-scale, multidimensional data. We begin with the mathematical preliminaries of tensor basics and multilinear algebra. We then describe the two most widely used techniques: CANDECOMP/PARAFAC (CP) decomposition which yields a sum of rank-one tensor components, and Tucker decomposition which provides a core tensor transformed by factor matrices along each mode. After reviewing the underlying models and algorithms, we discuss various applications of tensor decompositions in domains like web data mining, neuroscience, signal processing, and recommender systems. Through detailed numerical examples on real datasets, we demonstrate how tensor methods can uncover latent components and interactions in large datasets. We also address computational challenges and software tools available. Overall, this paper highlights the tremendous potential of tensor decompositions for gaining fundamental insights and hidden structure from big multidimensional data [8].

Tensor Preliminaries

A tensor is formally defined as a multidimensional array, where an Nth order tensor belongs to the tensor product of N vector spaces, each representing a distinct mode or

aspect of the tensor. For instance, a matrix is a 2nd order tensor, characterized by rows and columns as its modes, while a 3rd order tensor comprises three modes. The order or dimensionality of a tensor is interchangeably termed as its way or mode [9].

Figure 2.



Several fundamental concepts in tensor algebra contribute to a comprehensive understanding of tensors. Fibers, akin to higher-order analogs of matrix rows and columns, are derived by fixing all indices except one. Slices, on the other hand, denote 2D sections of a tensor and are defined by fixing all but two indices. Tensor multiplication extends the principles of matrix multiplication to higher orders [10]. The rank of a tensor is identified as the minimum number of rank-one tensors essential to generate the tensor as their summation. The mode-n product of a tensor with a matrix corresponds to the matrix multiplication of each mode-n fiber with the matrix. Additionally, the Kronecker product serves as an implementation of the tensor product of two matrices [11]. The primary objective of tensor decompositions lies in obtaining low-rank approximations of higher-order tensors. This is achieved by representing tensors as sums of outer products of vectors, essentially rank-one tensors. Such decompositions facilitate compression and dimensionality reduction. Among the various techniques, the CP (Candecomp/Parafac) and Tucker decompositions emerge as the most widely adopted methods for achieving this goal. These methods play a crucial role in applications where handling high-dimensional data efficiently is paramount, providing a formalized approach to tensor analysis and manipulation.
Table 1 provides a summary of key tensor concepts and notation [12].

Table 1: Tensor terminology and notation

Tensor Concept	Definition
Order	Number of dimensions (modes/ways) of a tensor
Rank	Minimum number of rank-one tensors required to generate a tensor
Mode	Dimensionality of a tensor, analogous to matrix rows/columns
Fibers	Higher-order analogue of matrix rows/columns
Slices	2D sections of a tensor obtained by fixing all but two indices

Tensor multiplication	Generalization of matrix multiplication to tensor contraction
Kronecker product	Tensor product of two matrices
Mode-n product	Multiplication of a tensor by a matrix in mode-n

CP Decomposition

The CANDECOMP/PARAFAC (CP) decomposition, developed in the 1970s, expresses a tensor as the sum of rank-one component tensors. Mathematically, for an Nth order tensor X, the CP decomposition is:

$$X = \sum Rr = 1\lambda rar \circ br \circ cr \circ \ldots$$

Where λr are weights, and ar, br, cr, ... are factor vectors. R is the rank, i.e. number of components. The symbol ∘ denotes vector outer product. Each component tensor on the right-hand side is the outer product of N factor vectors, hence rank-one [13].

The CP model can be interpreted as expressing the higher-order tensor X as a sum of R rank-one tensors. Each component captures a latent pattern in the data, formed by the outer product of factors along each mode [14]. CP decomposition thus provides a compressed representation via dimensionality reduction, with the factors containing the essential information.

Estimating the CP model involves computation of the factor matrices A, B, C such that the model best approximates the original tensor in a least squares sense. Various algorithms exist for fitting the CP model, such as alternating least squares and gradient-based methods. The model optimization can be sensitive to initialization and may converge to local optima [15].

The key advantages of CP decomposition are its simplicity, unique decomposition under mild conditions, and interpretability. The uncompressed form directly provides the latent patterns and relationships. It is also scalable to large datasets. However, difficulty in computation and non-uniqueness of solutions can be limitations for some applications.

Tucker Decomposition

The Tucker decomposition expresses a tensor via a core tensor transformed by a matrix along each mode. For a 3rd order tensor X, the Tucker model is:

$$X = G\ x1\ A\ x2\ B\ x3\ C$$

Where G is the core tensor, and A, B, C are factor matrices for the three modes. The core tensor captures the interaction between the factors. The number of components in each mode is given by the dimensionality of the corresponding factor matrix.

Unlike CP, the Tucker decomposition is not unique. Different rotations of the factor matrices can generate the same reconstructed tensor. A variant called Higher-order SVD (HOSVD) provides a structured Tucker model by using singular vectors of the mode-n unfoldings as factor matrices.

Estimating the Tucker model involves optimization of the core tensor and factor matrices [16]. Algorithms like higher-order orthogonal iteration and alternating least squares are commonly employed. Tucker models can be computed more efficiently

than CP, but are less interpretable. The core tensor may also have higher storage costs than CP for sparse data.

The Tucker decomposition provides a more flexible model than CP, with the core tensor capturing interactions. It remains interpretable via the factors, and can handle sparse, incomplete datasets. However, non-uniqueness and rotational freedom make the results more dependent on algorithm initialization and design choices.

Applications and Examples

Tensor decompositions have emerged as powerful tools with diverse applications across various domains. Their ability to analyze complex, multidimensional data has been particularly valuable in fields such as chemometrics, signal processing, neuroscience, web mining, computer vision, and recommender systems. In social network analysis, where the data forms a 3-way tensor with users, friends, and interactions, Canonical Polyadic (CP) and Tucker models have proven effective in uncovering latent communities through their respective latent factors [17]. Additionally, tensor regression based on decomposed features facilitates the prediction of links within the network. In the realm of neuroimaging, tensor decomposition has been applied to fMRI data, which typically involves four modes - voxels, time, subjects, and conditions [18]. CP extraction reveals spatially distributed, task-related source signals, while Tucker models identify local brain regions and interactions. The versatility of tensor decompositions extends to recommender systems, where user-item ratings are represented as a user x item x context tensor. Tensor factorization enhances recommendation accuracy by incorporating multidimensional effects, surpassing the limitations of traditional matrix-based methods.

Text mining benefits from tensor decompositions as well, particularly in the extraction of latent topics and their corresponding word distributions from document-term matrices. Furthermore, joint analysis of text and citations using tensors has been shown to improve topic coherence [19]. In computer vision, tensor methods prove valuable for extracting intrinsic image features for recognition and classification, with Tucker models adept at learning multilinear transformations for achieving view and illumination invariance.

Chemometrics, a field dealing with the analysis of chemical data, utilizes CP to resolve chemical mixtures from spectrometric data by representing factors as pure component spectra and concentrations. Meanwhile, Tucker models excel in identifying chemical interactions within complex datasets. To exemplify the practical applications of tensor decompositions, CP and Tucker models were applied to two distinct datasets [20]. The Enron Email Network, a dataset characterized by email senders, receivers, and time modes, underwent CP decomposition to identify latent communities and their temporal interaction patterns. The results, as presented in Table 2, showcase the effectiveness of tensor decomposition in revealing hidden structures within the data [21]. Similarly, the application of Tucker models to the Enron Email Network extracted key actors and their connections within each community. In the context of a brain fMRI study, the dataset comprising voxels, timepoints, and stimuli underwent CP decomposition to cleanly separate spatially distributed patterns of brain activity corresponding to each stimulus, as outlined in Table 3. Concurrently, Tucker models were employed to find localized regions and their interactions in response to

different stimuli. These examples underscore the ability of CP and Tucker models to derive latent patterns and interactions from multidimensional datasets, offering interpretable structure and dimensionality reduction [22].

Tensor decompositions, through CP and Tucker models, have demonstrated their efficacy in handling diverse and complex datasets across various domains. The ability to uncover latent patterns, communities, and interactions makes tensor methods scalable and well-suited for mining big data, particularly in situations involving large, incomplete datasets where traditional methods fall short [23].

Table 2: CP decomposition on Enron email tensor

Community 1	Community 2	Community 3	
Senior Executives	Traders	Legal Department	
Temporal Trends	Declining over Time	Spike During Crisis	Peaks on Quarter End

Table 3: Tucker decomposition on fMRI tensor

Brain Region A	Brain Region B	Brain Region C
Activates for Stimulus 1	Activates for Stimulus 2	Activates for Stimulus 3
Interacts with Region B	Interacts with Region A	Interacts with Region C

Computational Aspects

Tensor decompositions, although powerful for extracting latent features from multidimensional data, present intricate computational and algorithmic challenges that demand careful consideration. The scalability of tensor methods is a primary concern, particularly as they exhibit poor performance when confronted with high dimensionality and large datasets. The handling of massive data necessitates the utilization of high-performance computing resources, making scalability a critical aspect of tensor decomposition algorithms [24]. Another challenge arises from the uniqueness and rotations associated with tensor decompositions. The Canonical Polyadic (CP) decomposition lacks a guarantee of uniqueness, while the Tucker decomposition allows rotational freedom. Efficient algorithms must be designed to impose constraints on the solutions, ensuring meaningful and interpretable results. Addressing the issues of uniqueness and rotations is crucial for enhancing the reliability and applicability of tensor decomposition methods.

Initialization and convergence represent additional hurdles in the application of tensor decomposition techniques. The quality of the obtained results heavily depends on the initialization process, and algorithms may become trapped in local optima. To mitigate these challenges, it is imperative to develop robust initialization strategies to guide the algorithm toward optimal solutions. Overcoming convergence issues is essential for ensuring the efficiency and effectiveness of tensor decomposition algorithms [25]. Sparse and missing data pose significant challenges in the context of real-world datasets. Many real-world datasets exhibit sparsity, with numerous zero entries. Tensor decomposition algorithms must be designed to handle sparse and missing data effectively, ensuring that the presence of zeros does not compromise the accuracy and reliability of the extracted latent features.

Model selection is a critical aspect of tensor decomposition, requiring careful consideration during the design phase. Choosing the appropriate rank and determining the number of components significantly impacts the performance and interpretability of the model [26]. Rigorous validation processes are essential to make informed decisions about model parameters, ensuring that the tensor decomposition accurately captures the underlying structure of the data. Several software tools have been developed to facilitate the implementation of various tensor algorithms. Notable examples include the Tensor Toolbox in MATLAB, TensorFlow, PyTorch, and TensorLy. These tools provide efficient and scalable implementations, allowing researchers and practitioners to apply tensor decomposition methods to diverse datasets. Leveraging these software tools is crucial for overcoming the computational challenges associated with tensor decompositions and enhancing the accessibility of these techniques to a broader audience.

Despite these challenges, tensor decompositions remain exceptionally useful for extracting latent structures from complex, multidimensional datasets. Ongoing research efforts are focused on addressing the aforementioned challenges to improve the scalability, uniqueness, and interpretability of tensor models. The ability of tensor methods to fuse information from diverse modes positions them as invaluable tools for mining insights from large-scale, complex data. As advancements in computational methods continue, tensor decompositions are expected to play a pivotal role in unraveling the hidden patterns within massive and intricate datasets, contributing to the advancement of various scientific and industrial domains.

Conclusion

Tensor decompositions serve as a robust and efficient framework for revealing latent patterns within multidimensional big data. This method involves representing higher-order arrays as multilinear models, facilitating compression and dimensionality reduction while preserving crucial structural information. One prominent technique within tensor decomposition is the Canonical Polyadic (CP) decomposition, which stands out for its ability to extract interpretable components and interactions. Through CP decomposition, the original tensor is approximated as a sum of rank-one tensors, each corresponding to a unique component [27]. This decomposition not only aids in reducing the dimensionality of the data but also provides insights into the underlying structures and relationships present in the dataset. In addition to CP decomposition, the Tucker decomposition represents another noteworthy approach in tensor analysis. Unlike CP, Tucker allows for more flexibility by introducing a core tensor that models interactions among different modes of the original tensor. The core tensor captures the shared information among these modes, offering a more nuanced representation of complex relationships within the data. The versatility of the Tucker decomposition lies in its ability to adapt to various data structures, making it a valuable tool for uncovering hidden patterns in real-world datasets with diverse tensor structures.

The combination of CP and Tucker decompositions enhances the analytical capabilities for extracting meaningful insights from multidimensional data. While CP excels in isolating individual components and their interactions, Tucker provides a broader perspective by considering the interplay between different modes. This synergy creates a comprehensive toolkit for researchers and practitioners seeking to understand the complex relationships embedded in large-scale datasets [28]. Moreover, tensor decompositions find practical applications in a wide range of fields, such as image and signal processing, neuroscience, and social network analysis. In

image processing, for example, tensors can represent multi-dimensional pixel arrays, and tensor decompositions aid in extracting meaningful features and patterns. Similarly, in neuroscience, where data often exhibits complex interactions across multiple dimensions, tensor decompositions can reveal hidden structures and relationships, contributing to a better understanding of brain function [29].

The utility of tensor decompositions becomes particularly evident in the context of big data, where datasets are characterized by high dimensionality and intricate interdependencies. By employing tensor decomposition techniques, analysts can effectively reduce the dimensionality of these datasets, facilitating more manageable and interpretable analyses. The inherent ability of tensor decompositions to capture essential information while discarding redundant details is crucial for handling the challenges posed by the ever-increasing size and complexity of contemporary datasets [30]. We have highlighted diverse applications of tensor methods in areas ranging from neuroscience to social networks and recommender systems. These demonstrate their effectiveness in finding low dimensional structure and performing knowledge discovery on large datasets with complex interactions between modes. Ongoing algorithmic advances are improving the capability to handle massive, sparse data. This makes tensor decompositions extremely valuable in the current era of multidimensional big data across domains [31].

Tensor methods enable mining latent relationships and interactions that often escape conventional matrix-based analysis. By leveraging tensor algebra and decompositions, we can uncover previously unknown structures and insights in large, multidimensional datasets. This helps derive actionable intelligence and supports predictive modeling. Tensor decompositions will continue to be indispensable tools for extracting knowledge from the wealth of big data being generated in diverse settings.

References

[1] F. Cannarile, M. Compare, F. Di Maio, and E. Zio, “A clustering approach for mining reliability big data for asset management,” *Proc. Inst. Mech. Eng. O. J. Risk Reliab.*, vol. 232, no. 2, pp. 140–150, Apr. 2018.

[2] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Context-aware query performance optimization for big data analytics in healthcare,” in *2019 IEEE High Performance Extreme Computing Conference (HPEC-2019)*, 2019, pp. 1–7.

[3] K. M. L. Jones, A. Rubel, and E. LeClere, “A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics,” *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 10, pp. 1227–1241, Oct. 2020.

[4] A. Hsu, W. Khoo, N. Goyal, and M. Wainstein, “Next-generation digital ecosystem for climate data mining and knowledge discovery: A review of digital data collection technologies,” *Front. Big Data*, vol. 3, p. 29, Sep. 2020.

[5] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications surveys & tutorials*, 2015.

[6] Y. Wang, B. Wang, and Y. Huang, “Comprehensive analysis and mining big data on smart E-commerce user behavior,” *J. Phys. Conf. Ser.*, vol. 1616, no. 1, p. 012016, Aug. 2020.
[7] X. Zhao, “A study on the application of big data mining in e-commerce,” in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2018.
[8] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Approximate query processing for big data in heterogeneous databases,” in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 5765–5767.
[9] W. S. Chen and Y. K. Du, “Using neural networks and data mining techniques for the financial distress prediction model,” *Expert Syst. Appl.*, 2009.
[10] J. Pei, “Some New Progress in Analyzing and Mining Uncertain and Probabilistic Data for Big Data Analytics,” in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 2013, pp. 38–45.
[11] S. Batistič and P. der Laken, “History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations,” *Br. J. Manag.*, vol. 30, no. 2, pp. 229–251, Apr. 2019.
[12] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, “Federated query processing for big data in data science,” in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 6145–6147.
[13] D. B. Ventura, “Exploring the Perceptions, Influences, and Sociodemographic Determinants of Sustainable Fashion among Consumers in Colombia,” *IJRAI*, vol. 5, no. 3, pp. 1–14, Mar. 2015.
[14] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, 2014.
[15] C. K.-S. Leung, R. K. MacKinnon, and F. Jiang, “Reducing the search space for big data mining for interesting patterns from uncertain data,” in *2014 IEEE International Congress on Big Data*, 2014, pp. 315–322.
[16] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” *SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Sep. 2017.
[17] J. K.u and J. M.David, “Issues, Challenges and Solutions : Big Data Mining,” in *Computer Science & Information Technology (CS & IT)*, 2014.
[18] C. K. S. Leung and F. Jiang, “A data science solution for mining interesting patterns from uncertain big data,” *International Conference on Big Data …*, 2014.
[19] A. Lieto, C. Battaglino, D. P. Radicioni, and M. Sanguinetti, “A Framework for Uncertainty-Aware Visual Analytics in Big Data,” *CEUR Workshop Proc.*, vol. 1510, pp. 146–155, Nov. 2015.
[20] A. Nassar and M. Kamal, “Machine Learning and Big Data Analytics for Cybersecurity Threat Detection: A Holistic Review of Techniques and Case Studies,” *Intelligence and Machine Learning …*, 2021.
[21] Anjala, “Algorithmic assessment of text based data classification in big data sets,” *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. SP4, pp. 1231–1234, Mar. 2020.
[22] J. Liu, J. Li, W. Li, and J. Wu, “Rethinking big data: A review on the data quality and usage issues,” *ISPRS J. Photogramm. Remote Sens.*, vol. 115, pp. 134–142, May 2016.

[23] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Big data in cloud computing review and opportunities," *arXiv preprint arXiv:1912.10821*, 2019.
[24] J. J. Horton and P. Tambe, "Labor economists get their microscope: Big data and labor market analysis," *Big Data*, vol. 3, no. 3, pp. 130–137, Sep. 2015.
[25] G. Ilieva, T. Yankova, and S. Klisarova, "Big data based system model of electronic commerce," *Trakia Journal of Science*, vol. 13, no. Suppl.1, pp. 407–413, 2015.
[26] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015.
[27] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, and S. Fosso-Wamba, "The role of Big Data in explaining disaster resilience in supply chains for sustainability," *J. Clean. Prod.*, vol. 142, pp. 1108–1118, Jan. 2017.
[28] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *Miss. Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.
[29] F. Jiang and C. K. Leung, "A Data Analytic Algorithm for Managing, Querying, and Processing Uncertain Big Data in Cloud Environments," *Algorithms*, vol. 8, no. 4, pp. 1175–1194, Dec. 2015.
[30] D. Agrawal, P. Bernstein, E. Bertino, and S. Davidson, "Challenges and opportunities with Big Data 2011-1," 2011.
[31] A. Nassar and M. Kamal, "Ethical Dilemmas in AI-Powered Decision-Making: A Deep Dive into Big Data-Driven Ethical Considerations," *IJRAI*, vol. 11, no. 8, pp. 1–11, 2021.