The Role of Artificial Intelligence in Data Analytics

Pranay Mungara

Abstract: ***Recent advances in sensor networks and the Internet of Things (IoT) technologies have led to the gathering of an enormous scale of data. The exploration of such huge quantities of data needs more efficient methods with high analysis accuracy. Artificial Intelligence (AI) techniques such as machine learning and evolutionary algorithms able to provide more precise, faster, and scalable outcomes in big data analytics. Despite this interest, as far as we are aware there is not any complete survey of various artificial intelligence techniques for big data analytics. The present survey aims to study the research done on big data analytics using artificial intelligence techniques. The authors select related research papers using the Systematic Literature Review (SLR) method. Four groups are considered to investigate these mechanisms which are machine learning, knowledge - based and reasoning methods, decision - making algorithms, and search methods and optimization theory. A number of articles are investigated within each category. Furthermore, this survey denotes the strengths and weaknesses of the selected AI - driven big data analytics techniques and discusses the related parameters, comparing them in terms of scalability, efficiency, precision, and privacy. Furthermore, a number of important areas are provided to enhance the big data analytics mechanisms in the future.***

Keywords: Artificial Intelligence, Data analytics, Big data, IoT

1. Introduction

The fourth Industrial Revolution (I4.0) brought out many disruptive technologies that have substantially transformed existing systems and brought up new business models and processes across diverse domains and sectors [**1**]. Alongside, decision - making has increasingly been decentralized to computer systems, notably with the dawn of Artificial Intelligence (AI) and Big Data Analytics (BDA) [2]. AI mimics human intelligence and cognitive abilities with the assistance of machines or computers [3]. In [4] contend that AI can make decisions that are sometimes superior to those made by humans. Big Data refers to the enormous volume of data that cannot be managed and processed by conventional data management methods [5]. BDA refers to analyzing such massive datasets to emanate actionable insights and value [6]. Executing AI and BDA delivers a competitive edge to an organization with tangible growth prospects in its social and corporate status [7]. It is now possible to derive actionable insights from the billions of datasets sired every minute using AI - powered tools [8]. AI, Machine Learning (ML), and supporting BDA are growing at an unprecedented rate, opening up preposterous opportunities to enrich the performance of various industries, research capabilities, and businesses.

In line with the UN SDG#12 goal of achieving sustainable production and consumption through scientific, innovative and technological capabilities, emphasized the applications of AI in sustainable business models (SBMs). To achieve sustainable growth and development in a disruptive environment, many organizations have already started adapting to AI and/or other tech - driven solutions [9]. Demonstrate how AI has been applied in the sustainable agri - food industry and their supply chains (AI for sorting of food, supply - chain optimization, food security, hygiene standard, and food and drink preparations). In [10] compile literature and exhibit how wireless networking devices can improve agricultural activities. By connecting UN SDG #3 (Healthy lives and well - being for all), In [11] show how BDA and visualization can be applied in medical research to achieve sustainable healthcare. In a similar line, [12] show the recent sustainable technologies for older adults to promote healthy ageing and the social inclusion of the elderly using a smart environment. In [13] show how open innovation, e - market places, and sustainability are inter - connected from sustainable logistics using I4.0 and big data environment. Similarly, in [14] explored the literature on last - mile logistics optimization techniques and found that machine learning techniques and mixed methods are widely used in sustainable or smart logistics. On the other hand, in [15] demonstrated how digital technologies would have significant environmental impacts due to high carbon dioxide emissions associated with the energy consumption required to generate and process large amounts of data; mineral extraction for, and manufacturing of, technological components; and e - waste. In [16] identified that AI, IoT, circular economy, BDA, AR, and VR emerged as major trends in tourism research, broadening our understanding of how technology can shape the future of sustainable tourism.

Also argue that ‘Technological Intelligence, ’ the capacity to relish and adapt to technological advancements, is compulsory for all businesses. The machine - readable data furlough and steadily rising computation power and storage capacity have significantly impacted many sectors. Hence, businesses must dig through the scope of AI and BDA applications and map out their apparent vulnerability to disruptions. So far, many domains have opened up to AI applications over the past decade. Such applications include AI, ML, and Deep Reinforcement Learning (DRL) for Smart city and sustainable operations [17]; Supply chain management [18]; Medical diagnosis and treatment; Pandemic response and prediction modelling [19]; Engineering and constructions safety Power quality assurance; Commercial banking and stock market predictions; Enhanced marketing and customer experience, and many others. Similarly, BDA, with various AI algorithms, improved computing power and cloud storage, improved decision - making quality and added new value to various fields. These emerging fields of BDA study include smart logistics and IoT, precision agriculture and smart farming; sustainable architecture; consumer analytics and marketing transformation and many diverse areas [20].

2. Literature Review

This Special Issue consists of fourteen articles covering different aspects of machine learning and artificial intelligence.

This study focuses on creating a machine learning model that can predict the likelihood of chronic kidney disease using publicly available data [21]. The data underwent several preprocessing steps, including the imputation of missing values, balancing through the SMOTE algorithm, and scaling of features. The chi - squared test was utilized to select the most relevant and highly correlated features. The machine learning model was built using a combination of supervised learning techniques, with support vector machine (SVM) and random forest (RF) achieving the lowest false - negative rate and highest test accuracy of 99.33% and 98.67%, respectively. SVM was found to perform better than RF upon validation through 10 - fold cross - validation.
This study represents the first attempt to examine selected design research publications using a sophisticated method called "text mining" [22]. This method generates results based on the presence of specific research terms (i. e., keywords), which provides a more reliable outcome compared to other approaches that rely on contextual information or authors' perspectives. The primary objective of this research is to increase awareness and understanding of design research, and to identify potential future research directions by addressing gaps in the literature. Based on the literature review, it can be concluded that the field of design research still lacks a unifying theory. Text mining, with its features, enhances the validity and generalizability of the results compared to other methods in the literature. The text mining technique was applied to collect data from 3553 articles from 10 journals, utilizing 17, 487 keywords. This research explores new topics in the field of design concepts, drawing the attention of researchers, practitioners, and journal editorial boards. The key categories analysed and presented in this paper provide insights into the growth and decline in various fields in the domain of design.

This paper presents a novel deep learning approach for detecting student emotions [23]. The main objective of the study is to explore the relationship between teaching practices and student learning, based on emotional impact. The system uses facial recognition algorithms to gather information from online platforms and image classification techniques to identify the emotions of students and teachers. Two deep learning models are compared for their performance, and the results show promising outcomes, as discussed in the Experimental Results section. The proposed system is validated using an online course with students, and the results indicate that the technique operates effectively. Various deep learning techniques are applied for emotional analysis, including transfer learning for a pre - trained deep neural network, which increases the accuracy of the emotion classification stage. The results of the experiment demonstrate that the proposed method is promising, as discussed in the Experimental Results section.

This paper proposes a deep learning solution for detecting masks worn in public to prevent the spread of coronavirus [24]. The system, designed for real - time use with a webcam, utilizes an ensemble method for high accuracy and improved detection speed. Transfer learning on pre - trained models and rigorous testing on objective data resulted in a dependable and cost - effective solution. The findings indicate the effectiveness of the solution in real - world settings, contributing to pandemic control. Compared to existing methods, the proposed solution achieves improved accuracy and performance metrics, such as specificity, precision, recall, and F measure, in three - class outputs. A careful balance is maintained between the number of parameters and processing time.

This study proposes a deep learning method for the classification and analysis of scientific literature using convolutional neural networks (CNNs) [25]. The research is divided into three dimensions, publication features, author features, and content features, with explicit and implicit features forming a set of scientometric terms. The CNN model uses weighted scientometric term vectors to achieve dual - label classification of literature based on its content and methods. The study showcases the effectiveness of the proposed model through an application example from data science and analytics literature, with results showing improved precision, recognition, and F1 score compared to other machine learning classification methods. The proposed scientometric classification model also exhibits higher accuracy than deep learning classification using only explicit and dominant features. This study offers a guide for fine - grained classification of scientific literature and provides insight into its practical application.

This research aims to help science students identify butterfly species without causing harm to the insects during analysis [26]. The study employs transfer learning with neural network models to classify butterfly species based on images. The dataset consists of 10, 035 images of 75 butterfly species and 15 unusual species were selected for the study, with various orientations, photography angles, lengths, and backgrounds. The imbalanced class distribution in the dataset resulted in overfitting, which was addressed with data augmentation. Transfer learning was applied using several convolutional neural network architectures, including VGG16, VGG19, MobileNet, Xception, ResNet50, and InceptionV3. The models were evaluated based on precision, recall, F measure, and accuracy. The results showed that the InceptionV3 architecture provided an accuracy of 94.66%, which was superior to all other architectures. This work proposes a new approach for identifying glaucoma from fundus images using a deep belief network (DBN), optimized by the elephant - herding optimization (EHO) algorithm [27]. The system is designed to be tested on various datasets, which can help to improve the accuracy of glaucoma diagnosis.

This paper examines 66 machine learning models using a two - stage evaluation process [28]. The evaluation was performed on a real - world dataset of European credit card frauds and used stratified K - fold cross - validation. Out of 330 evaluation metrics, the All K - Nearest Neighbors (AllKNN) undersampling technique with CatBoost (AllKNN–CatBoost) was found to be the best model, achieving an AUC of 97.94%, recall of 95.91%, and F1 score of 87.40%. The AllKNN–CatBoost model was compared to relevant studies and was found to outperform previous models.

This research presents a hybrid data analytics framework that combines convolutional neural networks and bidirectional long short - term memory (CNN - BiLSTM) to examine the effect of merging news events and sentiment analysis with financial data on stock trend prediction [29]. Two real - world case studies were conducted using data from the Dubai Financial Market between 1 January 2020 and 1 December 2021, in the real estate and communications sectors. The results demonstrate that incorporating news events and sentiment analysis with financial data improves the accuracy of stock trend prediction. The CNN–BiLSTM model achieved an improvement of 11.6% in the real estate sector and 25.6% in communications compared to benchmarked machine learning models.

This study introduces a four - layer model and proposes a hybrid imputation method (HIMP) for filling in multi - pattern missing data, including non - random, random, and completely random patterns [30]. HIMP starts by imputing non - random missing data patterns and then dividing the resulting dataset into two datasets with random and completely random missing data patterns. Next, different imputation methods are applied to each dataset based on the missing data pattern. The final dataset is created by merging the best imputed datasets from random and completely random patterns. The effectiveness of HIMP was evaluated using a real dataset named IRDia that had all three missing data patterns. HIMP was compared to other methods using accuracy, precision, recall, and F1 score with different classifiers, and the results showed that HIMP outperformed other methods in imputing multi - pattern missing values.

This paper presents a new Whale Optimization Algorithm (EWOA) to solve Optimal Power Flow (OPF) problems, with the aim of improving exploration capability and maintaining a balance between exploration and exploitation [31]. The movement strategy of whales in the EWOA is improved through the introduction of two new techniques: (1) encircling the target using Levy motion and (2) searching for the target using Brownian motion, which work in conjunction with the traditional bubble - net attacking method. To evaluate the performance of EWOA - OPF, it is compared with six well - known optimization algorithms in solving both single - and multi - objective OPF problems under system constraints. The comparison results show that the EWOA - OPF outperforms the other algorithms and provides better solutions for both single - and multi - objective OPF problems.

In this review, the authors examine the advancements and applications of the Harris Hawk Optimizer (HHO), a robust optimization technique that has gained popularity in recent years [32]. Through experiments conducted on the Congress on Evolutionary Computation (CEC2005) and CEC2017, HHO is compared to nine other state - of - the - art algorithms, showing its efficacy and effectiveness. The paper provides a comprehensive overview of HHO and delves into future directions and areas for further investigation of new variants of the algorithm and its widespread use.

This paper provides a comprehensive overview of effective communication techniques for space exploration of ground, aerial, and underwater vehicles [33]. The study not only summarizes the challenges faced in trajectory planning, space exploration, optimization, and other areas, but also highlights the future directions for research. Aiming to fill the gap in the literature for those interested in path planning, this paper includes optimization strategies for terrestrial, underwater, and airborne applications. The study covers numerical, bio - inspired, and hybrid methodologies for each dimension discussed. The goal of this paper is to establish a centralized platform for publishing research on autonomous vehicles on land and their trajectory optimizations, airborne vehicles, and underwater vehicles.

3. AI - Driven Big Data Analytics Mechanisms

Classification and review of the selected big data analysis studies are performed based on the AI subfields used in big data analytics. Figure 1 shows the taxonomy of the big data analytics techniques based on the AI subfields, and categorizes the articles investigated in this survey within those categories. The presented taxonomy has four main categories, including machine learning, knowledge - based and reasoning methods, decision - making algorithms, and search methods and optimization theory [34].

Paper ID: SR24506151808 DOI: https://dx.doi.org/10.21275/SR24506151808

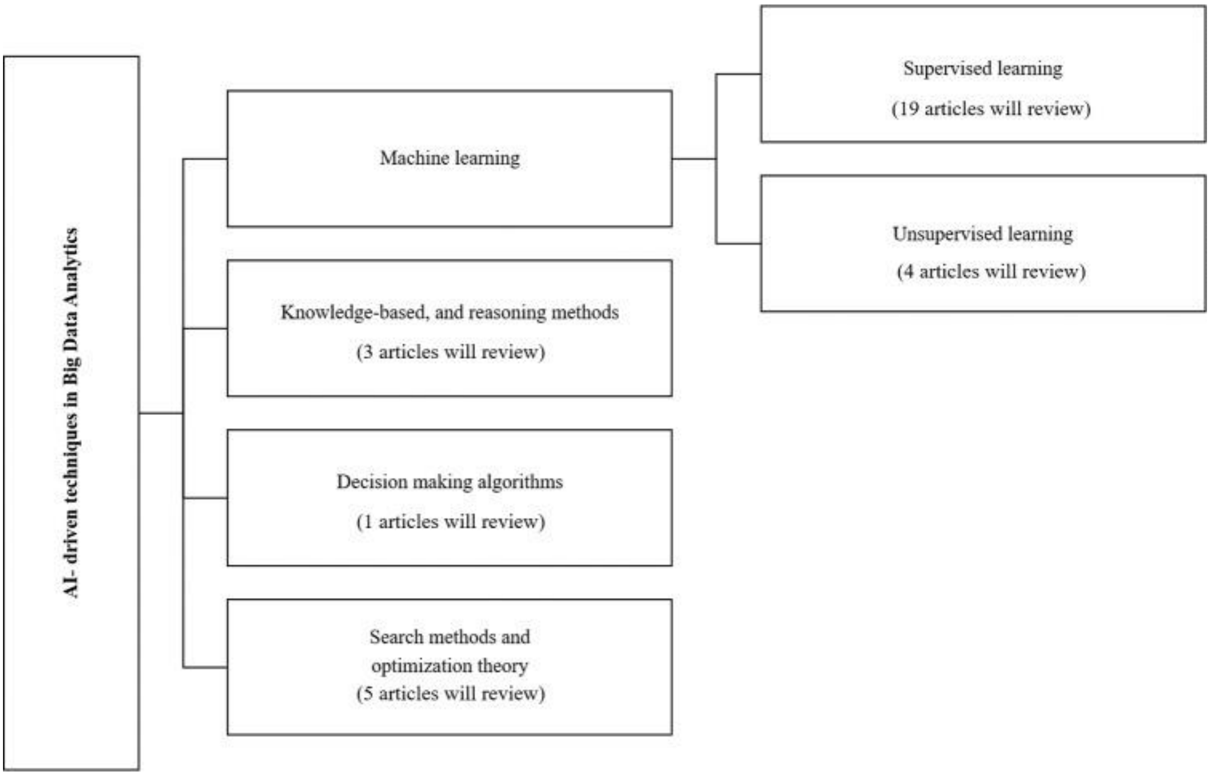

Figure 1: Schematic diagram of classification of AI

Furthermore, the four most significant qualitative parameters are defined to assess each big data analysis method and recognize its benefits and drawbacks, as follows:

- **Scalability**: The mechanism's ability to adapt to rapid changes without compromising the quality of the analysis.
- **Efficiency**: It denotes the ratio of the method to the overall time and cost need.
- **Precision**: This is detected with various parameters like data errors, and the predictive ability of algorithms.
- **Privacy**: It defines the practices which safeguard that the data is only used for its intended purpose.

The papers are overviewed and compared with mechanism goals in the last step.

3.1 Machine learning mechanisms

Machine learning algorithms can be divided into two main classes including supervised learning and unsupervised learning. The first class needs a lot of manual effort to put the data in a proper format to learn algorithms. The unsupervised learning algorithms can discover hidden patterns in huge amounts of unlabelled data.

Supervised learning

The aim of a supervised learning algorithm is to forecast the right label for newly presented input data using another dataset. In this learning method, a set of inputs and outputs is presented and the relation among them is found while training the system. The main objective of supervised learning is to model the dependency between the input features and the target prediction outputs. As shown in Fig.2, input examples are categorized into a known set of classes [35].

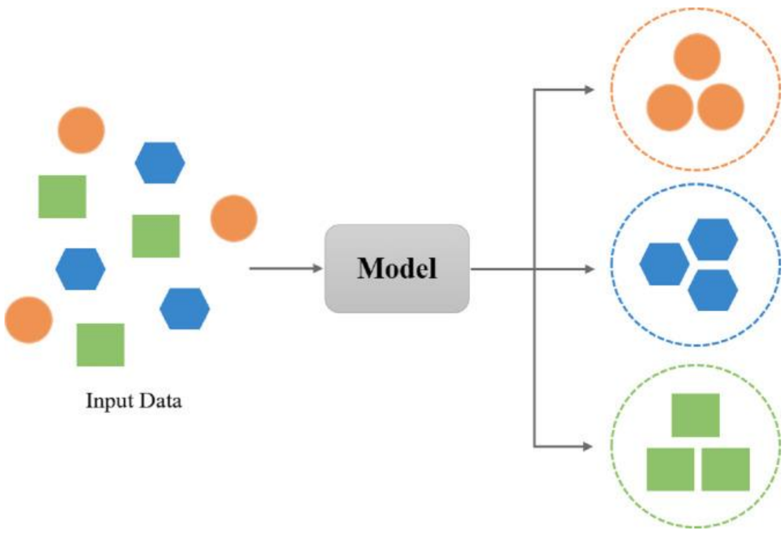

Figure 2: Overview of supervised learning.

In [36] proposed a novel platform for fraud detection named, Scalable Real - time Fraud Finder (SCARFF). The proposed platform uses Kafka, Spark, and Cassandra big data tools along with a machine learning technique to process streaming data. The machine learning engine composed of a weighted ensemble that employs two types of classifiers based on random forest. It deals with imbalanced data, non - stationarity, and feedback latency. The results indicate that the efficiency, accuracy, and scalability of the presented framework is satisfactory over a big stream of transactions.

In [37] presented a predictive approach on demonetization data using a support vector machine, called PAD - SVM. Preprocessing, descriptive analysis, and prescriptive analysis are three stages of the proposed PAD - SVM. Cleaning the data, handling the missing data fields, and splitting the essential data from the tweets are performed in the first stage. Identifying the most influential individual and performing analytical functionalities are two key functions of the descriptive analysis stage. Semantic analysis is also performed in the second stage. The present mindset of people and the reaction of society to the problem is predicted using predictive analysis. The authors performed a series of experiments and confirmed the performance of the proposed method in terms of execution time and classification error.

In [38] proposed several data mining and deep learning methods for visualization and trend prediction of criminal data. The authors discovered various interesting facts and patterns from the criminal data of San Francisco, Chicago, and Philadelphia datasets. The proposed method has lower complexity in comparison with LSTM. Based on the predictive results of the article, the superior performance of the Prophet model and Keras stateful LSTM is confirmed as compared to traditional neural networks.

Accurate and timely forecasting popularity of television programs is of great value for content providers, advertisers, and broadcast television operators. Traditional prediction models require a huge amount of samples and long training time, and the precision of predictions for programs with high peaks or severe decrease in popularity is poor. Proposed an enhanced prediction method based on trend detection. The authors used a random forest model after clustering the trends using the Dynamic Time Wrapping (DTW) algorithm and K - medoids clustering. For new programs, the GBM classifier applied to assign them to the existing trends. According to the trial outcomes, the introduced model obtains better prediction results with a combination of prediction values from the trend - specific models and classification probabilities. The results also revealed that the forecasting period is effectively reduced compared to the current forecasting methods.

Big data produced by social media is a great opportunity to extract valuable insights. With the growth of the data size, distributed deep learning models are efficient for analysing social data. Henceforth, it is essential to improve the performance of deep learning techniques. In [39] presented a novel efficient technique for sentiment analysis. The authors tried to adopt fastText with Recurrent Neural Network (RNN) variants to represent and classify textual data. Furthermore, a distributed system based on distributed machine learning has been proposed for real - time analytics. The performed trials prove that the presented method outperforms Long Short - Term Memory (LSTM), Bidirectional Long Short - Term Memory (BiLSTM), and Gated Recurrent Unit (GRU) methods in terms of classification accuracy. Also, it can handle large scale data for sentiment analysis.

Nowadays, the urban network has produced a huge amount of data. Therefore, some security challenges arise because of the private data gathering by smart devices. In [40] tried to discover the abnormal behaviour of insiders to avoid urban big data leakage. The authors developed various deep learning methods to analyse deviations among realistic actions and the normalcy of daily activities. Abnormal activities are recognized using a Multi - Layer Perceptron (MLP) based on the computed deviations. According to the trial outcomes, the proposed method can learn the normal pattern of behaviours and identify abnormal activities with high precision.

Internet traffic is growing rapidly in the age of multimedia big data. Therefore, data processing and network overload are two key challenges in this context. Proposed a hybrid - stream model to solve these challenges for video analysis. It contains data preprocessing, data classification, and data - load - reduction modules. A modified version of the CNN method is developed to evaluate the importance of each video frame to improve classification accuracy. The outcomes confirmed that the proposed model reduces data load, controls the video input size, and decreases the overload of the network. The outcomes also confirmed the effective reduction of processed video without compromising the quality of experience. Also, it observed that the model has a good performance for the continuous growth of large multimedia data as compared to other traditional models.

In [41] proposed a novel model for smart healthcare information systems using machine learning algorithms. The proposed model includes four layers. The data source layer handles heterogeneous data sources. The data storage layer manages the storage optimization process. Various techniques like indexing and normalization have been used to make optimal use of system resources. Different data security and privacy techniques such as data masking, granular control over data access, activity monitoring, dynamic encryption, and endpoint validation are used in the data security layer. Finally, machine learning methods used in the application layer for early diagnosis of the disease. Based on the trial outcomes of the article, the accuracy of the proposed model improved by using fuzzy logic and information theory.

In [42] introduced a novel health status prediction system by applying machine learning models on big data streams. The presented system built using Apache Spark and deployed in the cloud environment. The user sends his health qualities and the system forecasts the user's health status in real - time. A decision tree model is created from the existing healthcare data and applied to streaming data for health status prediction. The presented architecture leads to the time and cost - efficiency of the introduced system. The privacy of data is overcome by using a secondary Twitter account.

In [43] developed new big data technologies and machine learning methods to identify diabetes disease. First, the data

Paper ID: SR24506151808 DOI: https://dx.doi.org/10.21275/SR24506151808

is gathered from a huge data set, and the MapReduce model is used to efficiently combine the small chunk of data. Then, the normalization procedure is used to eliminate the noise of the collected data. Also, an ant bee colony algorithm is applied to select the statistical features. The chosen features are trained using the SVM with a multilayer neural network. The associated neural network is applied to classify the learned features. The results revealed that the SVM neural network provides high accuracy, sensitivity, and less error rate.

Detection of COVID - 19 based on the analysis of chest X - ray and Computed Tomography (CT) scans, has attracted the attention of researchers. COVID - 19 medical scans analysis using machine learning algorithms provides an automated and effective diagnostic tool. In [44] proposed a multi - task pipeline model based on deep neural networks for COVID - 19 medical scans analysis. An Inception - v3 deep model fine - tuned using multi - modal learning in the first stage. A Convolutional Neural Network (CNN) architecture is used to identify three types of manifestations in the second stage. Transfers learning from another domain of knowledge to generate binary masks for segmenting the regions related to these manifestations are performed in the last stage. Based on the trial results, the proposed framework enhances efficiency in terms of computational time. Furthermore, the proposed system has higher accuracy compared to the recent literature.

4. Conclusion

The state - of - the - art mechanisms in the field of big data analytics is surveyed in this article. According to the performed study, we introduced a taxonomy for AI - driven big data analytics mechanisms. The selected 32 articles are investigated in four main categories including machine learning, knowledge - based and reasoning methods, decision - making algorithms, and search methods and optimization theory. The advantages and disadvantages of each of these mechanisms have been investigated. The machine learning - based mechanisms use a learning method to adapt the automated decisions. Efficiency and precision as the major factors are improved in most of the machine learning - based mechanisms. However, the use of incomplete and inconsistent data may produce incorrect results. The search - based optimization methods used various objective functions to find an optimal solution from a number of alternative solutions. These methods have high efficiency and high precision. Although, these methods are not scalable enough. The knowledge - based and reasoning mechanisms improve the analytics quality using the knowledge base. The major advantage of knowledge - based mechanisms is their relative simplicity of development. Although coverage for different scenarios is lower, whatever scenarios are covered by these mechanisms will provide high accuracy.

References

[1] Di Vaio, A.; Boccia, F.; Landriani, L.; Palladino, R. Artificial intelligence in the agri - food system: Rethinking sustainable business models in the COVID - 19 scenario. Sustainability **2020**, 12, 4851. [**Google Scholar**] [**CrossRef**]

[2] López - Robles, J. R.; Otegi - Olaso, J. R.; Gómez, I. P.; Cobo, M. J.30 years of intelligence models in management and business: A bibliometric review. Int. J. Inf. Manag.2**019**, 48, 22–38. [**Google Scholar**] [**CrossRef**]

[3] Sjödin, D.; Parida, V.; Palmié, M.; Wincent, J. How AI capabilities enable business model innovation: Scaling AI through co - evolutionary processes and feedback loops. J. Bus. Res.2**021**, 134, 574–587. [**Google Scholar**] [**CrossRef**]

[4] Chen, H.; Chiang, R. H.; Storey, V. C. Business intelligence and analytics: From big data to big impact. MIS Q.2**012**, 1, 1165–1188. [**Google Scholar**] [**CrossRef**]

[5] Di Vaio, A.; Palladino, R.; Hassan, R.; Escobar, O. Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. J. Bus. Res.2**020**, 121, 283–314. [**Google Scholar**] [**CrossRef**]

[6] Davenport, T. H.; Ronanki, R. Artificial intelligence for the real world. Harv. Bus. Rev.2**018**, 96, 108–116. [**Google Scholar**]

[7] Davenport, T.; Guha, A.; Grewal, D.; Bressgott, T. How artificial intelligence will change the future of marketing. J. Acad. Mark. Sci.2**020**, 48, 24–42. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[8] Colson, E. What AI - Driven Decision - Making Looks Like. Harvard Business Review.2019. Available online: **https: //hbr. org/2019/07/what - ai - driven - decision - making - looks - like** (accessed on 15 January 2023).

[9] Lichtenthaler, U. Extremes of acceptance: Employee attitudes toward artificial intelligence. J. Bus. Strategy **2020**, 41, 39–45. [**Google Scholar**] [**CrossRef**]

[10] Wiener, M.; Saunders, C.; Marabelli, M. Big - data business models: A critical literature review and multiperspective research framework. J. Inf. Technol.2**020**, 35, 66–91. [**Google Scholar**] [**CrossRef**]

[11] Ardito, L.; Scuotto, V.; Del Giudice, M.; Petruzzelli, A. M. A bibliometric analysis of research on Big Data analytics for business and management. Manag. Decis.2**019**, 57, 1993–2009. [**Google Scholar**] [**CrossRef**]

[12] Batistič, S.; van der Laken, P. History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. Br. J. Manag.2**019**, 30, 229–251. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[13] Khanra, S.; Dhir, A.; Mäntymäki, M. Big data analytics and enterprises: A bibliometric synthesis of the literature. Enterp. Inf. Syst.2**020**, 14, 737–768. [**Google Scholar**] [**CrossRef**]

[14] Wamba, S. F.; Bawack, R. E.; Guthrie, C.; Queiroz, M. M.; Carillo, K. D. A. Are we preparing for a good AI society? A bibliometric review and research agenda. Technol. Forecast. Soc. Chang.2**021**, 164, 120482. [**Google Scholar**] [**CrossRef**]

[15] Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W. M. How to conduct a bibliometric analysis: An overview and guidelines. J. Bus. Res.2**021**, 133, 285–296. [**Google Scholar**] [**CrossRef**]

[16] Hu, F.; Liu, W.; Tsai, S. - B.; Gao, J.; Bin, N.; Chen, Q. An empirical study on visualizing the intellectual

structure and hotspots of big data research from a sustainable perspective. Sustainability **2018**, 10, 667. [**Google Scholar**] [**CrossRef**] [**Green Version**]
[17] Abdollahi, A.; Rejeb, K.; Rejeb, A.; Mostafa, M. M.; Zailani, S. Wireless Sensor Networks in Agriculture: Insights from Bibliometric Analysis. Sustainability **2021**, 13, 12011. [**Google Scholar**] [**CrossRef**]
[18] Liao, H.; Tang, M.; Luo, L.; Li, C.; Chiclana, F.; Zeng, X. - J. A bibliometric analysis and visualization of medical big data research. Sustainability **2018**, 10, 166. [**Google Scholar**] [**CrossRef**] [**Green Version**]
[19] Morato, J.; Sanchez - Cuadrado, S.; Iglesias, A.; Campillo, A.; Fernández - Panadero, C. Sustainable technologies for older adults. Sustainability **2021**, 13, 8465. [**Google Scholar**] [**CrossRef**]
[20] Cano, J. A.; Londoño - Pineda, A.; Castro, M. F.; Paz, H. B.; Rodas, C.; Arias, T. A Bibliometric Analysis and Systematic Review on E - Marketplaces, Open Innovation, and Sustainability. Sustainability **2022**, 14, 5456
[21] Swain, D.; Mehta, U.; Bhatt, A.; Patel, H.; Patel, K.; Mehta, D.; Acharya, B.; Gerogiannis, V. C.; Kanavos, A.; Manika, S. A Robust Chronic Kidney Disease Classifier Using Machine Learning. Electronics **2023**, 12, 212. [**Google Scholar**] [**CrossRef**]
[22] Nusir, M.; Louati, A.; Louati, H.; Tariq, U.; Abu Zitar, R.; Abualigah, L.; Gandomi, A. H. Design Research Insights on Text Mining Analysis: Establishing the Most Used and Trends in Keywords of Design Research Journals. Electronics **2022**, 11, 3930. [**Google Scholar**] [**CrossRef**]
[23] AlZu'bi, S.; Abu Zitar, R.; Hawashin, B.; Abu Shanab, S.; Zraiqat, A.; Mughaid, A.; Almotairi, K. H.; Abualigah, L. A Novel Deep Learning Technique for Detecting Emotional Impact in Online Education. Electronics **2022**, 11, 2964. [**Google Scholar**] [**CrossRef**]
[24] Ai, M. A. S.; Shanmugam, A.; Muthusamy, S.; Viswanathan, C.; Panchal, H.; Krishnamoorthy, M.; Elminaam, D. S. A.; Orban, R. Real - Time Facemask Detection for Preventing COVID - 19 Spread Using Transfer Learning Based Deep Neural Network. Electronics **2022**, 11, 2250. [**Google Scholar**] [**CrossRef**]
[25] Daradkeh, M.; Abualigah, L.; Atalla, S.; Mansoor, W. Scientometric Analysis and Classification of Research Using Convolutional Neural Networks: A Case Study in Data Science and Analytics. Electronics **2022**, 11, 2066. [**Google Scholar**] [**CrossRef**]
[26] Fathimathul Rajeena, P. P.; Orban, R.; Vadivel, K. S.; Subramanian, M.; Muthusamy, S.; Elminaam, D. S. A.; Nabil, A.; Abualigah, L.; Ahmadi, M.; Ali, M. A. A novel method for the classification of butterfly species using pre - trained CNN models. Electronics **2022**, 11, 2016. [**Google Scholar**] [**CrossRef**]
[27] Ali, M. A. S.; Balasubramanian, K.; Krishnamoorthy, G. D.; Muthusamy, S.; Pandiyan, S.; Panchal, H.; Mann, S.; Thangaraj, K.; El - Attar, N. E.; Abualigah, L.; et al. Classification of Glaucoma Based on Elephant - Herding Optimization Algorithm and Deep Belief Network. Electronics **2022**, 11, 1763. [**Google Scholar**] [**CrossRef**]
[28] Alfaiz, N. S.; Fati, S. M. Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics **2022**, 11, 662. [**Google Scholar**] [**CrossRef**]
[29] Daradkeh, M. K. A Hybrid Data Analytics Framework with Sentiment Convergence and Multi - Feature Fusion for Stock Trend Prediction. Electronics **2022**, 11, 250. [**Google Scholar**] [**CrossRef**]
[30] Nadimi - Shahraki, M. H.; Mohammadi, S.; Zamani, H.; Gandomi, M.; Gandomi, A. H. A Hybrid Imputation Method for Multi - Pattern Missing Data: A Case Study on Type II Diabetes Diagnosis. Electronics **2021**, 10, 3167. [**Google Scholar**] [**CrossRef**]
[31] Nadimi - Shahraki, M. H.; Taghian, S.; Mirjalili, S.; Abualigah, L.; Elaziz, M. A.; Oliva, D. EWOA - OPF: Effective Whale Optimization Algorithm to Solve Optimal Power Flow Problem. Electronics **2021**, 10, 2975. [**Google Scholar**] [**CrossRef**]
[32] Hussien, A. G.; Abualigah, L.; Abu Zitar, R.; Hashim, F. A.; Amin, M.; Saber, A.; Almotairi, K. H.; Gandomi, A. H. Recent Advances in Harris Hawks Optimization: A Comparative Study and Applications. Electronics **2022**, 11, 1919. [**Google Scholar**] [**CrossRef**]
[33] Mir, I.; Gul, F.; Mir, S.; Khan, M. A.; Saeed, N.; Abualigah, L.; Abuhaija, B.; Gandomi, A. H. A Survey of Trajectory Planning Techniques for Autonomous Systems. Electronics **2022**, 11, 2801. [**Google Scholar**] [**CrossRef**]
[34] Russell & Norvig (2020) Russell S, Norvig P. Artificial intelligence a modern approach.4th edition Hoboken: Prentice Hall; 2020.
[35] Kotsiantis, Zaharakis & Pintelas (2007) Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering.2007; 160 (1): 3–24
[36] Carcillo et al. (2018) Carcillo F, Pozzolo AD, Borgne Y - A, Caelen O, Mazzer Y, Bontempi G. Scarff: a scalable framework for streaming credit card fraud detection with spark. Information Fusion.2018; 41: 182–194. doi: 10.1016/j. inffus.2017.09.005
[37] Kannan et al. (2019) Kannan N, Sivasubramanian S, Kaliappan M, Vimal S, Suresh A. Predictive big data analytic on demonetization data using support vector machine. Cluster Computing.2019; 22 (6): 14709–14720. doi: 10.1007/s10586 - 018 - 2384 - 8
[38] Feng et al. (2019) Feng M, Zheng J, Ren J, Hussain A, Li X, Xi Y, Liu Q. Big data analytics and mining for effective visualization and trends forecasting of crime data. IEEE Access.2019; 7: 106111–106123. doi: 10.1109/ACCESS.2019.2930410.
[39] Hammou, Lahcen & Mouline (2020) Hammou BA, Lahcen AA, Mouline S. Towards a real - time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics. Information Processing & Management.2020; 57 (1): 102122. doi: 10.1016/j. ipm.2019.102122. [CrossRef] [Google Scholar]
[40] Tian et al. (2020) Tian Z, Luo C, Lu H, Su S, Sun Y, Zhang M. User and entity behavior analysis under urban big data. ACM Transactions on Data Science.2020; 1 (3): 1–19.
[41] Kaur, Sharma & Mittal (2018) Kaur P, Sharma M, Mittal M. Big data and machine learning based secure

healthcare framework. Procedia Computer Science.2018; 132: 1049–1059. doi: 10.1016/j. procs.2018.05.020.

[42] Nair, Shetty & Shetty (2018) Nair LR, Shetty SD, Shetty SD. Applying spark based machine learning model on streaming big data for health status prediction. Computers & Electrical Engineering.2018; 65: 393–399. doi: 10.1016/j. compeleceng.2017.03.009.

[43] AlZubi (2020) AlZubi AA. Big data analytic diabetics using map reduce and classification techniques. The Journal of Supercomputing.2020; 76 (6): 4328–4337. doi: 10.1007/s11227 - 018 - 2362 - 1

[44] El - bana, Al - Kabbany & Sharkas (2020) El - bana S, Al - Kabbany A, Sharkas M. A multi - task pipeline with specialized streams for classification and segmentation of infection manifestations in COVID - 19 scans. PeerJ Computer Science.2020; 6: e303. doi: 10.7717/peerj - cs.303