Modern Business Data Analysis and Data Visualization: A Real-Time Fusion Study

Suji Priya J[1*], Vijayadharsan S[1], Vasumathi A[1] , and Rethika S[1]

[1]Master of Computer Application, Sona College of Technology, Salem, India

Abstract—In contemporary data science and analytics, data clustering is a small bucket that divides computation among various child nodes. The network's capacity, specialized tools, and applications that cannot be trained quickly are among these methods' drawbacks. In addition, the IoT-formed Big Data raw data can result in highly heterogeneous and unstructured data. This kind of data is difficult to analyze for real-time analytics. Real-time analytical challenges can be reduced by making computational values available locally rather than via distributed resources. Most of the time, it takes a long time and a lot of money to run these teams and skill sets. As an alternative, provide tools that let end users, professionals in the industry, and data scientists directly create and deploy complex data analytics application solutions with less technical knowledge. It highlights key advantages, disadvantages, and potential future directions by contrasting various current research and practice approaches to assisting end users with data analytics.

Keywords—Data Analytics, Data Science, Data Clustering, Data Visualization, and Real-time Analytics

1. Introduction

Data analytics plays a pivotal role in big data as it enables businesses to extract valuable insight to optimize As data continues to be produced and shared at magnanimous levels, businesses are clamouring to leverage the advantages it offers. By adopting sophisticated data management techniques, companies can gain a competitive edge in the market. Predictive analytics and manufacturing analytics are two powerful forms of data analytics that enable organizations to gain a deeper understanding of the underlying trends in their data. This allows them to make better decisions and drive better business outcomes.

In today's data-driven world, companies across various industries rely on advanced data analytics to better understand their businesses. Among these analytics, predictive analytics and manufacturing analytics stand out as two powerful tools for extracting meaning from extensive data. Predictive analytics uses statistical algorithms to identify patterns and make informed predictions about future events. Manufacturing analytics optimizes production processes to improve efficiency and reduce costs. By leveraging these types of advanced analytics, businesses can unlock valuable insights from their data to make better decisions. This will enable them to gain a competitive advantage in their respective markets.

Advanced data analytics has transformed our understanding and insights from complex data sets. Data mining uncovers hidden patterns and valuable insights. To achieve such miraculous results, Artificial Intelligence and Automation methods have been game changers. The new methods have elevated the possibilities of Machine Learning, Data modeling, neural networks, data visualization, predictive modeling, and beyond. In this world of advanced data analytics, the sky’s the limit. Businesses that capitalize on such cutting-edge methodologies will have a substantial competitive edge over their competitors. The use of various tools can combine data to yield valuable predictions and insights. These insights are crucial in identifying market demands and adapting to changing business conditions.

*Corresponding author: sujipriya@sonatech.ac.in

The business landscape has evolved rapidly in recent years as emerging technologies have transformed how we collect, interpret, and act on data. This shift has rendered traditional analytics and BI tools obsolete, unable to keep pace with data-driven demands. Nevertheless, advanced analytics companies have given businesses an edge over their competitors. This has enabled them to gain deeper insights, uncover hidden patterns, and take more informed decisions based on predictive modeling and real-time analytics. These innovative companies are revolutionizing the industry, sparking an era of growth and success that promises to shape the future of business intelligence.

a. Future prediction: The actions take place rapidly when it deals with greater confidence about the outcomes in the future aspects. It helps the organization to make decisions with a perspective view and get a deep search into the market trends, customer priority, and activities of the key business. Implementation of advanced analytics into the organizations may lead to faster growth of the respective organization so that predictions could be perfect.
b. Reduce risks: Risky decisions and cost factors could be avoided after analyzing and predicting by using advanced analytics. When a business finds out its cost avoidance method, then the business grows. The only motive of every organization is to develop to its extent and be competitive with the upcoming organization. To identify and manage the risk, it also provides information about the past, present, and future for better understanding.
c. Problem-solving analysis: Problem-solving analysis is not a simple one to solve a problem according to the problem statement, which needs to be thought with a presence of mind. For these cases, the traditional BI is not supportable or recommendable. Instead of that, advanced analytics could be used. This gives the solution as well as better accuracy to the problematic statements.

2.Related works

Mahyuddin K M Nasution, et.al. determines the base of science begins from the roots of mathematics. To build more on the science project the main thing is the self-interest of an individual and group discussion. Structured and systematic data is included in data science. Many relationships are related to the statistics of computer science. It aids to reveal the limitations and obstacles to computer science, being extended and shows that science is not directly dependent on other science [1]. Dan Puiu, et.al. examines the connections that are made between the humans of the real world entity and the digital world. The world is revolving to grow technology to its extent thinking. The CityPulse framework helps large-scale systems, data analytics, and other large platforms to service the best in its methods. Intelligent data aggregation, event detection, quality assessment, contextual filtering, and decision support are the things that differ from existing systems. It aids in the easy way of communication made to the digital systematic world [2]. Sohail Jabbar1, et.al. discovers that traditional big data analytics uses a data clustering technique that is not efficient enough to describe the process and to give better results for the system. These systems bring problems in network capacity and specialized tools. It aids to break down all these issues which is already in progress and which may further occur. So it uses tree methods such as relational, semantical, and big data-based data and metadata with their enhanced capabilities [3]. Ravi Vatrapu, et.al. proposed that text analysis, social network analysis, social complexity analysis, and social simulations of these four things are categorized into the computational social science approach. It aids in the development of all these analyses into the newly updated methodology called social set analysis. Philosophies of computational social science, theory of social data, and conceptual and formal models of social data are the framework for the social set analysis [4]. L. Erhan, et.al. determines the deployment of the Internet of Things and digital technologies into society and that is gathering a large amount of data, saved to the database for the future service of the respective systems. The data collected are from the sensors and the user provides for the future aspects. By implementing the techniques like machine learning and data science, the accuracy of the data is calculated more relatable to the situation and it could maintain the information on a large scale [5].

Mingchen Feng, et.al. discovers a large amount of data for the different patterns after analyzing and identifying the approach of Big Data Analytics. It aids in the implementation of Big data analysis in criminal data, where visualization is made. This technique could be maintained by the techniques like data mining and deep learning. It results in better than the neural network model [6]. Tamer Z. Emara, et.al. describes the storing of data in centers are becoming huge. So the companies use multiple data centers for storing and retrieving the data from the database. It aids the newest system, the Random Sample Partition data model which is developed in recent

times for the betterment of worst-case scenarios. It converts the big data center into multiple data blocks. The result shows the efficiency of the process [7]. Jiangcheng Zhu, et.al. examines the newest technology that will combine Neural Networks and Data Assimilation. It is based on the structural model that proposes the framework for the Neural Network (NN) with the support of Data Assimilation (DA). The DA model is worked with the support of the Kalman filter [8].

Kenneth Li-Minn Ang et al. proposed that mobile technologies are growing day by day, while mobile increases the data that also increased. To manage a huge amount of data, there is a method, which is Big Data Analytics. It aids in the implementation of big data in higher education and learning techniques. It uses various techniques such as learning management systems (LMS), massive open online courses (MOOC), learning object repositories (LOR), OpenCourseWare (OCW), and open educational resources (OER) [9]. Norita Ahmad et al. has proposed that the data is stored in a huge amount and is mainly focused on data science. Data science is the one that manages every single bit of data that needs to be allocated. It not only accepts the companies with good understanding but it aims to solve the problem and make the right decision at the right time [10]. Mujthaba G.M et al. describe when a user sends a request or response to the respective data, it is not easier to find among a large amount of data, which makes them more complicated to make use of the data. It aids in the techniques such as Artificial Intelligence, Machine Learning, Deep learning, etc., and the operations such as data cleaning, data processing, data modeling, data visualization, and data presentation techniques. It gives accuracy in finding the data [11]. Peerapon Kamlangpuech et al. determines the popularity of Computer Science (CS), Information Technology (IT), and Data Science (DS), which is most commonly used by school children, students, and people. So now many courses could be gained from both online/offline methods which are comfortable to them. It aids in the process of analyzing the CS course content, called the CSCDA system (Computer Science Course Description Analysis system). It checks the content with matching text that shows the similarity and dissimilarities that belongs to the two CS courses [12].

Heonho Kim et al. have proposed that data mining models are used to analyze the patterns of the time-series database. The patterns analyzed are later revealed for risk prediction, system management, and decision-making. It aids the new concept of flexible periodic patterns, which gives better efficiency to the upcoming progress. Efficient periodic pattern mining (EPPM) and flexible periodic pattern mining (FPPM) are the techniques used in the proposed scheme [13]. HouriehKhalajzadeh et al. has examined the challenges faced in data analytics. Semantics and other uses are used for mining and communicating with data visualization techniques. Mostly the data mining techniques are used for the research purpose and assist in integrating processes. It aids the survey that they have examined throughout the database [14]. Danda B. Rawat et al. have proposed the information gathered is the way to develop the knowledge of the system, and this has become an issue because of the large amount of data to the relative performance of the data mining. For future examinations, the data are stored and gathered. It leads to the protection of data given by cyber-security for the huge amount of data [15]. Jeremy Greenwald, Art Frank, et al. determine there has been a lot of debate about whether or not using static code attributes to learn defect predictors is beneficial. Defect predictors like "McCabes versus Halstead versus lines of code counts" have been the subject of previous research. Since how the attributes are used to build predictors is much more important than which attributes are used, we demonstrate here that such debates are irrelevant. Additionally, contrary to previous pessimism. If prior research had focused on attribute subsets rather than learning methods as a whole, it would not have found reliable predictors; Our overall conclusion is that evaluating defect learning methods with just one data set and one learner is no longer sufficient [16].

A.L.Sayeth Saabith et al. have proposed well-studied methods of mining frequent itemsets, and association rule mining (ARM) reveals attractive correlations between variables in large datasets. One of ARM's most widely used algorithms, the Apriori algorithm, finds association rules in massive datasets by collecting itemsets that occur frequently. The initial Apriori algorithm was designed for sequential (computer or single node) environments. Tools for data mining can anticipate future trends and actions, enabling decision-makers in a variety of fields to take proactive, knowledge-driven actions. As of late, with the quick development of data innovation, how much information has dramatically expanded in different fields. The majority of big data comes from Internet-based businesses and everyday activities [17]. Ahmedamine Fariz et al. have proposed the idea of "cooperative agents" and, by extension, multi-agent systems at the heart of distributed artificial intelligence, which stipulates that entities with a certain amount of autonomy must be able to perceive and respond to their surroundings. These frameworks are turning out to be increasingly more fundamental in numerous application fields because of the way that they can take care of the issues of intricacy and conveyance, particularly concerning huge frameworks, for example, information mining. The extraction of knowledge from multiple databases, regardless of their physical location, is known as distributed data mining (DDM); It enables partial

analyses of data extracted from distinct distributed sites and sends the various partial results to other sites to produce the final result [18]. A.Pradeepa et al. discovered a new large-scale classifier that was developed using mining later. The Map Reduce simulator was made so that the proposed a priori algorithms' scalability on MaReduce could be evaluated. Associative rule mining inherits MapReduce's scalability to thousands of processing nodes and massive datasets. It employs a counting-based hybrid approach among miners to locate frequent item sets. Classifiers based on integrating classification and association rule mining can be more accurate and efficient than those based on conventional methods. MapReduce-based association rule mining, which is used to extract strict rules from large datasets, was recently introduced by [19].

Dr (Mrs). Sujni Paul et al. has proposed one of the best Association Rule mining algorithms is Rakesh Agarwal's Apriori algorithm. Additionally, the majority of parallel algorithms are based on Ii. The problem of association rule discovery is an ideal one to solve on multiple processors simultaneously due to the enormous and high-dimensional datasets that are typically available as input. The primary reason is that single processors are constrained by memory and CPU speed. Parallel data mining, distributed data mining, incremental data mining, and an optimized distributed association rule mining algorithm are discussed in this paper. The goal of data mining is to efficiently handle a large set of evolving and distributed data. The issues and current research on parallel and distributed data mining are the subjects of our discussion in this paper. The adaptability of some fundamental data mining algorithms, such as clustering, decision trees, and the discovery of frequent patterns. We have identified two methods for performing distributed data mining and attempted to highlight the bandwidth and network latency advantages of employing mobile agents in client-server-based methods [20]. Shashikumar G. et al. describe regular algorithms that are used to extract information from large data sets as part of the data mining process. Big Data refers to this extensive collection of market or business-related data. The exchange of data that tends to increase the enterprise's value is the central focus of business intelligence (BI). It is preferable to comprehend how organizations view big data and the extent to which they are currently utilizing it to benefit their business rather than collecting information on what organizations are doing. Organizations are now beginning to investigate methods for processing and analyzing this massive amount of data. In addition to BI, meteorology, petroleum exploration, and bioinformatics are among the scientific fields where big data and data mining are gaining popularity. Software, hardware, and sophisticated algorithms are required to support this data sequence [21].

M.Jayasree et al. have proposed the difficulty in locating rules of association between products in a large database of sales transactions. To solve this problem, we present two brand-new algorithms that are fundamentally distinct from the existing ones. These algorithms outperform known algorithms by a factor ranging from three for small problems to more than an order of magnitude for large ones, according to empirical testing. Additionally, we demonstrate how a hybrid algorithm known as AprioriHybrid can be created by combining the best aspects of the two proposed algorithms. AprioriHybrid scales linearly, as demonstrated by scale-up tests [22]. Rakesh Agrawal Ramakrishnan Srikant et al. describe one of the most well-known methods for gaining knowledge from data as Frequent Item Set Mining (FIM). When applied to Big Data, the combinatorial explosion of FIM methods becomes even more problematic. Fortunately, effective solutions to this issue are already available thanks to recent advancements in parallel programming. However, these tools come with their own set of technical issues, such as costs for inter-communication and balanced data distribution. In this paper, we look into whether FIM techniques can be used with the Map Reduce platform [23]. Sandy Moens et al. examine there has been a lot of debate about whether or not using static code attributes to learn defect predictors is beneficial. Defect predictors like "McCabes versus Halstead versus lines of code counts" have been the subject of previous research. Since how the attributes are used to build predictors is much more important than which attributes are used, we demonstrate here that such debates are irrelevant, likewise, in opposition to earlier negativity [24]. Amit K et al. describe the well-studied methods of mining frequent itemsets and association rule mining (ARM), revealing attractive correlations between variables in large datasets. One of ARM's most widely used algorithms, the Apriori algorithm, finds association rules in massive datasets by collecting itemsets that occur frequently. The initial Apriori algorithm was designed for sequential (computer or single node) environments. Tools for data mining can anticipate future trends and actions, enabling decision-makers in a variety of fields to take proactive, knowledge-driven actions. The amount of data in various fields has recently increased exponentially as a result of the rapid growth of information technology. The majority of big data comes from Internet-based businesses and everyday activities [25].

Mahesh A. et al. describe a new large-scale classifier developed using mining later. The Map Reduce simulator was made to evaluate the proposed a priori algorithms' scalability on MapReduce. Associative rule mining inherits MapReduce's scalability to thousands of processing nodes and massive datasets. It employs a

counting-based hybrid approach among miners to locate frequent item sets. A coordinating arrangement and affiliation rule mining can create more effective and exact classifiers than customary strategies. MapReduce-based association rule mining, which is used to extract strict rules from large datasets, was recently introduced by [26]. Abinaya K et al. describe regular algorithms used to extract information from large data sets as part of data mining. Big Data refers to this extensive collection of market or business-related data. The exchange of data that tends to increase the enterprise's value is the central focus of business intelligence (BI). It is preferable to comprehend how organizations view big data and the extent they currently utilize it to benefit their business rather than collecting information on what organizations are doing. Organizations are now investigating methods for processing and analyzing this massive data. In addition to BI, meteorology, petroleum exploration, and bioinformatics are among the scientific fields where big data and data mining are gaining popularity. Software, hardware, and sophisticated algorithms are required to support this data sequence [27]. AnbumalarSmilin et al. determine the model and predict the future with new challenges thanks to datasets like these, which offer us unparalleled opportunities. To forecast the future, it is, therefore, necessary to be aware of these flaws and the possibilities presented by these massive data sets. These days, huge data sets, such as big data, are growing due to technological advancements. Give a general overview of this subject, such as its current condition, controversy, and difficulties predicting the future. This paper characterizes some of these issues, utilizing delineations with applications from different regions. A brand-new term: is used to identify the datasets that were collected. However, because of their size and complexity, we cannot extract those datasets using our current methods or data mining software tools [28]. R. Hemamalini et al. examine the Big Data e-Health Service application that promises to improve efficiency, cost-effectiveness, and quality throughout heart disease healthcare. This application includes an information-driven model and an interest-driven total of data sources. As e-Health heart disease emerges as one of the primary drivers of innovation, Big Data is transforming healthcare and business. For Big Data applications in the e-Health service domain, investigate BDeHS (Big Data e-Health Service). Introducing the new knowledge of Big Data for understanding complex, growing, large-volume data sets with multiple independent sources. The HACE theorem describes the characteristics of the big data revolution and how to carry out the operation from a data mining perspective [29].

Table 1: Comparative Analysis

S.no	Techniques	Merits	Demerits
1	IoT	a. Reduced costs b. High efficiency and productivity c. Increased mobility and agility	a. Security issues b. Complexity c. Sometimes corrupted
2	Fuzzy	a. Robust b. Precise inputs	a. Dependent on human knowledge and expertise
3	RDB	a. Easily queried b. Consistent DB Design	a. High volume transaction b. Require storage
4	RDF	a. Consistent framework b. Metadata c. Standard syntax	a. Required memory
5	Linear Regression	a. Simple and easy to implement b. Easy to interpret the output coefficient	a. Low efficiency
6	BDA	a. Improved decision making b. Reduced costs c. Enhanced customer service	a. Security hazard b. Adherence

3. Major important factor

Show the information to the outer world is a complex task in big data analytics. It explores the hidden pattern and correlates the data, trends, and marketing. Customer priority is the information that could support organizations to make decisions well.

Advanced data analytics techniques and technologies could answer business operations and performance questions. And this could also do the process, such as gathering new information and analyzing the data sets. In organizations, data-driven decisions are taken by big data analytics software systems because this could develop the outcomes of business-related projects. It could compete with the other growing technologies when the efficiency is good. If the current approach feels like drowning means, it could adapt to the upcoming technologies.

A. Composition of data analytics

Arranging the data is about the generate, capturing, and storing in various formats, but when it comes to the analysis, it's different; all created data formats will not be equal. The data analytics structure could be designed or formatted with the basic building blocks of the algorithm. Such as,

- Rows
- Columns or fields
- Binning and Histograms
- Distributions and outliers
- Data types
- Pivot and Unpivot data
- Wide, tall data
- Normalization

B. Usage of data analytic frameworks

A framework describes the outline of the product, i.e., the outer part is practiced for data analytics management. The fundamental goal of the data analytics framework is to support enterprises in describing the greatest value from the information. The framework for the solutions may vary between the organizations, according to the problem statement.

Uses:

- The performance of the data process is to be measured for future aspects.
- The developed product must be based on the problems people face in the living world.
- Maintenance of the system is to be monitored so that the prediction will result positively.

C. Data analytics tools

Data analysis is the collection and analysis of the data of a business. This process could be improved by using the tools such as software and programs.

- MonkeyLearn
- RapidMiner
- KNIME
- Talend
- Airtable
- ClicData
- Qlik

D. The architecture of the DA

Fig 1**:** Architecture for current data analytics

To enhance data processing in data analytics, multiple technical layers are used and are commonly called umbrella terms. These layers are meticulously designed to expedite data analysis and storage. Fig. 1 explores these layers in detail, shedding light on their system protocols and tools. These layers also offer a stream of functions that can efficiently organize, collect, store, and use applications and tools at any given time. This allows users to analyze data easily and conveniently.

Business is increasingly reliant on data. There are, however, risks associated with that reliance. That's because business data analytics is a major issue when it comes to security breaches, while data management has emerged as a recent problem. Enterprises must focus on dispelling myths and implementing data analytics to understand fusion datasets. During the process, the server may fail and cybercriminals could exploit vulnerabilities to compromise business intelligence. For these reasons, it's vital to improve data visualization and analysis processes through precise definitions and enhanced security measures to safeguard valuable information.

The architecture for data discovery is x

- designed to handle the ingestion
- processing
- analysis, and a large, complex database system,

All of which can be effectively carried out with batch processing.

E. Levels of data analytics

There are three levels in data analytics, which are listed below.

i. Physical level
ii. Conceptual level
iii. External level

Physical level: The physical level, often referred to as the internal level, constitutes the foundational tier within the three-tier database architecture. At this fundamental stratum, the database's data storage mechanism is elucidated. Here, data is stored as binary bits, residing on external hard drives at its most rudimentary form. On a more advanced front, it can be likened to the directory or folder housing the data file. Additionally, the physical level encompasses discussions regarding strategies pertaining to encryption and compression techniques, which play a pivotal role in managing and safeguarding the data.

Conceptual level: The logical level, also recognized as the conceptual level, delves into the database's abstract structure. It provides users with a high-level view of how the database appears conceptually and elucidates the relationships among different data tables. At this level, the specific storage details of the database become inconsequential, as the focus is purely on the overarching design and how users perceive and interact with the data.

External level: At the zenith of the three-tier structure lies the external layer, often referred to as the view level. This topmost tier brings users closest to the database, offering them tailored perspectives. Here, the intricacies of the underlying data remain concealed, and only curated views of the pertinent database information are presented to users. As a result, users can interact with the database according to their specific needs, granting them the flexibility to access and perceive the data in diverse ways.

Within the realm of data management, analysis stands as a pivotal force, serving to purge superfluous data and uphold the sanctity of data integrity. Through the art of data visualization, enterprises can swiftly discern aberrations and zero in on security breaches. Incorporating blockchain methodologies further fortifies businesses, enhancing traceability, fortifying the cocoon of sensitive data, and fortifying data privacy. Conversely, hash graph techniques offer a potent shield, effectively thwarting potential data breaches and vulnerabilities. In the contemporary digital landscape, where the orchestration of sturdy data management solutions holds the key to prosperity, the adept utilization of these formidable tools is paramount.

This architecture has ingeniously crafted a pairing mechanism that acts as the key to unlock the full potential data functions, thereby finely tuning time management to perfection. Moreover, it employs collaborative filtering, seamlessly orchestrating unsupervised data flows while safeguarding the confidentiality of authors. By harnessing these cutting-edge techniques, the system doesn't just optimize operational efficiency; it also places user privacy and security on a paramount pedestal. This sophisticated amalgamation of methods epitomizes a holistic approach to handling invaluable data, ultimately elevating productivity without compromising the fortress of high-level security. Refer to Table 2 for a detailed algorithm comparison. The figure 2 shows the prediction rates of the algorithms.

Table 2: Comparison with other Algorithms / Methods

Dataset	**Algorithm**	**Min. Support**				
		100	150	200	250	300
T10I4D100K	Dist-Eclat	45	99	100	200	250
	BigFIM	56	110	110	220	220
	ClustBigFIM	56	110	120	210	240
	Collaborative filtering Screening Algorithm	99	152	185	245	280

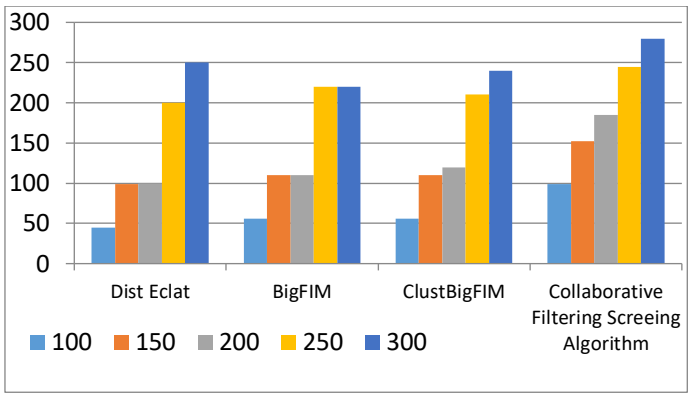

Fig 2: Prediction Rates

4.Conclusion

The realm of data science has been ignited by a multitude of esteemed experts. In the ever-evolving landscape of scientific endeavors, one can observe a tapestry of responses from both the general public and fellow scientists, all woven together through the diverse terminologies used to propose ideas and extend invitations. Interestingly, while data has played a role in scientific pursuits for eons, the formal discipline of data science is relatively new, emerging as a distinct field of inquiry.

To fully appreciate the current state of the data science, it's imperative to delve into its historical trajectory. Within this domain, data assumes two primary guises: the objective and the subjective, affording multifaceted avenues for analysis and comparison. It's worth noting that the statistical significance of data hinges on their completeness, accuracy, and availability. When we integrate this fusion of data, concepts like classification and regression in statistical analysis become more readily digestible. These facets underscore the critical importance of scrutinizing the current landscape of data science as it continues to unfurl its profound impact on our world.

References

1. Nasution, Mahyuddin KM, Opim Salim Sitompul, and Erna Budhiarti Nababan. Data science. Journal of Physics: Conference Series. Vol. **1566**. No. 1. IOP Publishing, (2020).
2. Anagnostopoulos, Eleftherios. Bus Scheduling including Dynamic Events. (2017).
3. Jabbar, Sohail, et al. A methodology of real-time data fusion for localized big data analytics. IEEE Access 6, PP. 24510-24520 (2018).
4. Vatrapu, Ravi, et al. Social set analysis: A set theoretical approach to big data analytics. Ieee Access 4, PP. 2542-2571 (2016).
5. Erhan, Laura, et al. Analyzing objective and subjective data in social sciences: Implications for smart cities. *IEEE Access* 7, PP. 19890-19906 (2019).
6. Feng, Mingchen, et al. Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access* 7, PP. 106111-106123 (2019).
7. Emara, Tamer Z., and Joshua Zhexue Huang. "Distributed data strategies to support large-scale data analysis across geo-distributed data centers." *IEEE Access* 8, PP. 178526-178538 (2020).
8. Zhu, Jiangcheng, et al. Model error correction in data assimilation by integrating neural networks. *Big Data Mining and Analytics* 2.2, PP. 83-91 (2019).
9. Ang, Kenneth Li-Minn, Feng Lu Ge, and Kah Phooi Seng. Big educational data & analytics: Survey, architecture and challenges. *IEEE access* 8, PP. 116392-116414 (2020).

10. Ahmad, Norita, Areeba Hamid, and Vian Ahmed. Data science: Hype and reality. *Computer* 55.2, PP. 95-101 (2022).
11. Mujthaba, G. M., et al. Data Science Techniques, Tools and Predictions. *International Journal of Recent Technology and Engineering (IJRTE)* 8.6 (2020).
12. Kamlangpuech, Peerapon, and Komate Amphawan. A new system for analyzing contents of computer science courses. *2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)*. IEEE, (2020).
13. Kim, Heonho, et al. Periodicity-oriented data analytics on time-series data for intelligence system. *IEEE Systems Journal* 15.4, PP. 4958-4969 (2020).
14. Khalajzadeh, Hourieh, et al. Survey and analysis of current end-user data analytics tool support. *IEEE Transactions on Big Data* 8.1, PP. 152-165 (2019).
15. Rawat, Danda B., Ronald Doku, and Moses Garuba. Cybersecurity in big data era: From securing big data to data-driven security. *IEEE Transactions on Services Computing* 14.6, PP. 2055-2072 (2019).
16. Menzies, Tim, Jeremy Greenwald, and Art Frank. Data mining static code attributes to learn defect predictors. *IEEE transactions on software engineering* 33.1, PP. 2-13 (2006).
17. Saabith, AL Sayeth, Elankovan Sundararajan, and Azuraliza Abu Bakar. Parallel implementation of apriori algorithms on the Hadoop-MapReduce platform-an evaluation of literature. *Journal of Theoretical and Applied Information Technology* 85.3, PP. 321 (2016).
18. Rghioui, A. N. A. S. S., et al. Symmetric cryptography keys management for 6lowpan networks. *Journal of Theoretical and Applied Information Technology* 73.3, PP. 336-345 (2015).
19. Pradeepa, A., and A. S. Thanamani. Parallelized Comprising for Apriori algorithm using Mapreduce framework. *International Journal of Advanced Research in Computer and Communication Engineering* 2.11, PP. 4365-4368 (2013).
20. Rokhman, Nur, and Amelia Nursanti. The MapReduce Model on Cascading Platform for Frequent Itemset Mining. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 12.2, PP. 149-160 (2018).
21. Totad, Shashikumar G., et al. Scaling data mining algorithms to large and distributed datasets. *Intl J Database Manag Syst* 2.2, PP. 26-35 (2010).
22. Jayasree, M. Data Mining: Exploring Big Data Using Hadoop and Map Reduce. *International Journal of Engineering Science Research-IJESR,* Vol. **04**. No. 1 (2013).
23. Agrawal, Rakesh, and Ramakrishnan Srikant. Fast algorithms for mining association rules. *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215 (1994).
24. Moens, Sandy, Emin Aksehirli, and Bart Goethals. Frequent itemset mining for big data. *2013 IEEE international conference on big data*. IEEE, (2013).
25. Tyagi, Amit K., R. Priya, and A. Rajeswari. Mining big data to predicting future. *International Journal of Engineering Research and Applications* 5.3, PP. 14-21 (2015).
26. Mahesh, A. Shinde#1, K. P. Adhiya. Frequent Itemset Mining Algorithms for Big Data using MapReduce Technique, ICGTETM (2016).
27. Abinaya, K. Data mining with big data e-health service using map reduce. *International Journal of Advanced Research in Computer and Communication Engineering* 4.2, PP. 123-127 (2015).
28. Anbumalar Smilin.V1, S.P.Siddique Ibrahim. A Survey on Different Techniques for Mining Frequent Itemsets, Vol. **5**. Issue 1 (2016).
29. Hemamalini, R., and L. Josephine Mary. An analysis on multi-agent based distributed data mining system. *International Journal of Scientific and Research Publications* 4.6, PP. 1-6 (2014).
30. Yoo, Kwan-Hee, Carson K. Leung, and Aziz Nasridinov. Big data analysis and visualization: challenges and solutions. *Applied Sciences* 12.16, PP. 8248 (2022).