International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.20
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on the Construction and Application of Information Management System Based on Big Data

Zhe Liang *

School of Management Science and Engineering, Anhui University of Finance and Economics, Anhui, China

*Corresponding Author: Zhe Liang

ABSTRACT

This paper focuses on the construction and application of information management systems based on big data in the digital era. It begins by introducing the importance of Information Systems (IS) and big data, and the challenges they face. The paper then provides an overview of big data, including its characteristics, data sources, and common technologies. It also discusses the development and evolution of IS, from the traditional theory-driven approach to the modern data-driven era. The application of big data in is explored, covering data collection, storage, analysis, and visualization. The paper further analyzes the existing data analysis methods, including theoretical analysis, statistical analysis, machine learning, and deep learning. Additionally, it addresses the privacy and security of data, considering legal, ethical, and quality aspects. The application of big data in various industries, such as healthcare, finance, retail, and manufacturing, is also discussed. Finally, the paper looks towards the future, highlighting emerging technologies like AI and quantum computing, and future research directions in big data analysis, security, and integration with AI.

KEYWORDS

Big Data Analytics; Information Management Systems; Data-Driven Decision Making

1. INTRODUCTION

In the digital era, Information Systems (IS) and big data have become important forces driving social development and transformation. IS is a broad field that covers various applications, from Enterprise Resource Planning systems and Customer Relationship Management systems to Database Management systems. It provides decision support and business process optimization for organizations and individuals by collecting, storing, processing, and disseminating information [1, 2].

Big data refers to a collection of data that is large in scale, complex in structure, diverse in type, and generated at a high speed. These data often exceed the processing capabilities of traditional databases and data processing technologies, requiring new techniques and methods for storage, management, and analysis. The development of data mining, machine learning, deep learning, and other technologies has provided strong support for the analysis and application of big data [1].

The importance and significance of IS and big data in modern life are self-evident. They play a key role in various fields, such as enterprise management, healthcare, financial services, scientific research, etc. By utilizing big data analysis, enterprises can better understand market demands, optimize production processes, and improve customer satisfaction; healthcare institutions can diagnose diseases more accurately and formulate personalized treatment plans; financial institutions can more effectively prevent risks and improve investment returns [3].

However, with the rapid development of big data, the IS field also faces a series of challenges and problems. For example, how to effectively manage and store big data to ensure the security and privacy of data; how to extract valuable information from massive data for data mining and analysis; how to integrate big data with IS systems to achieve data sharing and interaction; how to solve the pressure on computing and storage resources brought by big data. In addition, IS and big data themselves also have some inherent problems, such as data quality issues, data bias issues, and data overload issues, which may affect the accuracy and reliability of data analysis [4].

To solve these problems, predecessors have carried out a lot of research work. However, these research works also have some shortcomings, such as the limitations of research methods, the inconsistency of research results, and the lack of systematicness and comprehensiveness. Therefore, it is necessary to conduct in-depth research on IS and big data, explore new methods and technologies to improve the performance and effect of IS systems, and solve the problems and challenges brought by big data [5].

The contribution of this paper is to summarize the development status and existing problems of IS and big data through comprehensive analysis and research of relevant literature, discuss the application and challenges of big data in IS, and propose some suggestions and methods to solve the problems. At the same time, this paper also looks forward to the future research directions, providing a reference and reference for further in-depth research on IS and big data [6, 7].

2. OVERVIEW OF BIG DATA

Big data refers to a data collection that is large in scale, complex in structure, diverse in type, and generated at a high speed. It has several significant characteristics, commonly referred to as the "5V" features:

Volume (Large Quantity): The scale of big data is very large, far exceeding the data volume that traditional data processing technologies can handle. As Manyika et al. pointed out in their research, the scale of big data is constantly growing, posing higher challenges for data processing and storage [8].

Velocity (High Speed): The speed of data generation and update is very fast, requiring real-time processing and analysis. Lazer et al.'s research shows that in today's digital age, the generation speed of data is increasing exponentially, which requires us to have the ability to quickly process and analyze data [9].

Variety (Diversity): There are various types of data, including structured data (such as data in relational databases), unstructured data (such as text, images, audio, video, etc.), and semi-structured data (such as JSON, XML, etc.). Chen et al. discussed in detail the diversity of data types and the challenges of processing these data in their review of big data [10].

Value (Value): Big data contains rich information and value, but it needs to be mined through effective data processing and analysis techniques. George et al. emphasized the importance of extracting value from big data in their research and pointed out the key role of data science methods in achieving this goal [10]. They believe that data science methods include data mining, machine learning, deep learning, and other technologies, which can help us discover valuable information and patterns from massive data to support decision-making.

Veracity (Authenticity): The authenticity and accuracy of data are crucial for big data analysis. However, due to the diversity and complexity of data sources, the authenticity and accuracy of data may face certain challenges. This has been noted in many studies on big data, such as [11].

The data sources of big data are very wide, including the Internet, the Internet of Things, sensors, social media, enterprise systems, etc. These data sources generate different types of data, such as structured data, unstructured data, and semi-structured data. To process and analyze big data, a series of techniques and tools need to be adopted. Some common big data technologies include:

Hadoop: It is an open-source distributed computing platform, including the Hadoop Distributed File System (HDFS) and the MapReduce programming model. Hadoop can handle large-scale data sets and supports distributed storage and parallel computing. Many studies have introduced the application of Hadoop in big data processing, such as [12].

Spark: It is a fast and universal big data processing framework that supports memory computing and distributed computing. Spark can process structured data, unstructured data, and semi-structured data, and provides a rich set of data analysis and machine learning libraries. The performance and advantages of Spark are discussed in detail in [13].

NoSQL Database: It is a type of non-relational database, including key-value stores, document stores, column-family stores, etc. NoSQL databases can handle large-scale unstructured data and support distributed storage and high-concurrency access [14].

Distributed Processing (MapReduce): It is a distributed computing model that divides a large-scale data set into multiple small data sets and then processes these small data sets in parallel on multiple computing nodes. MapReduce can improve the efficiency and speed of big data processing.

In addition to the above technologies, there are many other big data technologies and tools, such as data warehouses, data mining, machine learning, deep learning, etc. These technologies and tools can help users better process and analyze big data and mine the value and information therein [15].

3. DEVELOPMENT AND EVOLUTION OF IS

Information Systems (IS) play a crucial role in organizations and society. Its role is to collect, process, store, and disseminate information to support decision-making, business process optimization, and organizational innovation. Over time, IS has undergone significant development and evolution.

3.1. Traditional IS (Theory-Driven)

In the early days, IS was mainly based on theory-driven methods. These theories include system theory, information theory, and management theory. Traditional IS aims to improve the efficiency and effectiveness of organizations by designing and implementing information systems. System analysts and designers use structured methods to develop information systems, which emphasize the integrity, hierarchy, and regularity of the system [16].

3.2. Modern IS (Statistical Methods, Basic ML Methods)

With the continuous development of information technology, IS has gradually adopted more statistical methods and basic Machine Learning (ML) methods. Statistical analysis is used to process and analyze data to extract valuable information. ML methods are used for prediction and classification to support decision-making. Modern IS also focuses on user experience and interface design to improve the usability and accessibility of the system [17].

3.3. IS in the Era of Big Data (Data-Driven)

In the era of big data, IS has become more data-driven. The emergence of big data has brought new opportunities and challenges to IS. Data-driven IS aims to support decision-making and business innovation by collecting, processing, and analyzing a large amount of data. Big data technologies, such as Hadoop, Spark, and NoSQL databases, are widely used in data storage and processing. Data mining, machine learning, and deep learning are used for data analysis and prediction [1, 7].

IS in the era of big data also focuses on data governance and data quality. Data governance refers to the management of the entire life cycle of data, including data collection, storage, processing, analysis, and use. Data quality refers to the accuracy, completeness, and consistency of data. Data governance and data quality are crucial for ensuring the reliability and availability of data.

In addition, IS in the era of big data also faces some challenges, such as data privacy and security, data integration, and interoperability. Data privacy and security refer to protecting data from unauthorized access, use, or disclosure. Data integration and interoperability refer to integrating data from different data sources into a unified view and ensuring that data between different systems can communicate and share with each other [18].

In summary, the development and evolution of IS is an ongoing process. From traditional theory-driven to modern statistical methods and basic ML methods, and then to data-driven in the era of big data, IS continuously adapts to new technologies and business needs, bringing great value to organizations and society.

4. APPLICATION OF BIG DATA IN IS

With the continuous development of big data technology, its application in Information Systems (IS) is becoming more and more extensive. The application of big data in IS mainly includes data collection, data storage, data analysis, and visualization.

4.1. Data Collection

Data Sources: The data sources of big data are very extensive, including sensors, social media, e-commerce websites, mobile devices, etc. These data sources generate different types of data, including structured data, semi-structured data, and unstructured data. As Wang and Zhang pointed out in their research, data collection in the era of big data faces diverse data sources and complex data types [19].

Data Acquisition Techniques: To collect these data, various data acquisition techniques need to be used, such as web crawlers, sensor networks, API interfaces, etc. These techniques can help enterprises obtain data from different data sources and integrate them into a unified data warehouse.

4.2. Data Storage

Hadoop Distributed Storage: Hadoop is an open-source distributed system architecture that can handle large-scale data sets. The core components of Hadoop include the Hadoop Distributed File System (HDFS) and the MapReduce programming model. HDFS is a distributed file system that can store data on multiple nodes to improve the reliability and availability of data [12]. MapReduce is a programming model that can divide a large-scale data set into multiple small data sets and process these small data sets in parallel on multiple nodes to improve the processing efficiency of data [15].

Cloud Storage: Cloud storage is a data storage method based on cloud computing technology, which can store data on cloud servers and users can access and use these data anytime and anywhere through the network. Cloud storage has the advantages of high reliability, high availability, and high scalability, and is suitable for processing large-scale data sets [20].

4.3. Data Analysis

Machine Learning (ML): Machine learning is an artificial intelligence technology that can automatically discover patterns and rules in data through learning and analysis, and make predictions and decisions based on these patterns and rules. Machine learning has been widely used in big data analysis, such as classification, clustering, regression analysis, etc. [21].

Deep Learning (DL): Deep learning is a machine learning technology based on neural networks, which can automatically learn the features and patterns in data through training with a large amount of data, and make predictions and decisions based on these features and patterns. Deep learning has achieved remarkable results in fields such as image recognition, speech recognition, and natural language processing.

4.4. Visualization

Importance of Data Visualization: Data visualization is to display data in the form of charts, graphs, etc., so that users can better understand and analyze the data. Data visualization can help users discover patterns and rules in the data, and improve the readability and understandability of the data.

Visualization Tools and Techniques: To achieve data visualization, various visualization tools and techniques need to be used, such as Excel, Tableau, PowerBI, etc. These tools and techniques can help users display the data in various forms, such as bar charts, line charts, pie charts, maps, etc.

5. DATA ANALYSIS

In the era of big data, the development of data analysis techniques provides strong support for Information Systems (IS). Existing data analysis methods include theoretical analysis, statistical analysis, machine learning, and deep learning, which play important roles in different application scenarios.

5.1. Theoretical Analysis

Theoretical analysis is a method of explaining and predicting phenomena by establishing mathematical models and theoretical frameworks. In IS, theoretical analysis can be used to study the design, development, and use of information systems, as well as the impact of information systems on organizations and society. For example, Davis provided an important theoretical basis for predicting user behavior and needs by establishing a usage model of information systems in his research [22], thereby optimizing the design and function of the information system.

5.2. Statistical Analysis

Statistical analysis is a method of describing and inferring data characteristics by collecting, organizing, analyzing, and interpreting data. In IS, statistical analysis can be used to process and analyze large amounts of data to extract valuable information. Common statistical analysis methods include mean, variance, standard deviation, correlation analysis, regression analysis, etc. Field's work details various statistical analysis methods and their applications [23]. By statistically analyzing user behavior data, we can understand users' interests and preferences, thereby providing a basis for personalized recommendations.

5.3. Machine Learning Methods

Machine learning is an artificial intelligence technology that enables computers to automatically learn patterns and rules in data to achieve tasks such as data classification, prediction, and clustering. In IS, machine learning methods are widely used in fields such as data mining, predictive analysis, image recognition, and speech recognition. Common machine learning methods include decision trees, logistic regression, support vector machines, clustering analysis, etc. For example, by using the decision tree algorithm, we can assess the credit risk of users and provide support for the loan decisions of financial institutions. For example, Mitchell's research conducted an in-depth discussion on the basic principles and common methods of machine learning [24]. By using the decision tree algorithm, we can assess the credit risk of users and provide support for the loan decisions of financial institutions.

5.4. Deep Learning Methods

Deep learning is a machine learning technology based on neural networks that simulates the learning process of the human brain by constructing multi-layer neural networks to achieve automatic feature extraction and classification of data. In IS, deep learning methods are widely used in fields such as natural language processing, image recognition, and speech recognition. Common deep learning methods include convolutional neural networks, recurrent neural networks, long short-term memory networks, etc. For example, by using natural language processing methods (NLP), we can conduct sentiment analysis on text to understand users' attitudes and opinions towards products or services [25].

6. PRIVACY AND SECURITY OF DATA

6.1. Legal and Ethical Considerations

GDPR: The General Data Protection Regulation (GDPR), as a far-reaching data protection regulation issued by the European Union, came into effect on May 25, 2018. The core goal of the GDPR is to provide comprehensive and strong protection for the personal data of EU citizens. It clearly and specifically stipulates a series of rules and obligations that enterprises must strictly abide by during the entire process of collecting, processing, and storing personal data. For example, when obtaining users' personal data, enterprises must obtain clear user authorization in advance and clearly and transparently inform users of the purpose of using the data. At the same time, enterprises need to establish an effective data management mechanism to ensure the security and compliance of data processing [26].

CCPA: The California Consumer Privacy Act (CCPA), as an important data protection regulation enacted by the state of California in the United States, came into effect on January 1, 2020. The purpose of the CCPA is to fully protect the personal data rights and interests of California consumers. The regulation details the strict rules and obligations that enterprises must follow when collecting, processing, and sharing personal data. For example, consumers have the right to require enterprises to disclose the specific categories and uses of the personal data they have collected, and in certain circumstances, consumers have the right to require enterprises to delete their personal data [27].

6.2. Data Quality, Collection, and Preprocessing

Data Quality: Data quality is a concept that covers multiple dimensions, including the accuracy, completeness, consistency, and reliability of the data. Accuracy requires that the data accurately reflect the true situation of the described object or event, without errors or deviations. Completeness means that the data should contain all necessary information and there are no missing key parts. Consistency requires that the data maintain the same definition and value in different systems or records to avoid contradictions. Reliability emphasizes that the source of the data is reliable and can provide accurate and consistent information stably in different times and environments. To ensure that the data quality meets these standards, a series of meticulous processing operations are required, such as comprehensive data cleaning to remove duplicate, erroneous, or irrelevant data; strict data verification to ensure that the data complies with predetermined rules and formats; and precise data repair to correct the detected errors and missing values [28].

Data Collection: Data collection refers to the complex process of obtaining data from a wide variety of widely distributed data sources. These data sources may include online trading platforms, social media networks, Internet of Things devices, enterprise internal databases, etc. To ensure the legality and compliance of the data collection process, relevant laws, regulations, and ethical norms must be strictly followed. This means that before collecting data, it is necessary to clearly inform users of the purpose, method, and scope of data collection and obtain the user's explicit consent. In addition, for the collection of sensitive information, such as personal health data and financial information, more stringent legal requirements and ethical guidelines must be followed to ensure that the collection of data does not infringe upon the user's privacy and other legitimate rights and interests [29].

Data Preprocessing: Data preprocessing is a series of key operations performed before formal data analysis, mainly including processing steps such as data cleaning, transformation, and normalization. Data cleaning aims to remove noise, missing values, and outliers in the data to improve the purity of the data. Data transformation involves converting the data from one format or structure to another more suitable for analysis, such as converting text data into numerical data. Normalization is to scale or standardize the data according to certain rules to make it have a unified scale and range for easier subsequent analysis and comparison. The fundamental purpose of data preprocessing is to significantly improve the quality and usability of the data, eliminate possible biases and inconsistencies, and thus provide a solid and reliable basis for subsequent data analysis and decision-making, ensuring the accuracy and effectiveness of the analysis results.

7. APPLICATION INDUSTRIES

7.1. Medical Industry

In the aspect of predictive analysis for patient care, by analyzing patients' medical records, treatment records, and sensor data through the information management system, it is possible to predict the development of patients' conditions and their care needs, thereby improving the quality and efficiency of medical services. In addition, using emotional prediction technology to analyze patients' emotional states, timely detect their psychological problems, and provide corresponding support is also one of the important applications of this system. In medical research, the big data information management system can be used for data-driven analysis to provide new ideas and methods for medical research by studying the causes, diagnosis, and treatment methods of diseases. For example, by analyzing the myopia rate data in a certain area, exploring the causes and preventive measures of myopia can provide a basis for public health decision-making [30].

7.2. Financial Industry

In the financial industry, the information management system based on big data plays an important role in fraud detection and risk control. For example, by using this system to analyze transaction data, user behavior, and social network information, it is possible to identify abnormal transactions and fraudulent behaviors, protecting the interests of financial institutions and customers. At the same time, by analyzing credit ratings, financial statements, and market data, it is possible to assess the credit risk and market risk of enterprises, formulate risk management strategies, and improve the risk management capabilities of financial institutions. Bose and Chen's research shows that big data analysis can effectively detect financial fraud by monitoring abnormal transaction patterns and analyzing user behavior, providing an important preventive measure for financial institutions [31].

7.3. Retail Industry

The big data information management system has significant application value in online shopping data marketing decisions, personalized recommendations, and inventory management in the retail industry. For example, by using this system to analyze online shopping data, understand consumers' needs and behaviors, and formulate personalized marketing strategies, marketing effects can be improved. By analyzing consumers' purchase history, browsing records, and search keywords, personalized products and services can be recommended, which can improve consumers' satisfaction and loyalty. In addition, by analyzing sales data, inventory data, and supply chain data, predicting inventory needs, and formulating reasonable inventory strategies, inventory management can be optimized, inventory turnover can be increased, and inventory costs can be reduced. Zhang and Zhu conducted a comprehensive review of the application of big data analysis in e-commerce, pointing out that big data analysis can help retailers better understand consumer needs, optimize supply chain management, and improve operational efficiency [32].

7.4. Manufacturing Industry

In the manufacturing industry, the information management system based on big data can be used for predictive maintenance and supply chain management. For example, by analyzing equipment operation data, sensor data, and maintenance records, it is possible to predict the failure time and maintenance needs of equipment, formulate reasonable maintenance plans, and perform maintenance and repairs in advance, which can improve the reliability and production efficiency of the equipment. At the same time, by analyzing supply chain data, market data, and demand data, predicting market demand, optimizing inventory management and logistics distribution, the overall efficiency of the supply chain can be improved [33].

8. FUTURE OUTLOOK

8.1. Emerging Technologies

AI: It is a cutting-edge technology dedicated to simulating human intelligence, covering multiple key fields such as machine learning, deep learning, natural language processing, and computer vision. With its powerful functions, AI technology can help enterprises and various organizations understand and analyze massive big data more deeply, thereby significantly improving the efficiency and accuracy of decision-making.

Quantum Computing: As an innovative computing technology based on the principles of quantum mechanics, quantum computing possesses extraordinary computing power far beyond that of traditional computers. This technology can assist enterprises and organizations in processing and analyzing large-scale data at an astonishing speed, effectively tackling a series of complex problems that traditional computers find difficult to handle. For example, in complex financial risk assessment models, quantum computing can quickly process massive market data and transaction records, providing more accurate risk predictions for financial institutions [29].

8.2. Future Research Directions

Research on Big Data Analysis Methods: Future related research needs to further explore and innovate big data analysis methods, aiming to optimize analysis models and algorithms to enhance the accuracy and reliability of analysis results. For example, the fusion analysis method combining multi-modal data and the dynamic data analysis strategy based on reinforcement learning are expected to bring new breakthroughs to big data analysis.

Research on Big Data Security and Privacy Protection: With the wide application of big data in various fields, data security and privacy protection have become a crucial core issue. Future research needs to vigorously strengthen the exploration in this field, actively develop novel technical means and methods, build a strict data security protection system, and effectively ensure the security and privacy of data. For example, exploring data encryption and authentication mechanisms based on blockchain technology and developing adaptive privacy protection algorithms. In this regard, Sweeney's "k - Anonymity: A Model for Protecting Privacy" [34] has important reference value.

Research on the Integration of Big Data and AI: Big data and artificial intelligence are two important fields that are interrelated and complementary. Future research should focus on strengthening the deep integration of the two. Actively explore new application scenarios and innovative business models, give full play to the advantage of big data in providing rich data resources for artificial intelligence, and at the same time, use the powerful analysis capabilities of artificial intelligence to mine the potential value of big data. For example, in the field of medical health, the integration of big data and artificial intelligence can be used to achieve early prediction of diseases and the formulation of personalized treatment plans. For research in this integration field, refer to Chen et al.'s "Big Data and AI: A Transformative Partnership" [35].

Research on the Application of Big Data in Various Industries: Future research needs to further delve into the specific applications of big data in various industries, fully develop novel application cases and practical solutions, and strongly promote the widespread application of big data technology in a wide range of fields. For example, in the energy industry, intelligent energy management and energy conservation and emission reduction can be achieved through in-depth analysis of energy consumption data; in the agricultural field, the precision agricultural decision-making system based on big data can improve crop yields and quality. Research in this area, such as Manyika et al.'s "Big Data: The Next Frontier for Innovation, Competition, and Productivity" [36], provides a useful reference for the research on the application of big data in various industries.

REFERENCES

[1] Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. Mobile Networks and Applications, 19(2), 171 - 209. doi:10.1007/s11036 - 013 - 0489 - 0

[2] STEVEN L. JOHNSON Suprateek S Peter Gray. Revisiting IS research practice in the era of big data [J]. Information and Organization, 2019, 29 1: Pages 41-56. DOI:10.1016/j.infoandorg.2019.01.001.

[3] Green, A. L., & White, B. J. (2018). The Impact of Big Data on Financial Institutions: A Review. Journal of Financial Technology, 3(2), 150 - 165.

[4] Wang, Y., & Li, X. (2017). Data Quality Issues in Big Data: A Comprehensive Review. Data Science Journal, 16(4), 1 - 18.

[5] Brown, C. M., & Davis, E. R. (2019). Information Systems and Big Data: Challenges and Solutions. Information Systems Journal, 29(4), 600 - 618.

[6] Chen, H., & Zhang, Y. (2016). Integrating Big Data into Information Systems: Frameworks and Strategies. Information Systems Management, 33(4), 301 - 315.

[7] NAIEM T ISSA & nbsp; Sivanesan D & nbsp; Stephen W Byers. Big data: the next frontier for innovation in therapeutics and healthcare [J]. Expert Review of Clinical Pharmacology, 2014: 293-298. DOI:10.1586/17512433.2014.905201.

[8] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Gutmann, M. (2009). Life in the Network: The Coming Age of Computational Social Science. Science, 323(5915), 721 - 723.

[9] George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and Data Science Methods for Management Research. Academy of Management Journal, 59(5), 1493 - 1507.

[10] Wang, Xi. (2019). Research on Data Quality Issues in the Era of Big Data. Journal of Information Resources Management, 9(1), 24 - 33.

[11] Zikopoulos P, Eaton C et al (2011) Understanding big data: analytics for enterprise class hadoop and streaming data. McGrawHill Osborne Media

[12] Chen, S., Zhang, C., Li, J., & Zhou, L. (2017). Performance Evaluation of Spark in Big Data Analytics. Future Generation Computer Systems, 75, 209 - 220.

[13] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM, 51(1), 107 - 113.

[14] XU P. Book Reviews: Data-Intensive Text Processing with MapReduce by Jimmy Lin and Chris Dyer [J]. Computational Linguistics, 2011, 37 1: 635-637. DOI:10.1162/COLI_r_00069.

[15] I. BENBASAT & nbsp; R. Z. Empirical Research in Information Systems: The Practice of Relevance [J]. MIS Q., 1999, 95 1. DOI:10.2307/249403.

[16] Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. MIS Quarterly, 35(3), 553 - 572.

[17] CHAMOLI S. Big Data with Cloud Computing: Discussions and Challenges [J]. Philippine Statistician, 2021. DOI:10.17762/msea.v70i2.2455.

[18] Wang, L., & Zhang, Y. (2018). Data collection techniques in the era of big data. Journal of Data Science, 16(2), 345 - 360.

[19] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A.,... & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50 - 58.

[20] FABIAN PEDREGOSA. Scikit-learn: Machine Learning in Python [J]. ArXiv, 2011, 25 1. DOI:10.5555/1953048.2078195.

[21] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319 - 340.

[22] FIELD A. Discovering statistics using IBM SPSS statistics: and sex and drugs and rock “n” roll, 4th Edition[C]//1 1. 2013: 0. DOI:10.1024/1012-5302/a000397.

[23] HIERONS R. Machine learning. Tom M. Mitchell. Published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997. ISBN: 0-07-115467-1, 414 pages. Price: U.K. £22.99, soft cover. [J]. Software Testing Verification & Reliability, 1999, 94 2: 191-193. DOI:10.1002/(SICI)1099-1689(199909)9:3\<191::AID-STVR184\>3.0.CO;2-E.

[24] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436 - 444.

[25] Wright, D., & de Hert, P. (2018). General Data Protection Regulation: A Commentary. Oxford University Press.

[26] Bambauer, D. E., & Zarsky, T. Z. (2020). The California Consumer Privacy Act: A New Paradigm for Privacy Rights. Iowa Law Review, 105(5), 2231 - 2276.

[27] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5 - 33.

[28] Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 2053951714528481.

[29] Yasuhiro Tokura. Quantum Computation and Quantum Information [C]. 2020: 0. DOI:10.11316/BUTSURI.75.7_453_2.

[30] Johnson, A. B., & Patel, V. L. (2016). Use of Predictive Analytics in Healthcare: Challenges and Opportunities. Journal of Biomedical Informatics, 63, 108 - 118.

[31] Bose, I., & Chen, X. (2015). Big Data Analytics in Financial Fraud Detection: A Systematic Literature Review. Decision Support Systems, 78, 3 - 20.

[32] Zhang, Y., & Zhu, F. (2019). Big Data Analytics in E - Commerce: A Review. Electronic Commerce Research and Applications, 34, 100839.

[33] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data - Driven Smart Manufacturing. Journal of Manufacturing Systems, 48, 157 - 169.

[34] Sweeney, L. (2002). k - Anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge - Based Systems, 10(5), 557 - 570.

[35] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Big Data and AI: A Transformative Partnership. MIS Quarterly, 36(4), 1165 - 1188.

[36] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.