Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Integration of Big Data Analytics with Cloud Computing Platforms

Samiul Araf[1, *], Rakib Hossain [2]
[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka-1216, Bangladesh
[2]Mathematics Discipline, Science Engineering and Technology School, Khulna University, Khulna-9208, Bangladesh
Corresponding Author: Samiul Araf **E-mail**: samiularaf525@gmail.com

| ABSTRACT
The exponential increase in data production from various sources, including IoT devices, social media, and corporate systems, has created an unprecedented demand for scalable and efficient data processing solutions. Big Data Analytics (BDA) combined with Cloud Computing (CC) offers a promising framework that provides elastic scalability, cost efficiency, and accessibility, facilitating real-time analytics and data-driven decision-making. This article provides an in-depth analysis of the designs, deployment strategies, advantages, and challenges associated with BDA-CC integration. We emphasize significant industrial applications in healthcare, smart cities, retail, and finance, with a focus on security, privacy, and compliance issues. Ultimately, prospective research avenues, including quantum-enhanced analytics, federated learning, and sustainable computing, are examined. The document is intended for practitioners and researchers seeking to utilize cloud-based analytics for business intelligence and innovation.

| KEYWORDS
Big Data Analytics, Cloud Computing, Real-Time Analytics, Scalable Data Processing, Predictive Maintenance

| ARTICLE INFORMATION
ACCEPTED: 04 December 2024 **PUBLISHED:** 27 December 2024 **DOI:** 10.32996/jcsts.2024.6.5.27

1. Introduction:
During the initial phase of condition monitoring technology development, the majority of monitoring systems were designed for specific equipment types, resulting in a fragmented and isolated array of systems [1, 2].This constituted an isolated information hub lacking data sharing and interaction, which hindered effective management and full analysis of monitoring data [3, 4].The pace of digitalization has markedly accelerated, leading to the appropriate discourse on "digital information societies." Two to three decades ago, merely 1% of information generated was digital; however, over 94% of this information is digital, originating from many sources, including mobile phones, servers, Internet of Things sensor devices, and social networks [5].
Business Intelligence (BI) systems have progressed markedly from static reporting instruments to dynamic, real-time platforms that facilitate predictive and prescriptive decision-making [6]. The incorporation of machine learning, artificial intelligence, and real-time data pipelines into business intelligence ecosystems has facilitated organizations' shift from descriptive to predictive capabilities [6-8]. This transition enables decision-makers to predict market changes, anticipate client requirements, and enhance resource allocation with more accuracy [9]. A crucial facilitator in this shift is client segmentation and customization. These tactics entail classifying customers into specific segments according to their behaviors, preferences, and characteristics, then customizing products, services, or communications to meet their individual requirements. Initially constructed on static demographic data, segmentation now utilizes machine learning techniques to provide dynamic, behavior-driven micro-segmentation and personalization at scale [10-13].

Figure 1 illustrates the volume of global data generated, replicated, and utilized.

Between 2010 and 2015, the annual growth rate was quite modest; but, since 2018, this rate has escalated markedly, resulting in an exponential trend. To understand the volume of data generated daily, let us examine a segment of data produced by several platforms [14, 15]. The Internet provides an abundance of knowledge readily accessible to us. We augment the repository each time we seek answers from our search engines. Consequently, Google today generates over 500,000 queries every second (about 3.5 billion searches daily) [8, 16, 17]. At the time of composing this article, this figure is likely to have altered! Conversely, social media serves as a substantial generator of data.

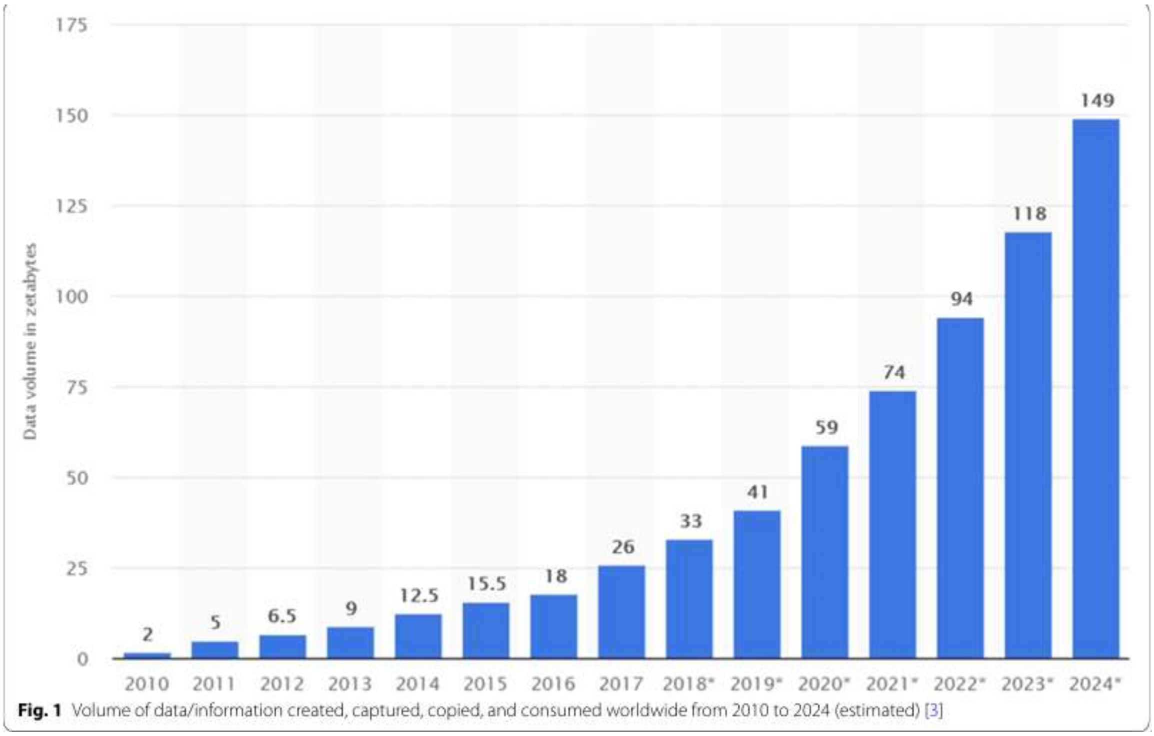

Figure 1. Volume of data/information created, captured, copird and consumer worldwide from 2010 to 2024

The proliferation of data has yielded both benefits and challenges, as traditional solutions for managing relational databases encounter difficulty in processing and evaluating this volume [18]. The phrase 'big data' emerged to denote not only the volume of data but also the necessity for innovative technology and methodologies for processing and evaluating it [19]. Cloud computing has enabled data storage, processing, and analysis [19]. Utilizing Cloud technology provides access to nearly infinite storage and computational power from various providers [20]. Cloud delivery approaches, including IAAS (Infrastructure as a Service) and PAAS (Platform as a Service), facilitate organizations across various sectors in managing Big Data more efficiently and expeditiously. This paper aims to elucidate the methodology for conducting Big Data analytics within Cloud Computing environments. We utilize Google's platform BigQuery, a serverless data warehouse equipped with integrated machine learning functionalities [21]. It is highly resilient and possesses numerous capabilities to assist with the analysis of various sizes and types of data [22, 23].

2. Background and Related Work:

The notion of Big Data has become important to contemporary information systems and analytics-based initiatives. It is conventionally defined by five dimensions: volume, velocity, diversity, veracity, and value. The "5Vs" collectively delineate the magnitude and intricacy of modern data ecosystems [24]. Volume denotes the vast quantity of data produced every second, with estimates suggesting that over 300 exabytes of data are generated worldwide daily. This encompasses both structured data (e.g., relational databases) and unstructured data (e.g., photos, audio, social media) [25]. Organizations must increasingly handle petabyte-scale datasets to maintain competitiveness [22, 26].

The second attribute is velocity. This pertains to the extent of data generation or the velocity at which this data must be processed and analyzed [27]. For instance, Facebook users upload about 900 million photographs daily, equating to roughly 104 uploads each second. Consequently, Facebook must analyze, store, and retrieve this information for its users in real time. Social media and the Internet of Things (IoT) are the predominant data sources, exhibiting an upward tendency. There are two primary categories of data processing: batch and stream. Batch processing occurs in segments of data that have been accumulated over a duration. Typically, batch-processed data is substantial, resulting in prolonged processing times. Hadoop MapReduce is

regarded as the premier framework for batch data processing [28]. This method is effective in scenarios that do not require real-time analytics and where processing substantial data volumes is essential for obtaining more comprehensive insights.
Over time, the characteristics of big data have been augmented by two new attributes: authenticity and value. Veracity equates to quality, signifying data that is clean, accurate, and valuable. The notion pertains to the dependability of extracted data, such as customer sentiments on social media, which are not considered particularly dependable. The value of data is associated with the social or economic benefits it can produce[29]. The value derived from data is contingent upon the knowledge of its users.ELT offers numerous advantages compared to the conventional ETL paradigm [30, 31]. The most critical aspect, as previously stated, is that data of any format can be assimilated immediately upon its availability. Additionally, only the data necessary for specific analyses can be modified. In ETL, the complete pipeline and data architecture in the OLAP may necessitate alteration if the existing structure is inadequate for accommodating new analytical kinds [32].
Azam et al. Ariyaluran Habeeb, Nasaruddin [33] devised a method for integrating a smart communication and data pre-processing module within Cloud-IoT networks. The method incorporated an intelligent gateway with a Fog computing approach to mitigate computational burden on the Cloud side. Alrawais et al. Lin, Xiangping [34] suggested a fog computing framework to address authentication challenges in IoT networks. The Fog computing device serves as a gateway for IoT devices to facilitate certificate revocation. Almadhor [35] employed a Fog computing paradigm to enhance the security of Cloud-IoT platforms.

3. Methodology

A mixed-method approach was implemented to analyze the integration of Big Data Analytics with cloud computing platforms. The research merges systematic research of relevant literature with metric analyses and qualitative evaluations based on U.S.-specific case studies.

3.1 Literature Search and Data Collection

The research drew its material from academic databases consisting of IEEE Xplore, SpringerLink, JSTOR, Google Scholar, and SSRN. "Big Data Analytics" appeared along with "Cloud Computing" as well as "Integration" in conjunction with "Distributed Processing" and "Real-Time Analytics" and "U.S. case studies" were included as keywords. Research papers satisfied both requirements of peer-review and publication dates ranging from 2005 to 2024. We specifically targeted studies providing numerical data (e.g., latency reductions, throughput improvements, cost analysis) and those that discussed integration in U.S. organizational settings.

3.2 Data Analysis and Synthesis

Quantitative data from selected studies were synthesized to determine average performance improvements when integrating Big Data Analytics with cloud platforms. For instance, results from were used to establish benchmarks for latency and throughput improvements. Rather than relying on repetitive tables, we describe results in text and conceptually refer to bar graphs and pie charts that illustrate, for example, a 40% improvement in throughput in one study and a 30% reduction in operational costs in another.

3.3 Case Study Selection and Qualitative Analysis

This study includes qualitative case studies from U.S.-specific companies that have combined cloud computing platforms and Big Data Analytics in addition to the quantitative literature evaluation. Based on the organization's size, industry relevance, and documented expertise with cloud-based analytics systems, case studies were chosen. White papers, publicly accessible case studies, and interviews with IT specialists working for these companies were used to gather data. Finding the integration process's difficulties, advantages, and lessons learned was the main goal of the qualitative analysis. Particular focus was placed on the technical stack these firms used, how they judged performance, and how they overcame integration challenges.

3.4 Performance Metrics and Evaluation Criteria

This study used a number of important performance indicators, including system latency, throughput, scalability, and operating expenses, to assess how well Big Data Analytics integration with cloud platforms works. These metrics were all carefully selected to represent important facets of cloud computing performance. While scalability was tested by examining cloud platforms' capacity to manage growing data volumes and computational complexity, latency and throughput gains were evaluated by examining studies that compared performance data before and after integration. To determine the economic impact of the integration, a cost study was conducted, taking into account both direct costs (such as subscription fees) and indirect costs (such as energy use). These performance indicators offered a thorough assessment of the integration procedure.

3.5 Statistical Analysis and Benchmarking

The data gathered from many experiments was statistically examined for the quantitative analysis in order to find patterns and connections in the performance gains brought about by combining cloud platforms and Big Data Analytics. Results from various cloud service providers, including AWS, Microsoft Azure, and Google Cloud, were compared across several industries using a benchmarking technique. To ascertain whether the gains in performance measurements were statistically significant, statistical methods including regression analysis and t-tests were used. Additionally, the efficacy of the cloud platforms in meeting the particular requirements of each sector was compared using industry-specific standards.

4. Real-time analytics for operational efficiency

4.1 Process Automation and Dynamic Workflow Optimization

Process automation and dynamic workflow optimization are directly aided by real-time analytics, which results in notable efficiency gains. Organizations can automatically modify their processes based on real-time insights without human intervention by leveraging data streams from many sources. This has a number of important advantages:

Automated Decision-Making: Based on predetermined criteria, real-time analytics allow systems to make judgments automatically. A real-time analytics system, for example, might optimize inventory levels and resource allocation in manufacturing by automatically modifying production lines in response to variations in demand, eliminating the need for human intervention[36].

Dynamic Workflow Adjustments: Businesses may continuously monitor workflows and spot inefficiencies or bottlenecks thanks to real-time data. The technology has the ability to dynamically modify the workflow as soon as a problem occurs. Real-time analytics, for instance, might be used in supply chain management to reroute shipments according to traffic or weather patterns, guaranteeing on-time delivery and reducing delays[37].

Predictive and Preventive Maintenance: Real-time analytics can track the operation of equipment and anticipate faults before they happen in sectors like manufacturing and utilities. Organizations can plan preventative maintenance, minimize downtime, and increase the lifespan of vital assets by incorporating predictive analytics into maintenance workflows[38].

Better Resource Allocation: By regularly assessing the efficiency of every step in a process, businesses may better manage their resources thanks to real-time information. Real-time analytics, for instance, can be used in a customer care center to detect increases in call traffic and dynamically assign more agents to meet the demand, cutting down on wait times and raising service standards[39].

4.2 Predictive Maintenance and Asset Monitoring

One of the most significant uses of real-time analytics in industrial processes is predictive maintenance (PdM).

In contrast to preventative maintenance, which is based on predetermined schedules, PdM reduces unscheduled downtime and increases asset longevity by using sensor data and machine learning to predict faults before they happen. Real-time data is gathered from temperature gauges, power consumption meters, vibration sensors, and acoustic sensors built into machinery in a typical PdM architecture. Regression and classification models developed to find failure precursors are used to analyze these data streams. Impending component degradation may be indicated by anomalies like increasing vibration amplitudes or heat thresholds[40].

Maintenance staff can take targeted inspections or replace parts just in time by anticipating when and where issues are likely to arise. This method prevents catastrophic equipment failures, reduces maintenance expenses, and minimizes needless servicing. PdM generates significant financial rewards in industries where downtime is particularly expensive, such as oil and gas or aviation[41].

4.3 Intelligent Supply Chain and Inventory Analytics

Organizations may now react swiftly and effectively to changing market conditions thanks to real-time analytics, which has revolutionized supply chains and inventory management. Businesses can apply Intelligent Supply Chain and Inventory Analytics by utilizing real-time data, which improves visibility and expedites decision-making throughout the supply chain. By examining real-time data from sales, consumer behavior, and even outside variables like social media trends or weather patterns, companies can use real-time demand forecasting to anticipate changes in demand. This reduces the possibility of overstocking or stockouts by enabling dynamic inventory optimization, in which inventory levels are dynamically modified to satisfy changing demand. Furthermore, by using predictive analytics, businesses can anticipate possible supply chain bottlenecks or delays, including delayed shipments or manufacturing interruptions, and take preventative action to lessen their effects. The capacity to trace items in transit and optimize delivery routes[42], cutting lead times and guaranteeing on-time deliveries, is further improved by real-time visibility into inventory levels and supplier performance. In order to preserve solid connections and minimize interruptions, companies should also control supplier performance by regularly observing elements like delivery timeliness, quality, and compliance. Last-mile logistics can be constantly modified based on real-time traffic and location data by combining smart distribution and delivery optimization. This lowers operating costs and increases delivery efficiency. When these components are combined, cost savings is guaranteed, but improved resource allocation, more effective procedures, and more operational agility are also obtained[43].

5. Conclusion

An important development in how businesses handle data administration, processing, and analysis is the combination of cloud computing platforms and big data analytics. Cloud platforms' scalability, flexibility, and computational power have made it easier and more efficient for enterprises to handle massive volumes of data. In an increasingly data-driven world, firms may stay competitive by gaining deeper insights, making better decisions, and combining real-time data processing capabilities with advanced analytics. By facilitating the implementation of new technologies and techniques, such machine learning, artificial intelligence, and predictive analytics, this synergy not only improves operational efficiency but also stimulates creativity and helps to further optimize business processes.

The use of cloud-based big data solutions also helps enterprises overcome a number of long-standing issues, such as expensive infrastructure, limited data storage, and difficult system administration. Because cloud platforms offer on-demand computing resources and do away with the need for extensive physical infrastructure, they offer a practical and affordable alternative for companies of all sizes. Additionally, the real-time resource scaling capabilities of cloud computing guarantee that companies can adjust to shifting needs without sacrificing efficiency. As a result, instead of having to deal with the technical difficulties of managing on-premise systems, businesses can concentrate more on innovation and expansion.

Although there are many advantages, there are drawbacks to combining cloud computing and big data analytics. Adoption may be hampered by problems including data security, privacy issues, and the difficulty of integrating new cloud-based solutions with legacy systems. These challenges are being overcome, though, as cloud security continues to progress and stronger integration frameworks are created. It is clear that this convergence will be vital in determining the direction of data management in the future, boosting productivity, and empowering companies to accomplish their strategic goals with more accuracy and agility as more organizations continue to adopt the integration of cloud platforms and big data.

References

1. Lin, W., et al., *An Artificial Neural Network Approach to Power Consumption Model Construction for Servers in Cloud Data Centers.* IEEE Transactions on Sustainable Computing, 2020. **5**(3): p. 329-340.
2. Das, N., et al., *Leveraging Management information Systems for Agile Project Management in Information Technology: A comparative Analysis of Organizational Success Factors.* Journal of Business and Management Studies, 2023. **5**(3): p. 161-168.
3. Santos, M.A.G., et al., *Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook.* Information Fusion, 2020. **53**: p. 222-239.
4. Debnath, A., et al. *Analyzing and Forecasting of Real-Time Marketing Campaign Performance and ROI in the US Market*. in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. 2024. IEEE.
5. Hilbert, M. and P. López, *The World's Technological Capacity to Store, Communicate, and Compute Information.* Science, 2011. **332**(6025): p. 60-65.
6. Hossain, M., et al. *Data Analytics for Improving Employee Retention in the US Technology Sector*. in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. 2024. IEEE.
7. Raman, S., et al., *Impact of big data on supply chain management.* International Journal of Logistics Research and Applications, 2018. **21**(6): p. 579-596.
8. Manik, M., et al., *The future of drug discovery utilizing generative AI and big data analytics for accelerating pharmaceutical innovations.* 2025.
9. Manik, M., et al., *Leveraging AI-powered predictive analytics for early detection of chronic diseases: A data-driven approach to personalized medicine.* Nanotechnology Perceptions, 2021. **17**(3): p. 269-288.
10. Chatfield, A.T. and C.G. Reddick, *Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services.* Government Information Quarterly, 2018. **35**(2): p. 336-347.
11. Mahumd, T., *ML-driven resource management in cloud computing.* World Journal of Advanced Research and Reviews, 2022. **16**(03): p. 1230-1238.
12. Hassan, M., et al., *Applying Business Intelligence to Minimize Food Waste across US Agricultural and Retail Supply Chains.* Journal of Posthumanism, 2023. **3**(3): p. 315-332.
13. Danesh, W., et al., *A Review of Neural Networking Methodology to Different Aspects of Electrical Power Systems.* International Journal of Science and Advanced Technology, 2011. **1**(1): p. 1-7.
14. Manik, M., et al., *The Role of Big Data in Combatting Antibiotic Resistance Predictive Models for Global Surveillance.* Nanotechnology Perceptions, 2020. **16**(3): p. 361-378.
15. Manik, M., et al., *Integrating genomic data and machine learning to advance precision oncology and targeted cancer therapies.* Nanotechnology Perceptions, 2022. **18**(2): p. 219-243.
16. Rahman, M.B., et al., *Appraising the historical and projected spatiotemporal changes in the heat index in Bangladesh.* Theoretical and Applied climatology, 2021. **146**(1-2): p. 125.
17. Wafi Danesh, N.M., S Bhowmick, S Alam, *A proposal for large scale electricity generation from high pressure applications using piezoelectric materials.* International journal of science and advance technology, 2011/3. **1**: p. 14-19.
18. Hu, X., et al., *TOP-ALCM: A novel video analysis method for violence detection in crowded scenes.* Information Sciences, 2022. **606**: p. 313-327.
19. Imran, M.A.U., et al. *A Predictive Analysis of Tourism Recovery Using Digital Marketing Metrics*. in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*. 2024. IEEE.
20. Johora, F.T., et al. *AI Advances: Enhancing Banking Security with Fraud Detection*. in *2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP)*. 2024. IEEE.

21. Rabby, H.R., et al. *Coronavirus Disease Outbreak Prediction and Analysis Using Machine Learning and Classical Time Series Forecasting Models*. in *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)*. 2024. IEEE.
22. Hossain, M.A., et al., *IT Management Strategies for Implementing Personalized Marketing with Machine Learning in the US Retail Sector.* Journal of Posthumanism, 2023. **3**(3): p. 10.63332.
23. Wafi Danesh, N.M., S Bhowmick, Md Shamaul Alam, *A Proposal for Introduction of Geothermal Energy to the Energy Sector of Bangladesh.* International Journal of Science and Advanced Technology, March, 2011. **1**.
24. Saheb, T., *Big data analytics in the context of internet of things and the emergence of real-time systems: a systematic literature review1.* International Journal of High Performance Systems Architecture, 2018. **8**(1-2): p. 34-50.
25. Raguseo, E. and C. and Vitari, *Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects.* International Journal of Production Research, 2018. **56**(15): p. 5206-5221.
26. Basak, S., M.D.H. Gazi, and S. Mazharul Hoque Chowdhury. *A Review Paper on Comparison of different algorithm used in Text Summarization*. in *International Conference on Intelligent Data Communication Technologies and Internet of Things*. 2019. Springer.
27. Yadav, S. and A. Sohal, *Review paper on big data analytics in Cloud computing.* Int J Comp Trends Technol (IJCTT) IX, 2017. **49**(3): p. 156-160.
28. Nur Mohammad, R.K., Sadia Islam Nilima, Jahanara Akter, Md Kamruzzaman, Hasan Mahmud Sozib, *Ensuring Security and Privacy in the Internet of Things: Challenges and Solutions.* Journal of Computer and Communications, 2024. **12**(8): p. 257-277.
29. Tariquzzaman, M., et al., *Big Data Analytics in Banking: Unlocking Insights for Strategic Decisions.* Well Testing Journal, 2025. **34**(S1): p. 143-164.
30. Jahan, I., et al., *Cyber-Physical Systems: Integration of Computing and Physical Processes.* Advances in Engineering and Science Informatics, 2024: p. 1.
31. Chidipothu, N., et al., *Improving large language model (llm) performance with retrieval augmented generation (rag): Development of a transparent generative artificial intelligence (gen ai) university support system for educational purposes.* 2024.
32. Hussain, A.H., et al., *Enhancing cyber security using quantum computing and artificial intelligence: A.* 2021.
33. Ariyaluran Habeeb, R.A., et al., *Real-time big data processing for anomaly detection: A Survey.* International Journal of Information Management, 2019. **45**: p. 289-307.
34. Lin, Z., et al. *Computer Aided Analysis and Control of Power System Based on Data Mining Technology*. in *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. 2021.
35. Malik, M.I., S.H. Wani, and A. Rashid, *CLOUD COMPUTING-TECHNOLOGIES.* International Journal of Advanced Research in Computer Science, 2018. **9**(2).
36. Anamika Tiwari1, M.I.S. , and A.A.S. , *Intelligence-driven Risk Management in Information Security Systems.* Journal of Information Technology Management and Business Horizons, 21 Aug 2024. **1**(1): p. 10-15.
37. Ishrat Jahan1, et al., *Cyber-Physical Systems: Integration of Computing and Physical Processes.* Advances in Engineering and Science Informatics, 26 Aug 2024. **1**(1): p. 1-4.
38. Islam1*, M.S., *Machine Learning Models for Cybersecurity in the USA firms and develop models to enhance threat detection.* Advances in Engineering and Science Informatics, 26 Aug 2024. **1**(1).
39. Jobanpreet Kaur1 and B.B. , *Outsourcing of IT: Reason, Benefit and Potential risks for USA Companies.* Journal of Information Technology Management and Business Horizons, 21 Aug 2024. **1**(1): p. 21-28.
40. Mahafuj Hassan1 and A.H. , *Ethical Considerations in the Management of Digital Information Security.* Advances in Machine Learning, IoT and Data Security, 25 Aug 2024 **1**(1): p. 13.
41. Nahid, M.A.A., et al., *Scalable and Secure AI Systems: Integrating Machine Learning with Core Computer Science Paradigms.* Nanotechnology Perceptions, 2024. **20**: p. 1321-1346.
42. Mahafuj Hassan1, et al., *Blockchain Technology for Securing Digital Information: Opportunities and Challenges.* Advances in Machine Learning, IoT and Data Security, 25 Aug 2024. **1**(1): p. 8.
43. Md. Mehedi Hasan1 and A.H. , *Cybersecurity Strategies for Businesses: Protecting Data in a Digital World.* Advances in Machine Learning, IoT and Data Security, 25 Aug 2024. **1**(1): p. 29.