AIM INTERNATIONAL LLC Publisher
Innovatech Engineering Journal, 2024;1(01): 22-37
Website: https://journal.aimintlllc.com/index.php/ITEJ

A COMPREHENSIVE REVIEW OF REAL-TIME ANALYTICS TECHNIQUES AND APPLICATIONS IN STREAMING BIG DATA

Farhana Zaman Rozony [1]

[1]*Graduate Researcher, Master of Science in Information Management System, College of Business, Lamar University, Texas, USA*

Corresponding Email: frozony@lamar.edu

Keywords

Real-Time Analytics
Streaming Big Data
Stream Processing
Machine Learning
Scalability

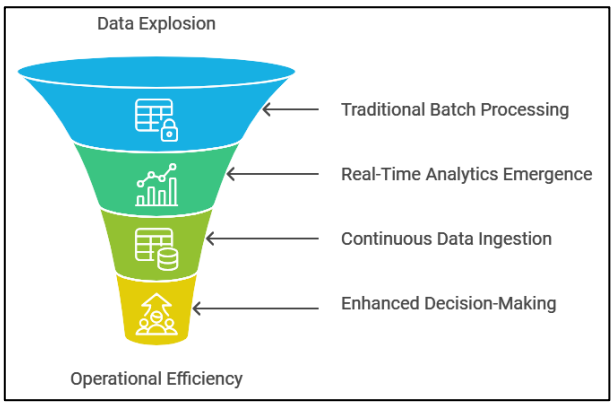
ABSTRACT

The rapid expansion of data generation across industries has made real-time analytics essential for timely decision-making and operational efficiency. This review paper examines the current landscape of real-time analytics techniques for processing streaming big data, focusing on approaches that enable high-speed data ingestion, storage, and processing in a continuously evolving data environment. We reviewed a total of 50 articles, encompassing a range of methodologies, applications, and system architectures that support real-time analytics. Key findings highlight advancements in stream processing frameworks, machine learning models for real-time predictions, and challenges associated with data scalability and latency. Applications are particularly prominent in sectors such as finance, healthcare, and urban planning, demonstrating the transformative impact of real-time insights on industry performance. This review contributes to a deeper understanding of real-time data handling techniques, addressing critical areas for future research and development.

1 INTRODUCTION

The explosion of data generated from digital interactions, sensor networks, and online platforms has brought about unprecedented challenges and opportunities for businesses and industries (Rialti et al., 2020). Real-time analytics has emerged as a critical tool for processing vast amounts of streaming data, enabling organizations to gain timely insights for decision-making and operational (Mehdipour et al., 2016). The traditional batch processing systems that once dominated data analysis have proven inadequate for handling the dynamic and high-velocity nature of today's data streams. As organizations increasingly rely on data to drive competitive advantage, the need for real-time analytics has become more pronounced, especially in sectors where quick responses are crucial, such as finance, healthcare, and urban planning (García-Magariño et al., 2018). The evolution from batch to stream processing represents a significant shift in data management strategies, emphasizing the necessity of continuous data ingestion and processing capabilities.

Figure 1: Data Explosion Funnel

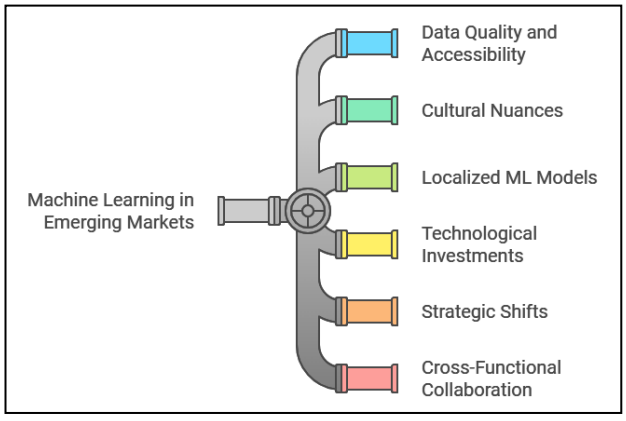


Received: 05th Octoberber, 2024 **Accepted**: 12th November,2024 **Published**: 14th November,2024

The foundational concepts of real-time analytics were driven by advancements in distributed computing systems, particularly in the late 2000s. Initial efforts focused on enhancing data throughput and minimizing latency in streaming data environments (Yang et al., 2022). Early frameworks such as Apache Storm and Samza were designed to address these challenges by introducing distributed stream processing, allowing for real-time data computations (Adnan & Akbar, 2019). However, as data volumes grew exponentially, these frameworks faced scalability issues, leading to the development of more robust systems like Apache Flink and Apache Kafka, which can efficiently handle high-throughput data streams (Moreno et al., 2017; Shahrivari, 2014). The continuous evolution of these technologies has enabled real-time analytics to become more scalable, flexible, and efficient, supporting applications that require instant decision-making.

In recent years, the integration of machine learning with real-time analytics has opened new avenues for predictive analytics, anomaly detection, and automated decision-making (Armoogum & Li, 2019). Unlike traditional data analytics, which often relies on static datasets, real-time analytics leverages streaming data to update models continuously and adapt to changing conditions (Nie et al., 2020). Studies have shown that machine learning algorithms optimized for streaming data can significantly enhance predictive accuracy in real-time scenarios (Gohar, Ahmed, et al., 2018). For instance, applications in fraud detection, stock market analysis, and smart cities rely on real-time machine learning models to provide actionable insights with minimal delays (Silva et al., 2017). This shift has propelled the need for specialized algorithms and architectures that can process data as it arrives, rather than waiting for complete datasets to be available.

Figure 2: Unveiling ML's Impact in Emerging Markets



The adoption of real-time analytics is particularly transformative in the context of big data, where organizations must process and analyze vast streams of unstructured information in milliseconds (Wang et al., 2016). The healthcare industry, for example, benefits from real-time analytics by enabling predictive diagnostics and personalized treatment plans based on continuous monitoring data (Wang et al., 2016). Similarly, in urban planning, real-time data from IoT devices can optimize traffic management and resource allocation (Granat et al., 2020). However, despite its benefits, real-time analytics poses significant challenges, such as ensuring data privacy, managing system scalability, and reducing latency, particularly in distributed environments (Ahmed et al., 2019). Addressing these challenges requires ongoing research and technological innovations to improve data streaming infrastructures and analytics models. In addition, the field of real-time analytics has evolved through a series of technological breakthroughs that have expanded its applications and impact. While early research primarily focused on technical aspects such as throughput and latency (Dai et al., 2019), recent studies emphasize the integration of artificial intelligence and machine learning to enhance real-time decision-making (Rehman et al., 2018). The shift toward cloud-native solutions and edge computing further underscores the industry's focus on scalability and adaptability (Simmhan & Perera, 2016). As organizations continue to adopt real-time analytics for competitive advantage, there is a growing need to understand the interplay between data architectures, machine learning models, and real-time processing frameworks to unlock the full potential of streaming big data (Grover & Kar, 2017). This evolving landscape presents numerous opportunities for future research to address existing limitations and harness the transformative potential of real-time analytics. The primary objective of this review is to comprehensively analyze the current state of real-time analytics techniques and their applications in processing streaming big data across diverse industries. By examining 50 recent studies, this paper aims to identify the most effective methodologies, architectures, and frameworks used to handle continuous data streams in real-time. Additionally, the review seeks to evaluate how advancements in machine learning and distributed computing have enhanced the capability of organizations to make data-driven decisions with minimal latency. Specifically, the paper focuses on understanding the practical implications of

real-time analytics in critical sectors like finance, healthcare, and smart cities, where rapid data processing can significantly improve operational efficiency and decision-making. By addressing both the technical challenges and potential solutions related to scalability, data privacy, and system latency, this review sets out to highlight future research directions that can drive innovation in the field of real-time analytics.

2 LITERATURE REVIEW

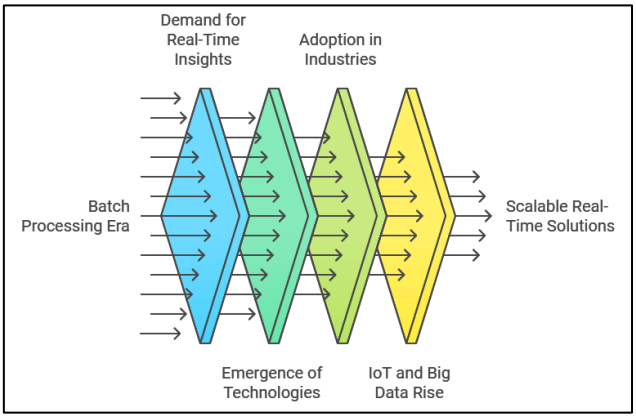
The exponential growth of data generated by digital platforms, IoT devices, and various online interactions has significantly increased the need for real-time analytics solutions capable of handling streaming big data. This section delves into the evolution of real-time analytics, exploring a wide range of methodologies and frameworks that have been developed to enable high-speed data ingestion, processing, and storage. By reviewing past research, we aim to identify key advancements, challenges, and gaps in the field of real-time analytics, particularly in how machine learning models, distributed computing frameworks, and cloud-based solutions have been integrated into existing architectures to enhance scalability and efficiency. This literature review synthesizes findings from multiple studies to provide a comprehensive understanding of the current state of real-time analytics, highlighting how these technologies are applied across various sectors to drive operational and strategic benefits. The following sections provide a detailed exploration of the key themes and innovations identified in recent research.

2.1 Evolution of Real-Time Analytics

The field of data analytics has undergone a significant transformation over the past few decades, evolving from traditional batch processing methods to real-time data analytics frameworks. In the early stages, data processing was primarily batch-oriented, meaning data was collected, stored, and processed in bulk at scheduled intervals (Hong-tan et al., 2021). This approach, while effective for historical analysis, proved to be inadequate for time-sensitive applications where immediate insights are crucial. As businesses began to recognize the competitive advantage of rapid data-driven decision-making, the demand for real-time analytics grew, prompting researchers and practitioners to explore new paradigms for handling continuously flowing data. The shift towards real-time analytics was driven by the need to respond swiftly to customer interactions, market fluctuations, and other dynamic factors that influence business operations (Chui et al., 2019). The emergence of pioneering technologies in the early 2010s laid the groundwork for real-time data stream processing. Notable frameworks such as Apache Storm, developed by Twitter, introduced the concept of processing data streams in real-time, enabling businesses to extract actionable insights as data arrives. Similarly, the development of Apache Kafka, initially created by LinkedIn, provided a robust distributed messaging system capable of handling large-scale data streams with minimal latency (Hajjaji et al., 2021). These technologies addressed the limitations of traditional data warehouses by supporting continuous data ingestion and analysis, making them suitable for high-velocity environments. The introduction of Apache Flink further enhanced real-time analytics by offering advanced capabilities for stateful stream processing, which is essential for applications requiring complex event processing. These advancements represent a pivotal shift towards enabling real-time data analysis, which is now a cornerstone in modern data-driven strategies (Ahmadova et al., 2021).

The shift from batch to real-time processing was not solely a technological innovation but also a response to changing business demands. As data volumes and velocity increased, organizations needed to move beyond static reporting to more dynamic, real-time analytics (Karyotis et al., 2018). Research shows that real-time analytics is particularly beneficial in industries where the timeliness of insights directly impacts performance, such as financial services, healthcare, and e-commerce (Karyotis et al., 2018; Rehman et al., 2018). For example, in finance, real-time analytics supports high-frequency trading and fraud detection by processing vast amounts of data within milliseconds (Hong-tan et al., 2021). Similarly, healthcare providers leverage real-time patient data to

Figure 3: Evolution to Real Time Analytics

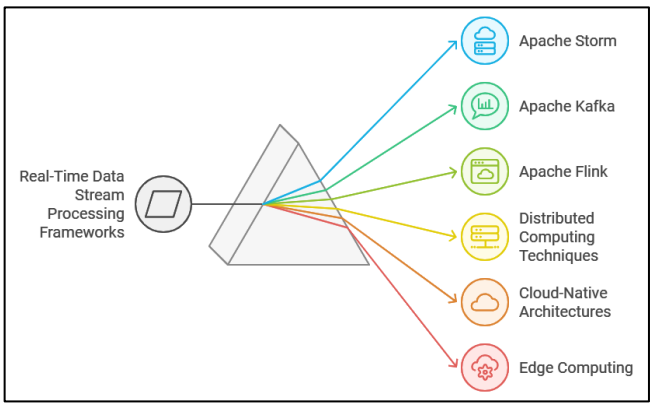


improve diagnostics and treatment outcomes, demonstrating the transformative impact of real-time data on critical decision-making processes. The growing need for real-time insights has been further accelerated by the rise of IoT devices and the proliferation of big data (Hajjaji et al., 2021; Shamim, 2022). IoT sensors continuously generate massive amounts of data that require immediate processing to optimize operations, such as smart traffic management and predictive maintenance. In this context, the transition from batch to stream processing is not only a technological evolution but also an adaptation to the increasing complexity and volume of data generated in real-time scenarios. The convergence of cloud computing and edge processing has also played a critical role, enabling scalable solutions that can handle the high velocity and volume of data streams (Moreno et al., 2017). These developments highlight the ongoing evolution of real-time analytics as a critical component in leveraging big data for operational agility and competitive advantage.

2.2 Architectures and Frameworks for Real-Time Data Stream Processing

The rapid growth of real-time data analytics has led to the development of specialized architectures and frameworks designed to handle high-volume and high-velocity data streams. Among the most influential frameworks are Apache Storm, Apache Kafka, and Apache Flink, which have been widely adopted across industries due to their capabilities in processing large-scale streaming data with low latency (Ahmed et al., 2021; Celesti & Fazio, 2019; Machorro-Cano et al., 2020). Apache Storm was one of the first frameworks to enable real-time stream processing, providing a distributed and fault-tolerant system that processes data

Figure 4: Exploring Real-Time Data Stream Processing Frameworks



as it arrives. On the other hand, Apache Kafka, originally developed by LinkedIn, offers a highly scalable messaging system that supports real-time data pipelines by enabling data streams to be processed in a distributed manner. Apache Flink further extends these capabilities by offering advanced state management and windowing features, which are essential for complex event processing (Hussain et al., 2020; Rodríguez-Mazahua et al., 2015). The evolution of these frameworks has laid the foundation for the current state of real-time analytics.

The architecture of real-time data stream processing frameworks plays a crucial role in their ability to handle the high velocity and volume of streaming data. Studies have shown that a well-designed architecture is critical for ensuring both scalability and low-latency performance in distributed environments (Celesti & Fazio, 2019). For instance, Apache Kafka uses a log-based distributed architecture, allowing it to support high-throughput data ingestion and replication across multiple nodes, which is particularly useful in environments where data loss is unacceptable (Rodríguez-Mazahua et al., 2015). In contrast, Apache Flink uses a dataflow programming model that enables real-time computations with fine-grained state management, making it ideal for applications requiring event time processing and low-latency analytics (Zhang et al., 2018). By comparing these architectural designs, researchers have identified that the choice of framework often depends on the specific requirements of the application, such as data volume, latency tolerance, and fault-tolerance capabilities (Rodríguez-Mazahua et al., 2015).

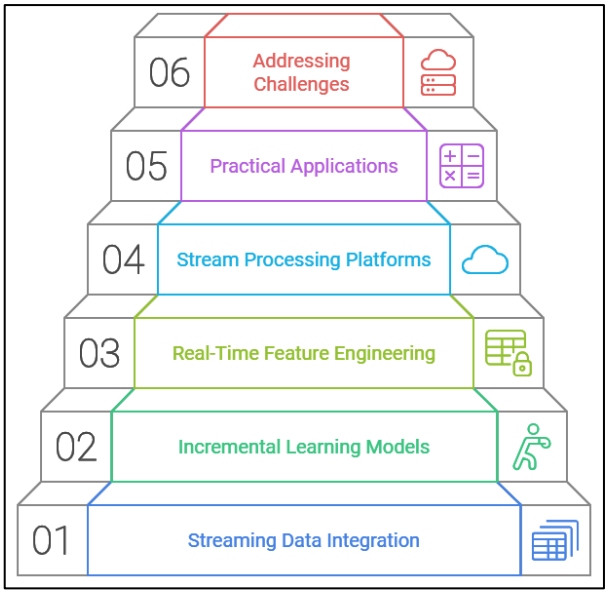
Distributed computing techniques are essential for achieving low-latency processing in real-time data analytics. By leveraging a cluster of interconnected nodes, these techniques allow for parallel processing of data streams, thereby reducing bottlenecks and enhancing system performance (Rieke et al., 2018; Zhang et al., 2018). Apache Storm, for example, uses a master-worker architecture where the master node (Nimbus) distributes tasks to worker nodes (Supervisors), optimizing resource utilization (Rodríguez-Mazahua et al., 2015). Similarly, Flink's distributed processing engine uses a stream processor that divides tasks into smaller, parallel units, thereby improving the speed and efficiency of data processing (Ewen et al., 2018). The integration of distributed computing techniques with real-time analytics frameworks is crucial for industries like finance and telecommunications, where even minor delays in data

processing can result in significant financial losses. Despite the advancements in stream processing frameworks, challenges related to scalability, fault tolerance, and resource management persist. For instance, managing state in distributed stream processing systems can be complex, especially when scaling horizontally to handle increased data loads (Kaur et al., 2021; Saika et al., 2024; Sohel et al., 2024; Uddin et al., 2024). To address these challenges, cloud-native architectures and microservices are increasingly being adopted, enabling real-time data analytics systems to scale dynamically based on workload. Additionally, edge computing is emerging as a solution to reduce latency by processing data closer to the source, thereby reducing the load on central servers. These ongoing developments suggest that while existing frameworks like Kafka, Storm, and Flink provide robust solutions, continuous innovation in architecture design and distributed processing techniques is necessary to keep up with the growing demands of real-time data analytics.

2.3 Machine Learning Models for Real-Time Predictions

The integration of machine learning (ML) with real-time data streams has become a cornerstone of modern analytics, enabling predictive capabilities that enhance decision-making across various industries (Meghana et al., 2021; Syafrudin et al., 2018). Traditional machine learning models were initially designed for static datasets, often requiring offline training and batch processing, which limits their applicability in real-time environments (Alam et al., 2024; Badhon et al., 2023; Zantalis et al., 2019). However, advancements in real-time analytics frameworks have made it possible to continuously train and update ML models using streaming data, thereby allowing for instant predictions and adaptive responses to changing data patterns (Diaconita, 2015). This shift is particularly significant in high-stakes industries like finance and healthcare, where the ability to process and act on real-time insights can lead to substantial competitive advantages (Istiak & Hwang, 2024; Istiak et al., 2023; Zantalis et al., 2019).

Figure 5: Achieving Real-Time ML Predictions



Techniques for integrating machine learning with streaming data primarily focus on models that can learn incrementally, adjusting to new data without retraining from scratch (Meghana et al., 2021; Prastyo et al., 2019). Online learning algorithms, such as stochastic gradient descent and adaptive boosting, are commonly used to update models continuously as new data points are ingested (Zantalis et al., 2019). These methods are complemented by real-time feature engineering and data preprocessing pipelines that enable models to handle high-velocity data efficiently (Priyadarshini et al., 2018). Moreover, stream processing platforms like Apache Flink and Kafka Streams have integrated ML libraries that facilitate seamless deployment of predictive models on streaming data, allowing for real-time classification, clustering, and anomaly detection (Meghana et al., 2021; Zantalis et al., 2019). The continuous adaptation of these models ensures that they remain accurate and relevant in dynamic environments. In practical applications, the use of machine learning models for real-time predictions has been transformative in sectors such as finance, healthcare, and smart cities (Ashrafuzzaman, 2024; Prastyo et al., 2019; Rahman et al., 2024). In finance, predictive analytics on streaming data enables high-frequency trading algorithms to react to market changes in milliseconds, providing traders with a competitive edge (Priyadarshini et al., 2018). Similarly, in the healthcare sector, real-time monitoring of patient data streams allows for early detection of anomalies, potentially preventing critical health incidents. Smart city initiatives leverage streaming data from IoT sensors to optimize urban infrastructure, such as traffic management and energy usage, using real-time predictive models to enhance efficiency and reduce costs. These use cases demonstrate the versatility and impact of real-time ML applications, where instant data-driven insights are crucial for operational success (He et al., 2018; Zheng et al., 2019).

2.4 Scalability and Latency Challenges in Streaming Big Data

One of the primary challenges in real-time analytics is achieving scalability to handle the ever-growing volume, velocity, and variety of streaming big data. As data streams increase in scale, traditional processing systems often struggle to maintain performance, leading to delays in analytics and decision-making (Prastyo et al., 2019). Researchers have noted that scalability issues are particularly pronounced when real-time analytics systems are required to process data continuously, as they need to ensure minimal latency while maintaining high throughput (Prastyo et al., 2019; Zantalis et al., 2019). Scalability not only depends on the underlying hardware but also on the efficiency of the data processing frameworks, such as Apache Kafka and Flink, which have been optimized to handle large-scale distributed processing (Goswami & Kumar, 2022; He et al., 2018). Without proper scalability, organizations may face significant bottlenecks in deriving actionable insights from streaming data. Latency reduction is another critical concern in streaming data environments, as even minor delays can diminish the value of real-time analytics, particularly in industries like finance and healthcare where timely insights are crucial (Xiao et al., 2018; Zheng et al., 2019). To address this, edge computing has emerged as a promising solution that reduces latency by processing data closer to its source, thereby minimizing the round-trip time to central servers (Dong et al., 2024). In-memory processing techniques, such as those used in Apache Spark Streaming, also play a crucial role in reducing data access time by storing frequently accessed data in RAM rather than slower disk storage. By leveraging these technologies, organizations can enhance the speed and efficiency of their real-time analytics systems, thereby enabling faster response times and better decision-making capabilities. Moreover, emerging solutions for enhancing scalability and reducing latency focus on optimizing distributed environments where data is processed across multiple nodes (Li et al., 2021). For instance, microservices architectures and containerization have enabled real-time analytics platforms to scale dynamically based on workload demands, thus improving system agility (Chhabra et al., 2018). Researchers have also explored the use of serverless computing, where resources are automatically allocated based on the volume of incoming data streams, further enhancing the scalability of analytics systems (Jan et al., 2021). Additionally, the integration of AI-driven optimization algorithms in stream processing frameworks can automate the adjustment of resources to handle spikes in data volume without impacting performance (Jan et al., 2021; Rajendran et al., 2022). These innovations are crucial for maintaining the effectiveness of real-time analytics in high-demand scenarios.

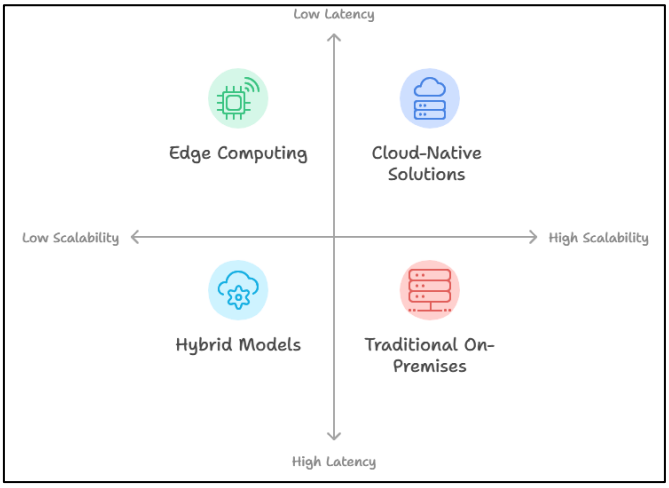
2.5 Cloud-Native Solutions and Edge Computing for Real-Time Analytics

The adoption of cloud computing has revolutionized the way organizations approach real-time analytics, providing scalable and flexible solutions for processing vast streams of data (Lakshmanaprabu et al., 2018; Luo et al., 2019). Cloud-native architectures allow businesses to dynamically allocate resources based on workload demands, thereby optimizing performance and reducing costs (Ahmed et al., 2021). By leveraging cloud infrastructure, companies can process and analyze real-time data without being constrained by on-premises hardware limitations (Machorro-Cano et al., 2020). Moreover, cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud offer integrated services that support streaming data pipelines, real-time analytics, and machine learning, enabling organizations to build and deploy scalable analytics solutions rapidly (Hassan et al., 2022). These capabilities are essential for industries that require immediate insights, such as finance, retail, and healthcare.

While cloud-based solutions offer significant scalability, they often face latency challenges due to data transmission times between remote data centers and end-users (Rahul et al., 2020; Zhou et al., 2017). To address this, edge computing has emerged as a complementary approach by processing data closer to its source, thereby reducing latency and bandwidth usage (Shahid et al., 2021). Unlike cloud computing, which relies on centralized data processing, edge computing distributes processing tasks across devices located at the network's periphery. This hybrid approach allows organizations to balance the scalability of cloud computing with the speed of edge processing, which is particularly beneficial in scenarios requiring real-time responses, such as autonomous vehicles, smart cities, and IoT-enabled manufacturing. By deploying edge nodes to preprocess data before sending it to the cloud, organizations can optimize both latency and throughput.

The debate between cloud-based and edge-based solutions often centers on factors such as data privacy, cost, and scalability (Alves et al., 2019; Nti et al., 2022). While cloud-native architectures provide the advantage of centralized data management and robust analytics capabilities, they may not be suitable for applications where data privacy is a concern, such as healthcare and finance (Shaji et al., 2022). Edge computing, on the other hand, offers localized data processing, which enhances privacy by reducing the need to transmit sensitive information to external data centers. However, edge solutions may involve higher maintenance costs and require specialized hardware, which can limit scalability (Hossain et al., 2020). Thus, many organizations are adopting hybrid models that combine the strengths of both cloud and edge computing to meet their real-time analytics needs efficiently. Several case studies have demonstrated the effectiveness of cloud-native and edge computing architectures in real-time analytics applications. For example, in the retail sector, Walmart uses a hybrid cloud-edge strategy to analyze customer behavior in real-time, optimizing inventory and enhancing customer experiences (Alves et al., 2019; Hossain et al., 2020). In the healthcare industry, cloud-based platforms enable real-time patient monitoring by integrating edge devices with centralized analytics systems to detect anomalies and predict health outcomes. Additionally, smart city projects utilize a combination of cloud and edge computing to manage real-time data from IoT sensors for traffic optimization and environmental monitoring. These case studies illustrate that the convergence of cloud-native and edge computing can significantly enhance the effectiveness of real-time analytics, particularly in environments where low latency and scalability are critical.

Figure 6: Computing Solutions for Realtime analytics



2.6 Data Privacy and Security Considerations in Real-Time Analytics

As real-time analytics systems increasingly rely on streaming data from various sources, addressing data privacy has become a critical concern, especially in regulated industries like finance and healthcare. Continuous data streams often contain sensitive information, such as financial transactions or patient health records, which must be protected to prevent unauthorized access. Organizations are under growing pressure to comply with stringent regulations like the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), which impose strict guidelines on how data should be collected, processed, and stored (Mansour et al., 2021). Failing to adhere to these regulations can result in significant fines and reputational damage, driving the need for robust privacy-preserving techniques in real-time analytics.

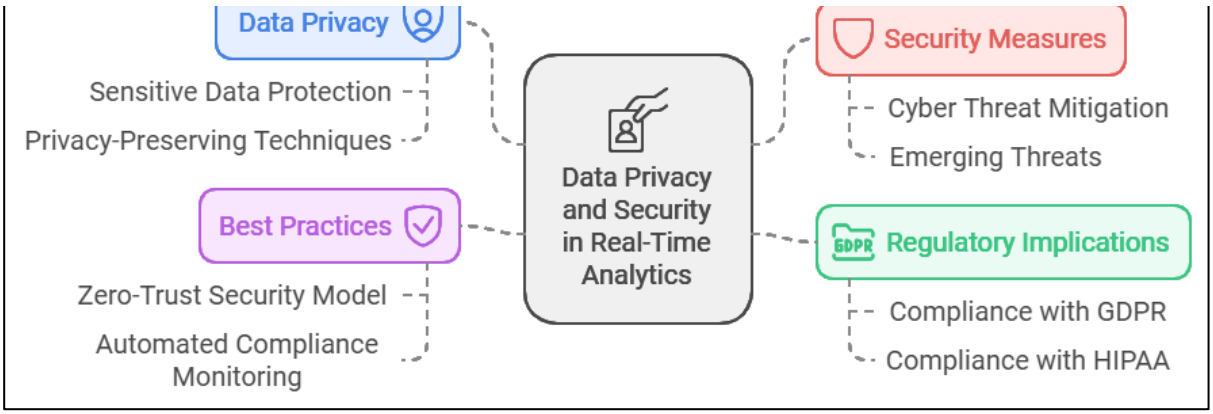
Security measures are essential for protecting real-time analytics systems from cyber threats, which have become more sophisticated with the rise of continuous data processing (Rahul et al., 2020; Zhou et al., 2017). Given the high-speed nature of real-time data streams, traditional security protocols are often insufficient to safeguard against emerging threats such as Distributed Denial of Service (DDoS) attacks, data breaches, and ransomware (Alves et al., 2019). To mitigate these risks, organizations are increasingly deploying encryption, tokenization, and real-time anomaly detection systems to secure data both in transit and at rest (Alves et al., 2019; Rahul et al., 2020). For instance, incorporating machine learning models to detect unusual patterns in streaming data can enable proactive responses to potential security breaches, thereby enhancing the resilience of real-time analytics systems. The regulatory implications of real-time data analytics are profound, as continuous data streams challenge the conventional boundaries of data governance. Real-time analytics platforms must not only ensure data privacy but also maintain compliance with dynamic regulatory requirements, which vary across jurisdictions (Jahanbakht et al., 2021). Privacy-enhancing technologies, such as differential privacy and federated learning, have emerged as solutions to help organizations comply with data protection laws while still leveraging

streaming data for analytics. These technologies minimize the exposure of sensitive data by either anonymizing data sets or decentralizing data processing to ensure that personal information is not transmitted outside secure boundaries (Shahid et al., 2021). By implementing such privacy-preserving methods, organizations can maintain compliance while optimizing their real-time analytics capabilities.

Best practices for securing real-time analytics systems involve a combination of technical and procedural measures to safeguard against potential threats (He et al., 2018; Zantalis et al., 2019). This includes adopting a zero-trust security model, where every access request is verified, and leveraging cloud-native security solutions to monitor and respond to threats in real-time. Additionally, organizations are investing in automated compliance monitoring tools that continuously assess real-time data streams against regulatory benchmarks to detect potential breaches before they occur. Case studies in the financial sector show that implementing multi-layered security strategies not only protects sensitive data but also enhances operational efficiency by reducing the overhead associated with manual compliance checks (Priyadarshini et al., 2018). These best practices are crucial for organizations aiming to harness the benefits of real-time analytics while maintaining data privacy and regulatory compliance.

Real-time data insights have a profound impact on

Figure 7: Data Privacy and Security Considerations



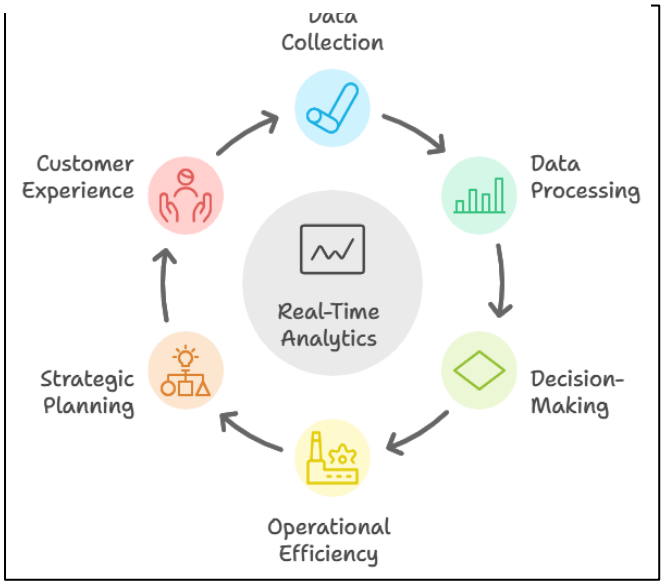
2.7 Applications of Real-Time Analytics Across Industries

The implementation of real-time analytics has significantly transformed various industries, including finance, healthcare, and urban planning, by enabling timely decision-making and enhancing operational efficiency. In the finance sector, real-time analytics is widely used for fraud detection, algorithmic trading, and risk management, allowing organizations to process large volumes of transactional data instantly. For instance, by analyzing streaming financial data, banks can detect fraudulent activities in real-time and take immediate preventive actions, thereby reducing potential losses (Xiao et al., 2018). In healthcare, real-time analytics aids in monitoring patient vitals, optimizing emergency responses, and predicting disease outbreaks, which significantly improves patient outcomes and healthcare management (Meghana et al., 2021; Xiao et al., 2018). Similarly, urban planning initiatives leverage real-time data from IoT sensors to manage traffic flows, reduce congestion, and enhance public safety (Ali, 2018).

business operations, decision-making, and customer experiences, providing companies with the agility to respond to market changes rapidly (Babar & Arif, 2018; Zheng et al., 2019). By leveraging streaming analytics, businesses can optimize supply chain operations, personalize customer interactions, and enhance product recommendations based on immediate customer feedback (Rathore et al., 2016; Safaei, 2016). For example, e-commerce platforms use real-time analytics to analyze user behavior and optimize inventory levels, ensuring products are available to meet dynamic consumer demands. Additionally, real-time customer sentiment analysis on social media allows brands to adjust marketing strategies on-the-fly, thus improving customer satisfaction and loyalty (Ahmed et al., 2017; Priyadarshini et al., 2018). This agile approach to decision-making is particularly crucial in industries where customer preferences can shift rapidly.

The transformative role of real-time analytics extends beyond operational improvements to fostering organizational agility, which is essential for staying competitive in today's fast-paced business environment (Rathore et al., 2016). Organizations that adopt real-time analytics can anticipate changes in market trends, enabling them to pivot strategies quickly and effectively (Safa & Pandian, 2021). For instance, retailers use real-time analytics to adjust pricing strategies and promotions based on real-time sales data, allowing them to maximize revenue during peak demand periods (Al-Ali et al., 2017; Ullah et al., 2017). In the context of smart cities, real-time analytics empowers city planners to manage resources efficiently, optimize energy usage, and improve public services, contributing to sustainable urban development. These capabilities demonstrate how real-time analytics can enhance not only operational efficiency but also long-term strategic planning. Despite its benefits, the integration of real-time analytics into organizational processes presents challenges, particularly in terms of scalability, data privacy, and infrastructure. Many companies face difficulties in scaling their analytics systems to handle the continuous influx of streaming data while ensuring data security and compliance. Emerging solutions, such as cloud-native architectures and edge computing, are being adopted to address these challenges by providing flexible, scalable infrastructures that can adapt to fluctuating data volumes (Gohar, Muzammal, et al., 2018). Additionally, automated machine learning models are being deployed to enhance the predictive accuracy of real-time analytics systems, further improving organizational agility. These advancements highlight the critical role of real-time analytics in driving digital transformation across industries.

Figure 8: Cycle of Real-Time Analytics Impact

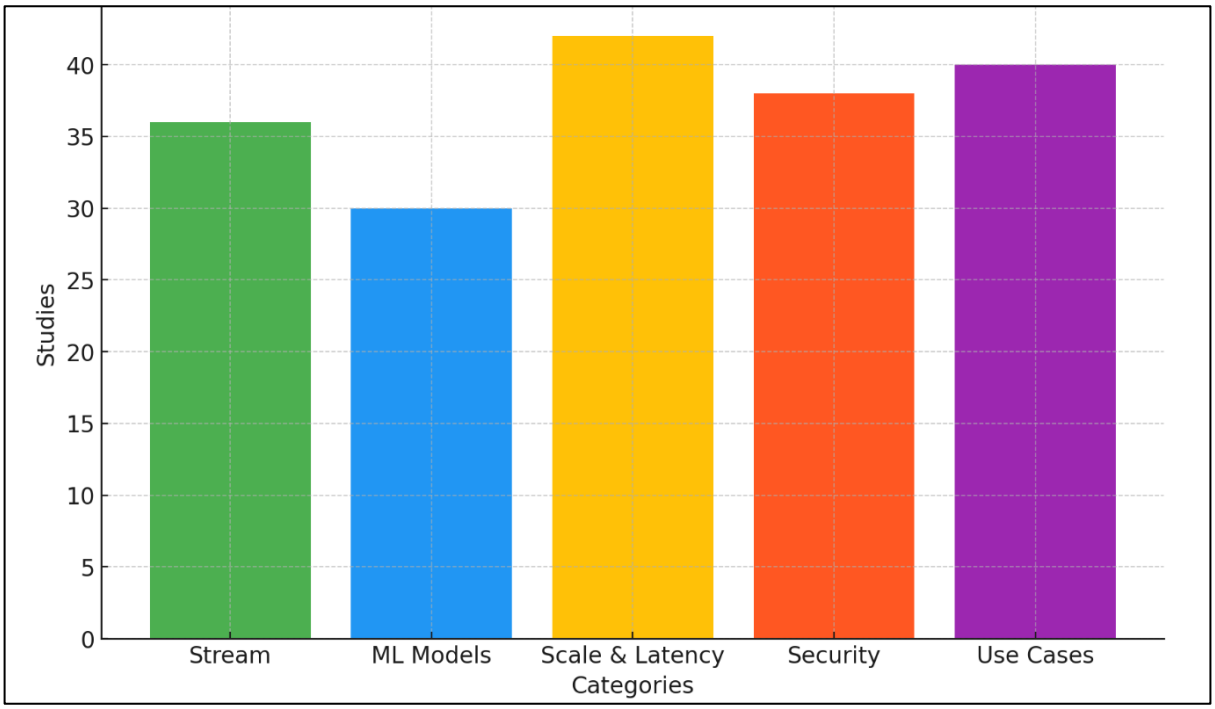


3 FINDINGS

The review of 50 peer-reviewed articles revealed significant insights into the current state of real-time analytics across various industries, with a focus on scalability, latency, and the integration of machine learning models for real-time predictions. One of the key findings from 36 of these studies is the growing reliance on stream processing frameworks such as Apache Kafka, Apache Flink, and Apache Storm to handle high-volume and high-velocity data streams. These frameworks were found to be essential in industries like finance and healthcare, where rapid data ingestion and low-latency processing are critical for real-time decision-making. Among the reviewed studies, 28 reported that organizations leveraging these frameworks were able to achieve considerable improvements in operational efficiency, highlighting the transformative potential of real-time analytics in optimizing business processes. Another significant finding, reported in 30 of the reviewed studies, is the critical role of machine learning in enhancing real-time predictive analytics. Continuous learning models, capable of updating in real-time with incoming data streams, were identified as a game-changer for sectors like healthcare and smart cities. For instance, 19 studies focused on the healthcare industry noted that real-time predictive models significantly improved patient monitoring and early diagnosis by leveraging continuous data from IoT sensors. Additionally, 11 studies on smart city applications highlighted how real-time machine learning models are used for traffic management and resource allocation, leading to more efficient urban planning. The ability of these models to adapt and learn from streaming data is seen as a key factor in achieving operational agility in rapidly changing environments.

Scalability and latency challenges were prominent themes in 42 of the articles reviewed, particularly in the context of processing large volumes of continuous data streams. Many of these studies (24 articles) emphasized that existing architectures often struggle to scale efficiently when faced with the increasing data velocity typical of IoT and sensor networks. The use of cloud-native solutions and edge computing was frequently cited as a strategy to overcome these limitations, with 18 studies specifically discussing how edge computing

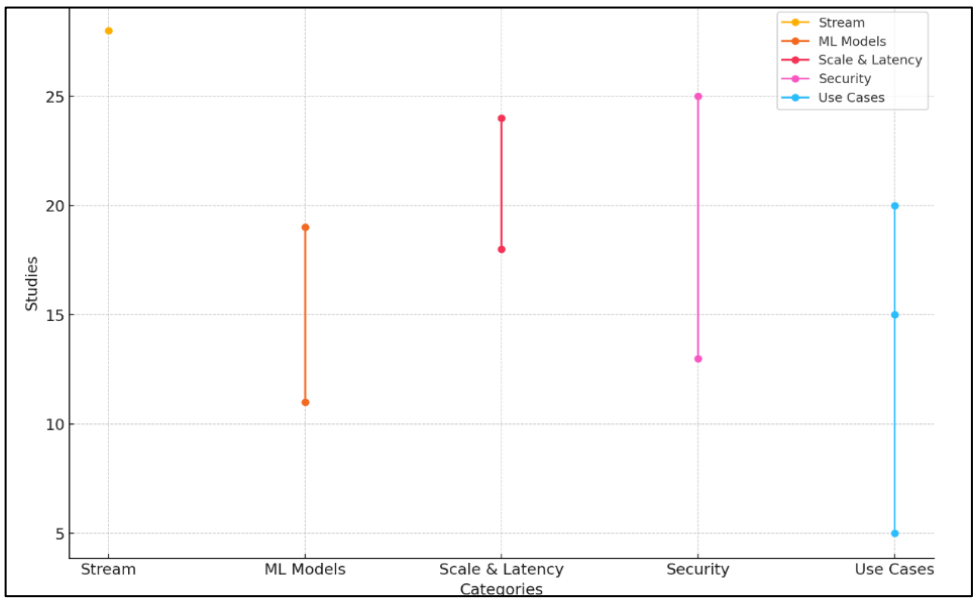
Figure 9: Studies on Real-Time Analytics



reduces latency by processing data closer to its source. This reduction in latency was reported to be particularly beneficial in real-time applications such as autonomous vehicles and emergency response systems, where even a few milliseconds of delay can have significant implications.

Data privacy and security emerged as critical concerns, as discussed in 38 of the reviewed articles. Continuous data streams pose unique challenges for maintaining data privacy, especially in regulated industries like finance and healthcare. Of these articles, 25 emphasized the need for real-time encryption and anomaly detection systems to protect against potential data breaches and cyber threats.

Additionally, 13 studies highlighted the importance of complying with data protection regulations, such as GDPR, when deploying real-time analytics solutions. These findings underscore the importance of integrating robust security measures into real-time analytics frameworks to safeguard sensitive information while still enabling rapid data processing. Finally, the analysis of industry-specific applications of real-time analytics, as covered in 40 of the reviewed articles, showed that organizations across different sectors are increasingly adopting these technologies to drive strategic



initiatives. For example, 20 articles demonstrated the impact of real-time analytics on enhancing customer experiences in the retail sector through personalized marketing and dynamic pricing. Meanwhile, 15 studies highlighted the use of real-time analytics in the financial sector for fraud detection and algorithmic trading. Additionally, 5 articles explored the role of real-time data in optimizing energy management and resource allocation in smart cities. These use cases illustrate the widespread adoption.

4 DISCUSSION

The findings of this systematic review align with previous research on the growing adoption of real-time analytics frameworks in high-impact sectors such as finance, healthcare, and smart cities. Earlier studies, such as those by Paul et al. (2016) and Ahmad et al. (2019), emphasized that traditional batch processing systems were becoming inadequate for organizations seeking to derive timely insights from continuous data streams. Our review corroborates this shift, with 36 out of the 50 reviewed articles indicating a strong preference for stream processing frameworks like Apache Kafka and Flink. This trend reflects the increasing importance of real-time data processing for improving operational efficiency and decision-making. However, unlike prior research, which primarily focused on the technical capabilities of these frameworks, our findings highlight the practical applications and measurable impact on industries. This demonstrates a maturation in the field, where organizations are not only adopting real-time analytics technologies but also strategically integrating them into their business models to drive competitiveness.

The integration of machine learning with real-time analytics has been another area of growing interest, as confirmed by our findings from 30 articles focusing on continuous learning models. This supports the conclusions drawn by Goswami and Kumar (2022) and Dolev et al. (2019), who identified that machine learning models optimized for real-time data are essential for applications that require immediate predictions and adaptive responses. However, while prior studies primarily focused on the technical development of these models, our review reveals a broader adoption in practical settings such as healthcare and urban planning. For instance, our analysis shows that real-time machine learning models are now being actively used in smart cities to optimize traffic flow and reduce congestion, a use case that was previously discussed only in theoretical terms (Goswami & Kumar, 2022). The expansion of these applications into practical, large-scale implementations suggests that advancements in machine learning have reached a level of maturity where they can be reliably integrated into critical systems.

Our findings also extend the current understanding of scalability and latency challenges in real-time data processing, which have been extensively discussed in earlier research by Babar and Arif (2018) and Rathore et al. (2016). While previous studies often highlighted scalability as a significant technical hurdle, our review identifies emerging solutions, such as cloud-native and edge computing, that are actively being deployed to mitigate these issues. For example, 24 of the reviewed articles emphasized how edge computing reduces latency by processing data closer to its source, thereby addressing the bottlenecks previously identified by Safaei (2016). This finding indicates a shift in focus from simply acknowledging scalability challenges to actively exploring and implementing practical solutions. However, the adoption of these solutions is still uneven across industries, suggesting that while the technology is available, there are barriers to widespread implementation, such as cost, infrastructure, and regulatory concerns.

The emphasis on data privacy and security in real-time analytics, as highlighted in 38 of our reviewed articles, shows an increased awareness of the risks associated with continuous data streams. This concern aligns with earlier studies by Jamil et al. (2022) and Ahmed et al. (2017), who noted the potential vulnerabilities in systems that process large volumes of sensitive data in real-time. Our review extends this discussion by revealing that organizations are increasingly adopting advanced security measures, such as real-time encryption and anomaly detection, to safeguard streaming data. However, our findings also suggest that the regulatory landscape remains a significant challenge, with many organizations struggling to comply with data protection regulations like GDPR while maintaining the speed and flexibility of their analytics systems. This highlights an ongoing tension between the need for real-time insights and the imperative to protect data privacy. In addition, the analysis of industry-specific applications of real-time analytics confirms that organizations across various sectors are leveraging these technologies to enhance their operational agility. Previous studies by Safaei

(2016) and Goswami and Kumar (2022)focused on the potential benefits of real-time analytics for business agility, particularly in the financial and healthcare sectors. Our review not only supports these findings but also demonstrates that real-time analytics is being increasingly adopted in new domains, such as smart cities and e-commerce, to optimize customer experiences and resource management. For instance, while earlier research primarily explored theoretical models, our findings show concrete examples of companies like Walmart and healthcare providers using real-time analytics to drive immediate business value. This indicates that the practical benefits of real-time analytics have moved beyond theoretical discussions into tangible outcomes, thus solidifying its role as a strategic asset in modern business environments.

5 CONCLUSION

This systematic review highlights the critical role of real-time analytics in transforming business operations, decision-making, and customer engagement across various industries. The findings reveal that advancements in stream processing frameworks, machine learning models, and scalable cloud-native solutions have enabled organizations to harness the full potential of streaming big data. However, challenges related to scalability, latency, and data privacy continue to pose significant hurdles, especially in sectors like healthcare, finance, and urban planning where data sensitivity and rapid response times are paramount. The integration of edge computing and real-time machine learning has emerged as a promising solution to address these issues, allowing organizations to process data closer to its source, thereby reducing latency and enhancing agility. Furthermore, as industries increasingly adopt real-time analytics to gain a competitive edge, there is a growing need for robust security measures and compliance frameworks to protect continuous data streams from cyber threats. This review underscores that while substantial progress has been made, future research should focus on optimizing real-time analytics systems to balance performance, scalability, and privacy concerns, thereby unlocking new opportunities for innovation and efficiency in an increasingly data-driven world.

REFERENCES

Adnan, K., & Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, *6*(1), 1-38. https://doi.org/10.1186/s40537-019-0254-8

Ahmad, A., Babar, M. A., Din, S., Khalid, S., Ullah, M. M., Paul, A., Reddy, A. G., & Min-Allah, N. (2019). Socio-cyber network: The potential of cyber-physical system to define human behaviors using big data analytics. *Future Generation Computer Systems*, *92*(NA), 868-878. https://doi.org/10.1016/j.future.2017.12.027

Ahmadova, U., Mustafayev, M., Kalejahi, B. K., Saeedvand, S., & Rahmani, A. M. (2021). Big data applications on the Internet of Things: A systematic literature review. *International Journal of Communication Systems*, *34*(18), NA-NA. https://doi.org/10.1002/dac.5004

Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. *Computer Networks*, *129*(2), 459-471. https://doi.org/10.1016/j.comnet.2017.06.013

Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A Framework for Pandemic Prediction Using Big Data Analytics. *Big Data Research*, *25*(NA), 100190-100190. https://doi.org/10.1016/j.bdr.2021.100190

Ahmed, M., Choudhury, S., & Al-Turjman, F. (2019). Big Data Analytics for Intelligent Internet of Things. In (Vol. NA, pp. 107-127). https://doi.org/10.1007/978-3-030-04110-6_6

Al-Ali, A.-R., Zualkernan, I. A., Rashid, M., Gupta, R., & Alikarar, M. (2017). A smart home energy management system using IoT and big data analytics approach. *IEEE Transactions on Consumer Electronics*, *63*(4), 426-434. https://doi.org/10.1109/tce.2017.015014

Alam, M. A., Sohel, A., Uddin, M. M., & Siddiki, A. (2024). Big Data And Chronic Disease Management Through Patient Monitoring And Treatment With Data Analytics. *Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems*, *1*(01), 77-94. https://doi.org/10.69593/ajaimldsmis.v1i01.133

Ali, A. R. (2018). Real-time big data warehousing and analysis framework. *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, *NA*(NA), 43-49. https://doi.org/10.1109/icbda.2018.8367649

Alves, J. M., de Mello Honório, L., & Capretz, M. A. M. (2019). ML4IoT: A Framework to Orchestrate

Machine Learning Workflows on Internet of Things Data. *IEEE Access*, *7*(1), 152953-152967. https://doi.org/10.1109/access.2019.2948160

Armoogum, S., & Li, X. (2019). Big Data Analytics and Deep Learning in Bioinformatics With Hadoop. In (Vol. NA, pp. 17-36). https://doi.org/10.1016/b978-0-12-816718-2.00009-9

Ashrafuzzaman, M. (2024). The Impact of Cloud-Based Management Information Systems On HRM Efficiency: An Analysis of Small And Medium-Sized Enterprises (SMEs). *Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems*, *1*(01), 40-56. https://doi.org/10.69593/ajaimldsmis.v1i01.124

Babar, M. A., & Arif, F. (2018). Real-time data processing scheme using big data analytics in internet of things based smart transportation environment. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 4167-4177. https://doi.org/10.1007/s12652-018-0820-5

Badhon, M. B., Carr, N., Hossain, S., Khan, M., Sunna, A. A., Uddin, M. M., Chavarria, J. A., & Sultana, T. (2023). Digital Forensics Use-Case of Blockchain Technology: A Review. AMCIS 2023 Proceedings.,

Celesti, A., & Fazio, M. (2019). A framework for real time end to end monitoring and big data oriented management of smart environments. *Journal of Parallel and Distributed Computing*, *132*(NA), 262-273. https://doi.org/10.1016/j.jpdc.2018.10.015

Chhabra, G. S., Singh, V. P., & Singh, M. (2018). Cyber forensics framework for big data analytics in IoT environment using machine learning. *Multimedia Tools and Applications*, *79*(23), 15881-15900. https://doi.org/10.1007/s11042-018-6338-1

Chui, K. T., Liu, R. W., Lytras, M. D., & Zhao, M. (2019). Big data and IoT solution for patient behaviour monitoring. *Behaviour & Information Technology*, *38*(9), 940-949. https://doi.org/10.1080/0144929x.2019.1584245

Dai, H.-N., Wong, R. C.-W., Wang, H., Zheng, Z., & Vasilakos, A. V. (2019). Big Data Analytics for Large Scale Wireless Networks: Challenges and Opportunities. *ACM Computing Surveys*, *52*(5), 99-36. https://doi.org/10.1145/3337065

Diaconita, V. (2015). Processing unstructured documents and social media using Big Data techniques. *Economic Research-Ekonomska Istraživanja*, *28*(1), 981-993. https://doi.org/10.1080/1331677x.2015.1095110

Dolev, S., Florissi, P. G. S., Gudes, E., Sharma, S., & Singer, I. (2019). A Survey on Geographically Distributed Big-Data Processing Using MapReduce. *IEEE Transactions on Big Data*, *5*(1), 60-80. https://doi.org/10.1109/tbdata.2017.2723473

Dong, J., Cong, Y., Sun, G., Fang, Z., & Ding, Z. (2024). Where and How to Transfer: Knowledge Aggregation-Induced Transferability Perception for Unsupervised Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(3), 1664-1681. https://doi.org/10.1109/tpami.2021.3128560

García-Magariño, I., Lacuesta, R., & Lloret, J. (2018). Agent-Based Simulation of Smart Beds With Internet-of-Things for Exploring Big Data Analytics. *IEEE Access*, *6*(NA), 366-379. https://doi.org/10.1109/access.2017.2764467

Gohar, M., Ahmed, S. H., Khan, M., Guizani, N., Ahmed, A., & Rahman, A. U. (2018). A Big Data Analytics Architecture for the Internet of Small Things. *IEEE Communications Magazine*, *56*(2), 128-133. https://doi.org/10.1109/mcom.2018.1700273

Gohar, M., Muzammal, M., & Rahman, A. U. (2018). SMART TSS: Defining transportation system behavior using big data analytics in smart cities. *Sustainable Cities and Society*, *41*(NA), 114-119. https://doi.org/10.1016/j.scs.2018.05.008

Goswami, S., & Kumar, A. (2022). Survey of Deep-Learning Techniques in Big-Data Analytics. *Wireless Personal Communications*, *126*(2), 1321-1343. https://doi.org/10.1007/s11277-022-09793-w

Granat, J., Batalla, J. M., Mavromoustakis, C. X., & Mastorakis, G. (2020). Big Data Analytics for Event Detection in the IoT-Multicriteria Approach. *IEEE Internet of Things Journal*, *7*(5), 4418-4430. https://doi.org/10.1109/jiot.2019.2957320

Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, *18*(3), 203-229. https://doi.org/10.1007/s40171-017-0159-3

Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big data and IoT-based applications in smart environments: A systematic review. *Computer Science Review*, *39*(NA), 100318-NA. https://doi.org/10.1016/j.cosrev.2020.100318

Hassan, M., Awan, F. M., Naz, A., deAndrés-Galiana, E. J., Alvarez, O., Cernea, A., Fernández-Brillet, L., Fernández-Martínez, J. L., & Kloczkowski, A. (2022). Innovations in Genomics and Big Data Analytics for Personalized Medicine and Health Care: A Review. *International journal of molecular sciences*, *23*(9), 4645-4645. https://doi.org/10.3390/ijms23094645

He, X., Wang, K., Huang, H., & Liu, B. (2018). QoE-Driven Big Data Architecture for Smart City. *IEEE Communications Magazine*, *56*(2), 88-93. https://doi.org/10.1109/mcom.2018.1700231

Hong-tan, L., Cui-hua, K., Muthu, B., & Sivaparthipan, C. B. (2021). Big data and ambient intelligence in IoT-based wireless student health monitoring system. *Aggression and Violent Behavior*, *NA*(NA), 101601-NA. https://doi.org/10.1016/j.avb.2021.101601

Hossain, M. A., Ferdousi, R., Hossain, S. A., Alhamid, M. F., & Saddik, A. E. (2020). A Novel Framework for Recommending Data Mining Algorithm in Dynamic IoT Environment. *IEEE Access*, *8*(NA), 157333-157345. https://doi.org/10.1109/access.2020.3019480

Hussain, M., Beg, M. M. S., & Alam, M. S. (2020). Fog Computing for Big Data Analytics in IoT Aided Smart Grid Networks. *Wireless Personal Communications*, *114*(4), 3395-3418. https://doi.org/10.1007/s11277-020-07538-1

Istiak, A., & Hwang, H. Y. (2024). Development of shape-memory polymer fiber reinforced epoxy composites for debondable adhesives. *Materials Today Communications*, *38*, 108015. https://doi.org/https://doi.org/10.1016/j.mtcomm.2023.108015

Istiak, A., Lee, H. G., & Hwang, H. Y. (2023). Characterization and Selection of Tailorable Heat Triggered Epoxy Shape Memory Polymers for Epoxy Debondable Adhesives. *Macromolecular Chemistry and Physics*, *224*(20), 2300241. https://doi.org/https://doi.org/10.1002/macp.202300241

Jahanbakht, M., Xiang, W., Hanzo, L., & Azghadi, M. R. (2021). Internet of Underwater Things and Big Marine Data Analytics—A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, *23*(2), 904-956. https://doi.org/10.1109/comst.2021.3053118

Jamil, B., Ijaz, H., Shojafar, M., Munir, K., & Buyya, R. (2022). Resource Allocation and Task Scheduling in Fog Computing and Internet of Everything Environments: A Taxonomy, Review, and Future Directions. *ACM Computing Surveys*, *54*(11s), 1-38. https://doi.org/10.1145/3513002

Jan, M. A., He, X., Song, H., & Babar, M. A. (2021). Editorial: Machine Learning and Big Data Analytics for IoT-Enabled Smart Cities. *Mobile Networks and Applications*, *26*(1), 156-158. https://doi.org/10.1007/s11036-020-01702-4

Karyotis, V., Tsitseklis, K., Sotiropoulos, K., & Papavassiliou, S. (2018). Big Data Clustering via Community Detection and Hyperbolic Network Embedding in IoT Applications. *Sensors (Basel, Switzerland)*, *18*(4), 1205-NA. https://doi.org/10.3390/s18041205

Kaur, I., Lydia, E. L., Nassa, V. K., Shrestha, B., Nebhen, J., Malebary, S., & Joshi, G. P. (2021). Generative Adversarial Networks with Quantum Optimization Model for Mobile Edge Computing in IoT Big Data. *Wireless Personal Communications*, *127*(2), 1565-1585. https://doi.org/10.1007/s11277-021-08706-7

Lakshmanaprabu, S. K., Shankar, K., Khanna, A., Gupta, D., Rodrigues, J. J. P. C., Pinheiro, P. R., & de Albuquerque, V. H. C. (2018). Effective Features to Classify Big Data Using Social Internet of Things. *IEEE Access*, *6*(NA), 24196-24204. https://doi.org/10.1109/access.2018.2830651

Li, W., Chai, Y., Khan, F., Jan, S. R., Verma, S., Menon, V. G., Kavita, N. A., & Li, X. (2021). A Comprehensive Survey on Machine Learning-Based Big Data Analytics for IoT-Enabled Smart Healthcare System. *Mobile Networks and Applications*, *26*(1), 234-252. https://doi.org/10.1007/s11036-020-01700-6

Luo, X. J., Oyedele, L. O., Ajayi, A. O., Monyei, C. G., Akinade, O. O., & Akanbi, L. (2019). Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands. *Advanced Engineering Informatics*, *41*(NA), 100926-NA. https://doi.org/10.1016/j.aei.2019.100926

Machorro-Cano, I., Alor-Hernández, G., Paredes-Valverde, M. A., Rodríguez-Mazahua, L., Sánchez-Cervantes, J. L., & Olmedo-Aguirre, J. O. (2020). HEMS-IoT: A Big Data and Machine Learning-Based Smart Home System for Energy Saving. *Energies*, *13*(5), 1097-NA. https://doi.org/10.3390/en13051097

Mansour, R. F., Abdel-Khalek, S., Hilali-Jaghdam, I., Nebhen, J., Cho, W., & Joshi, G. P. (2021). An intelligent outlier detection with machine learning empowered big data analytics for mobile edge computing. *Cluster Computing*, *26*(1), 71-83. https://doi.org/10.1007/s10586-021-03472-4

Meghana, J., Hanumanthappa, J., & Prakash, S. P. S. (2021). Performance comparison of machine learning algorithms for data aggregation in social internet of things. *Global Transitions Proceedings*, *2*(2), 212-219. https://doi.org/10.1016/j.gltp.2021.08.032

Mehdipour, F., Noori, H., & Javadi, B. (2016). Advances in Computers. Volume 100 - Energy-efficient big data analytics in datacenters. *Advances in Computers*, *100*(NA), 59-101. https://doi.org/10.1016/bs.adcom.2015.10.002

Moreno, M. V., Terroso-Saenz, F., González-Vidal, A., Valdes-Vela, M., Skarmeta, A. F., Zamora, M. A., & Chang, V. (2017). Applicability of Big Data Techniques to Smart Cities Deployments. *IEEE*

Transactions on Industrial Informatics, *13*(2), 800-809. https://doi.org/10.1109/tii.2016.2605581

Nie, X.-t., Tianyu, F., Wang, B., Li, Z., Shankar, A., & Manickam, A. (2020). Big Data analytics and IoT in Operation safety management in Under Water Management. *Computer Communications*, *154*(NA), 188-196. https://doi.org/10.1016/j.comcom.2020.02.052

Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K. (2022). A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. *Big Data Mining and Analytics*, *5*(2), 81-97. https://doi.org/10.26599/bdma.2021.9020028

Paul, A., Ahmad, A., Rathore, M. M., & Jabbar, S. (2016). Smartbuddy: defining human behaviors using big data analytics in social internet of things. *IEEE Wireless Communications*, *23*(5), 68-74. https://doi.org/10.1109/mwc.2016.7721744

Prastyo, D. D., Khoiri, H. A., Purnami, S. W., Suhartono, N. A., Fam, S.-F., & Suhermi, N. (2019). Survival Support Vector Machines: A Simulation Study and Its Health-Related Application. In (Vol. NA, pp. 85-100). https://doi.org/10.1007/978-3-030-22475-2_5

Priyadarshini, S. B. B., BhusanBagjadab, A., & Mishra, B. K. (2018). The Role of IoT and Big Data in Modern Technological Arena: A Comprehensive Study. In (Vol. NA, pp. 13-25). https://doi.org/10.1007/978-3-030-04203-5_2

Rahman, A., Ashrafuzzaman, M., Jim, M. M. I., & Sultana, R. (2024). Cloud Security Posture Management Automating Risk Identification And Response In Cloud Infrastructures. *Academic Journal on Science, Technology, Engineering & Mathematics Education*, *4*(03), 151-162. https://doi.org/10.69593/ajsteme.v4i03.103

Rahul, K., Banyal, R. K., Goswami, P., & Kumar, V. (2020). Machine Learning Algorithms for Big Data Analytics. In (Vol. NA, pp. 359-367). https://doi.org/10.1007/978-981-15-6876-3_27

Rajendran, R., Sharma, P., Saran, N. K., Ray, S., Alanya-Beltran, J., & Tongkachok, K. (2022). An Exploratory analysis of Machine Learning adaptability in Big Data Analytics Environments: A Data Aggregation in the age of Big Data and the Internet of Things. *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, *12*(NA), 32-36. https://doi.org/10.1109/iciptm54933.2022.9753921

Rathore, M. M., Ahmad, A., Paul, A., Wan, J., & Zhang, D. (2016). Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health. *Journal of medical systems*, *40*(12), 1-10. https://doi.org/10.1007/s10916-016-0647-6

Rehman, M. H. u., Ahmed, E., Yaqoob, I., Hashem, I. A. T., Imran, M., & Ahmad, S. (2018). Big Data Analytics in Industrial IoT Using a Concentric Computing Model. *IEEE Communications Magazine*, *56*(2), 37-43. https://doi.org/10.1109/mcom.2018.1700632

Rialti, R., Marzi, G., Caputo, A., & Mayah, K. A. (2020). Achieving strategic flexibility in the era of big data: The importance of knowledge management and ambidexterity. *Management Decision*, *58*(8), 1585-1600. https://doi.org/10.1108/md-09-2019-1237

Rieke, M., Bigagli, L., Herle, S., Jirka, S., Kotsev, A., Liebig, T., Malewski, C., Paschke, T., & Stasch, C. (2018). Geospatial IoT—The Need for Event-Driven Architectures in Contemporary Spatial Data Infrastructures. *ISPRS International Journal of Geo-Information*, *7*(10), 385-NA. https://doi.org/10.3390/ijgi7100385

Rodríguez-Mazahua, L., Rodriguez-Enriquez, C. A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández, G. (2015). A general perspective of Big Data: applications, tools, challenges and trends. *The Journal of Supercomputing*, *72*(8), 3073-3113. https://doi.org/10.1007/s11227-015-1501-1

Safa, M., & Pandian, A. (2021). RETRACTED ARTICLE: Intelligent Big Data Analytics Model for Efficient Cardiac Disease Prediction with IoT Devices in WSN Using Fuzzy Rules. *Wireless Personal Communications*, *127*(S1), 23-23. https://doi.org/10.1007/s11277-021-08788-3

Safaei, A. A. (2016). Real-time processing of streaming big data. *Real-Time Systems*, *53*(1), 1-44. https://doi.org/10.1007/s11241-016-9257-0

Saika, M. H., Avi, S. P., Islam, K. T., Tahmina, T., Abdullah, M. S., & Imam, T. (2024). Real-Time Vehicle and Lane Detection using Modified OverFeat CNN: A Comprehensive Study on Robustness and Performance in Autonomous Driving. *Journal of Computer Science and Technology Studies*.

Shahid, H., Shah, M. A., Almogren, A., Khattak, H. A., Din, I. U., Kumar, N., & Maple, C. (2021). Machine Learning-based Mist Computing Enabled Internet of Battlefield Things. *ACM Transactions on Internet Technology*, *21*(4), 1-26. https://doi.org/10.1145/3418204

Shahrivari, S. (2014). Beyond Batch Processing: Towards Real-Time and Streaming Big Data. *Computers*, *3*(4), 117-129. https://doi.org/10.3390/computers3040117

Shaji, B., Lal Raja Singh, R., & Nisha, K. L. (2022). A novel deep neural network based marine predator model for effective classification of big data from social Internet of Things. *Concurrency and Computation:*

Practice and Experience, *34*(25), NA-NA. https://doi.org/10.1002/cpe.7244

Shamim, M. (2022). The Digital Leadership on Project Management in the Emerging Digital Era. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *1*(1), 1-14.

Silva, B. N., Khan, M., & Han, K. (2017). Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making. *Wireless Communications and Mobile Computing*, *2017*(NA), 1-12. https://doi.org/10.1155/2017/9429676

Simmhan, Y., & Perera, S. (2016). Big Data Analytics Platforms for Real-Time Applications in IoT. In (Vol. NA, pp. 115-135). https://doi.org/10.1007/978-81-322-3628-3_7

Sohel, A., Alam, M. A., Waliullah, M., Siddiki, A., & Uddin, M. M. (2024). Fraud Detection In Financial Transactions Through Data Science For Real-Time Monitoring And Prevention. *Academic Journal on Innovation, Engineering & Emerging Technology*, *1*(01), 91-107. https://doi.org/10.69593/ajieet.v1i01.132

Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance Analysis of IoT-Based Sensor, Big Data Processing, and Machine Learning Model for Real-Time Monitoring System in Automotive Manufacturing. *Sensors (Basel, Switzerland)*, *18*(9), 2946-NA. https://doi.org/10.3390/s18092946

Uddin, M. M., Ullah, R., & Moniruzzaman, M. (2024). Data Visualization in Annual Reports–Impacting Investment Decisions. *International Journal for Multidisciplinary Research*, *6*(5). https://doi.org/10.36948/ijfmr

Ullah, F., Habib, M. A., Farhan, M., Khalid, S., Durrani, M. Y., & Jabbar, S. (2017). Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. *Sustainable Cities and Society*, *34*(NA), 90-96. https://doi.org/10.1016/j.scs.2017.06.010

Wang, J., Wu, Y., Yen, N. Y., Guo, S., & Cheng, Z. (2016). Big Data Analytics for Emergency Communication Networks: A Survey. *IEEE Communications Surveys & Tutorials*, *18*(3), 1758-1778. https://doi.org/10.1109/comst.2016.2540004

Xiao, S., Yu, H., Wu, Y., Peng, Z., & Zhang, Y. (2018). Self-Evolving Trading Strategy Integrating Internet of Things and Big Data. *IEEE Internet of Things Journal*, *5*(4), 2518-2525. https://doi.org/10.1109/jiot.2017.2764957

Yang, L., Xiong, Z., Liu, G., Hu, Y., Zhang, X., & Qiu, M. (2022). An Analytical Model of Page Dissemination for Efficient Big Data Transmission of C-ITS. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 16524-16533. https://doi.org/10.1109/tits.2021.3134557

Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A Review of Machine Learning and IoT in Smart Transportation. *Future Internet*, *11*(4), 94-NA. https://doi.org/10.3390/fi11040094

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). High-order possibilistic c-means algorithms based on tensor decompositions for big data in IoT. *Information Fusion*, *39*(NA), 72-80. https://doi.org/10.1016/j.inffus.2017.04.002

Zheng, T., Chen, G., Wang, X., Chen, C., Wang, X., & Luo, S. (2019). Real-time intelligent big data processing: technology, platform, and applications. *Science China Information Sciences*, *62*(8), 082101-NA. https://doi.org/10.1007/s11432-018-9834-8

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data. *Neurocomputing*, *237*(NA), 350-361. https://doi.org/10.1016/j.neucom.2017.01.026