

Navigating Healthcare Challenges Text Analytics, Data Integration, and Decision-Making in the COVID-19 Era

Shujana Madnira

School of Information Technology, Mapúa Institute of Technology, PHILIPPINES

e-mail : guyanaa931@gmail.com

Corresponding Autor: Shujana Madnira

Abstract

In the context of the COVID-19 pandemic, Integrated Healthcare Systems have emerged as crucial components in effectively managing healthcare challenges. This study delves into the multifaceted role of integrated systems, with a particular focus on the pivotal aspects of text analytics. An exploration of various applications of text analytics unfolds, shedding light on its diverse utility within the healthcare landscape. Extensive reviews of problems encountered by different organizations and insights gleaned from research contribute to a comprehensive understanding of the challenges faced by Health and Human Services (HHS). These challenges, intricately linked to issues such as hospital strains and consumers' personal experiences, are thoroughly examined to provide actionable solutions. A key emphasis is placed on the indispensability of data integration, and the abstract discusses how various analytic approaches can be strategically employed within a well-integrated database system. The nuances of implementing an integrated model are scrutinized, highlighting the primary challenges that organizations, particularly HHS, may encounter. Subsequently, potential solutions are presented, leveraging the power of OLAP to construct a dashboard tailored to address the identified problems. Beyond the technical intricacies, the abstract explores the ramifications of an integrated approach on decision-making processes within HHS. The discussion extends to the acceleration of decision-making possibilities, underlining the imperative need for timely and informed actions in the face of healthcare challenges. In essence, this study provides a nuanced exploration of the role of Integrated Healthcare Systems during the COVID-19 pandemic, incorporating insights from text analytics, data integration, and analytic methodologies. The findings aim to contribute valuable perspectives to healthcare organizations, particularly HHS, as they navigate and mitigate the complexities posed by the ongoing global health crisis.

Keywords— Pharmacy, Clinic, Text analytics, Data mining, OLAP, Sentiment analysis, Data integration.

1 Introduction

In the wake of the Covid-19 pandemic, the significance of healthcare has been thrust into the spotlight, particularly when the physical accessibility of necessary treatments is compromised. The need for a robust and responsive healthcare system has become more evident than ever. To ensure a consistent and high-quality healthcare delivery, the integration of advanced technological systems for comprehensive data management has become imperative [1]. In response to the ongoing crisis, healthcare organizations find themselves grappling with an unprecedented surge in data collection, surpassing the volumes witnessed in pre-pandemic times. The sheer magnitude of data being generated necessitates the implementation of well-integrated systems capable of efficiently handling and storing this wealth of information [2], [3]. The establishment of such integrated frameworks is paramount for ensuring the seamless storage, retrieval, and management of data crucial for analyses. An integral aspect of this technological evolution is its direct impact on patient care. The effective storage and management of data play a pivotal role in facilitating prompt and precise treatment strategies. In essence, a well-integrated data

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

infrastructure becomes the linchpin in delivering healthcare solutions that are not only efficient but also tailored to the unique needs arising from the challenges posed by the pandemic [4].

Within this landscape of evolving healthcare paradigms, the Department of Health and Human Services (HHS) emerges as a key player committed to addressing operational challenges through innovative analytical approaches. Leveraging the wealth of data housed within an integrated infrastructure, the HHS aims to navigate and solve complex issues inherent in the healthcare system [5], [6], [7]. The integration of diverse analytical methodologies positions the organization to extract actionable insights, enhancing its ability to make informed decisions that positively impact patient outcomes and organizational efficiency. As the healthcare sector continues to grapple with the ever-evolving challenges posed by the pandemic, the synergy between technological advancements, data integration, and analytical approaches championed by organizations like the HHS stands as a beacon of progress. It underscores the transformative potential of leveraging data in addressing operational inefficiencies and elevating healthcare delivery to new heights. In this era of heightened reliance on technology, the integration of data-driven solutions not only ensures the resilience of healthcare systems but also paves the way for a future where patient care is optimized through informed decision-making and advanced analytics.

1) *Analytical Perspectives.*

Due to the intricate nature of digital data sets, the journal underscores the critical importance of employing text analytics to extract, organize, and analyze pertinent medical data. This approach is essential for identifying patterns that enable the monitoring of patients' progress. Within the vast data collected by healthcare systems, a predominant portion exists in the form of unstructured textual information, including prescriptions, medical history records, and test results encapsulated in Electronic Health Records (EHR), as noted by Hyder and Nafees [8], the integration of Natural Language Processing (NLP) with machine learning techniques facilitates the extraction of insights from the free-text data within EHRs. This, in turn, enables the development of predictive analytics for patient management, pharmacovigilance, and clinical populations, as highlighted by Zhao Jorgetto [9], who also leverage web analytics for prediction, identification, and monitoring based on available information.

Schwemmer and Ziewiecki [10] emphasizes the increasing demand for social media analytic techniques in light of the substantial growth in social media and content usage. Utilizing information such as customer opinions, reviews, and product pricing from other businesses aids in understanding past events and predicting future trends. The article discusses the development of a framework for text mining, quantitative analysis, and sentiment analysis, employing machine learning to classify text responses into negative and positive categories.

Amid the current healthcare landscape, there is a pressing need for the overhaul of healthcare systems to address sustainability concerns. The potential shortage of drugs in pharmacies poses a significant risk to public health. Consequently, predictive models, employing stream analytics and machine learning algorithms, become crucial for estimating impending drug shortages and determining the most effective medicines for specific diseases [11].

Robakowska et al. [12] highlight the challenges posed by a high demand for minor ailment treatment in hospital emergency departments (ED), leading to an influx of doctor appointments and ED visits. To address this issue, geospatial analysis is employed to assess the mapping data and geographical distribution of healthcare centers, offering insights into dominant patterns in accessibility and distribution.

While data analytics technology is indispensable for extracting insights from unstructured data in the healthcare industry, it is not without its challenges. Hyder and Nafees [8] note issues with Natural Language Processing, particularly in implementing complex linguistic features, leading to less accurate clinical insights compared to annotated clinical documentation. Furthermore, Zhao Jorgetto [9] highlight the time-consuming and unreliable nature of analytics, especially in social media analytics, where integration of data from various platforms can lead to inefficiencies in analysis. These challenges underscore the need for continued advancements in analytics management to harness the full potential of data-driven insights in healthcare.

2 Problem Statement

1) *Strains in hospitals due to lack of community pharmacies*

Amid the ongoing pandemic, there is an urgent demand for healthcare services, particularly for COVID-19 treatments. This heightened demand has placed significant strains on hospitals, especially in rural areas. With medical facilities prioritizing COVID-19 patients in emergency departments (ED), individuals with minor ailments find it challenging to secure treatment in hospitals. Consequently, those with non-urgent medical conditions are compelled to seek care from local clinics and pharmacies. However, in certain states across the US, the distribution of pharmacies and clinics is insufficient, making it difficult for patients to access essential services. This scarcity

forces patients to rely on hospitals, contributing to increased COVID-19 infections and further burdening healthcare facilities. Recognizing this challenge, the Department of Health and Human Services (HHS) endeavors to address the issue through a comprehensive study.

To gather pertinent data, the HHS employs web analytics based on the traffic received on its website, originating from patients searching online for local pharmacies. The HHS website serves as a valuable resource by providing addresses of various clinics and pharmacies managed by the organization across multiple states. This web traffic, analyzed through web analytics, serves as a foundation for geospatial analytics. Geospatial analytics is utilized to identify the geographic locations of website visitors, particularly focusing on regions where patients are concentrated. This analysis offers insights into the areas where there is a higher demand for community pharmacies. By referring to existing organizational data containing details about pharmacies, including their locations, the HHS can strategically address the distribution gaps in states with high patient concentrations. Moreover, leveraging web analytics enables the implementation of stream analytics, providing real-time data on visitor locations. This live data facilitates the creation of dynamic and timely reports, allowing for the analysis of geolocation data. Through this multifaceted approach, the HHS aims to not only identify current challenges but also to proactively address emerging needs, ultimately improving the accessibility of healthcare services during the ongoing pandemic.

2) *Strains in hospitals due to lack of community pharmacies*

The surge in demand for pharmaceutical products and healthcare services during the pandemic has led to a notable shift in customer behavior compared to pre-pandemic times [13]. It is imperative to comprehend the sentiments, behaviors, and thoughts of customers regarding the use of pharmaceutical products and the acquisition of services from clinics and pharmacies in this altered landscape. This understanding is crucial for differentiating authentic personal experiences from feedback influenced by mainstream media. For instance, a recent controversy arose among consumers when negative comments emerged, undermining the benefits of vaccination. These comments were rooted in anti-vaccine propaganda disseminated by individuals lacking sufficient knowledge about the vaccines. The Department of Health and Human Services (HHS) aims to address such challenges through a strategic approach.

The HHS plans to leverage text mining techniques to extract unstructured texts from its social media pages and website. Subsequently, text analytics will be employed to transform this data into structured information containing feedback and complaints related to healthcare products and clinical services. Web analytics will further be conducted to analyze the duration visitors spend on the HHS website, providing insights into their behavior; a longer duration suggests a higher likelihood of genuine engagement. Crucially, social media analytics will play a pivotal role in understanding visitor sentiments. Employing sentiment analysis, the HHS aims to establish a sentiment ratio for individual feedback. By discerning the difference between the proportions of high and low sentiment ratios and considering the context of the feedback, the organization can effectively distinguish between genuine experiences and mere hearsay.

In summary, the HHS recognizes the importance of comprehending evolving customer sentiments and behaviors in the context of pharmaceutical products and healthcare services during the pandemic. Through the implementation of text mining, text analytics, web analytics, and social media analytics, the organization seeks to extract valuable insights, enabling a nuanced understanding of genuine experiences and addressing misinformation or misconceptions in the public discourse.

3 Text Data Retrieval Perspective

1) *Data Integration*

Data integration stands out as a highly impactful practice across scientific and commercial domains, particularly in healthcare. The meticulous control of healthcare data holds the potential to markedly enhance clinical outcomes and patient care within healthcare organizations [14]. Additionally, it facilitates the tracking of crucial work-related information for physicians, such as organizational affiliations, faculty appointments, and locations. Improved network communication within the data system not only enhances call center productivity but also enables real-time accessibility to appointment information. The judicious use of information technology is imperative to support decision-making processes for healthcare professionals and top management. Efforts have been directed towards leveraging various medical data sources for applications in clinical research and education. Amidst these advancements, hospitals find themselves grappling with an influx of patients seeking treatment for minor ailments, often stemming from limited access to local clinics or dissatisfaction with healthcare service providers.

To address this challenge, web analytics becomes indispensable in identifying specific regions where people encounter difficulties accessing clinical services. Web and text mining techniques prove invaluable in extracting unstructured text chunks, encompassing web and geospatial data related to individuals in need of healthcare services [15]. Given the dynamic nature of this data, daily updates underscore the importance of storing it in a well-integrated database for comprehensive analyses. Without a robust enterprise data infrastructure, the crucial steps involved in transforming and cleansing data are overlooked. This oversight results in the storage of redundant and inaccurate data, diminishing the business value of the information. Analyzing such unclean datasets leads to skewed results influenced by unforeseen patterns and trends, thereby compromising the decision-making process. A well-integrated data infrastructure serves as a remedy, eliminating undesired data and ensuring accurate and consistent analyses in the realms of web analytics, geospatial analysis, and stream analytics for web-related data.

2) *Text Analytics and Its Application*

Web crawling serves as a method of web mining, aiming to gather data pertaining to web content and usage. Its primary purpose is to extract meaningful information from unstructured text content within webpages, encompassing details about visitor activity, such as visits and time spent on each page. Social media analytics utilizes the unstructured text extracted through web crawling, analyzing social media content, including posts and comments, to identify entities and relationships. Natural Language Processing (NLP) is employed to discern ambiguities and sentiments, distinguishing genuine feedback from hearsay. Geospatial data is an integral component of the information extracted through web crawling, encompassing latitudes, longitudes, and states. This geospatial data is crucial for pinpointing areas lacking pharmacies and clinics through geospatial analysis within web analytics [7], [8], [11], [15]. Given the research's organization-oriented focus on identifying deficient clinics and pharmacies, a geospatial static approach is exclusively employed.

Stream analytics comes into play for live streaming of locations as feedback is provided or webpages are visited. In text analytics, specific business challenges are addressed by identifying keywords and utilizing them in a word cloud to efficiently filter and analyze data. Based on results from various analytical approaches, data mining is conducted through centroid clustering using Lloyd's algorithm [16], [17]. This facilitates the creation of clusters based on different classification scenarios, such as consumers providing diverse feedback about pharmacies and clinics.

3) *The Advantages of Text Retrieval*

Health policies and administration

- Allow detection of insurance fraud where doctors or patients unnecessarily claim drugs for a particular diagnosis.
- Predictive models in real-time fraud detection systems can identify what drug type or procedure is necessary for different diagnoses and when those conditions are violated.
- Having predictive models for fraud detection eases the financial burden of governments and insurance companies as well as preventing bankruptcy [18].

Decision support

- Allow clinicians to gain information on patients' diagnosis and treatment, especially for clinician with less experience.
- Reduce human subjectivity
- Effective management such as monitoring drug orders, clinical complexity management, and administrative work support.

Disease prediction

- Reduce medical error by improved prediction models based on patient data recorded in the EHR
- Minimise adverse drug reactions and side effects
- Predictive models support and provide medicine new knowledge and founding

4) *The Disadvantages of Text Retrieval*

High financial costs

- Adoption and implementation costs
 - Purchase and installation of hardware and software
 - Conversion of analogue records to electronic
 - End-user training
- Maintenance costs

- Regular upgrades for hardware and software

Disruption of workflow

- End-users (medical staff and providers) have to spend extra time learning how to use the system
- Researchers have estimated that EHR end-users spent 134.2 hours on implementation activities associated with getting and learning a new system [19].

Accurately analysing and extracting unstructured data.

- Unstructured format of medical records as well as its complexity makes it hard for them to be extracted and analysed
- Similar or contradicting findings from unstructured can be challenging for the classification process.

Patient Security and Privacy

- Legal and ethical concerns arise when using text mining to extract and analyse sensitive information about patients such as identity, name, banking details without consent, resulting in misconduct or misuse [20].

4 Challenges in Implementation

1) Data Extraction

Developing an OLAP system faces the challenge of data conversion, particularly when dealing with unstructured data like medical notes, clinical tests, observations, and feedback forms. To facilitate analysis, these forms need to be transformed into an electronic format. However, this process can be time-consuming, given the complexity of analyzing and extracting information from textual data. In addressing this issue, Arfaoui and Akaichi [21] suggest the utilization of an Optical Character Recognition program for data conversion. This program demonstrates a $\geq 99\%$ agreement between data collection forms and scanned data. The author emphasizes that the time spent on manual data entry surpasses the efforts required to develop a scanning system, input data, and verify its accuracy.

2) Low Currency

To effectively leverage the vast amounts of data collected in healthcare centers, the adoption of dashboards generated through Online Analytical Processing (OLAP) is becoming imperative. These dashboards present insights derived from OLAP, enabling data analysts to navigate the database and delve into specific subsets of the data [21]. While the use of OLAP-created dashboards holds significant benefits for the healthcare industry, there are challenges associated with their implementation. The primary challenge of OLAP systems lies in the delayed currency of data findings. This delay is attributed to the extended time required for data to undergo Extraction, Transformation, and Loading (ETL) processing into the OLAP environment. Consequently, OLAP systems face limitations in achieving real-time data analysis, given the resource-intensive and complex nature of their processes compared to conventional transactional data processing methods. OLAP environments necessitate the aggregation of data over time, with large-scale queries typically executed on a monthly or quarterly basis. As a result, the data produced in OLAP exhibits high latency, being considered "near real-time" at best [21].

5 The Dashboard Analysis

In response to the challenges encountered by HHS amid the pandemic, marked by the strain on hospitals and the need to analyze consumer behaviors, a dashboard has been developed. This dashboard is designed to facilitate various analytics and is powered by a cleaned and transformed dataset stored in a data warehouse. The aim is to establish an integrated information management system to effectively address the current healthcare scenario.

precisely identify locations where complaints regarding distance and crowded conditions originate. The results reveal that Colorado, Nevada, Oregon, and Utah are grappling with challenges related to hospital strains and a scarcity of pharmacies and clinics. The legitimacy of the feedback is underscored by a 0% bounce rate, indicating that individuals spent a substantial average time of 4.5 minutes researching and providing feedback. Therefore, for informed decision-making, HHS should prioritize the establishment of additional clinics in Colorado, Nevada, Oregon, and Utah, along with a targeted expansion of pharmacies in Nevada.

1) Distinguishing Legitimate comments from Hearsay

Among the individuals providing feedback, there are those who raise non-legitimate complaints. The pie chart highlights a sentiment score of 0.25, indicating a lower end of the spectrum. Notably, feedback with this sentiment score is consistently attributed to the same user with the userid "derie," situated in Wyoming—the sole visitor from that state. The user repetitively returns to the website, registering five instances of providing identical negative feedback, as depicted in the bar chart categorizing them as a returning visitor. Despite spending an average of three minutes on the HHS website, the user exhibits a 2.5% bounce rate, suggesting a departure during one of the visits. Identifying recurring sentiment scores from the same user raises concerns about the legitimacy of the feedback, categorizing it as potential spam. Such feedback may not warrant significant consideration in the decision-making process. Consequently, for informed decision-making, specific measures will be implemented to filter out these repetitive comments, ensuring they do not mislead other customers contemplating a visit to the particular pharmacy.

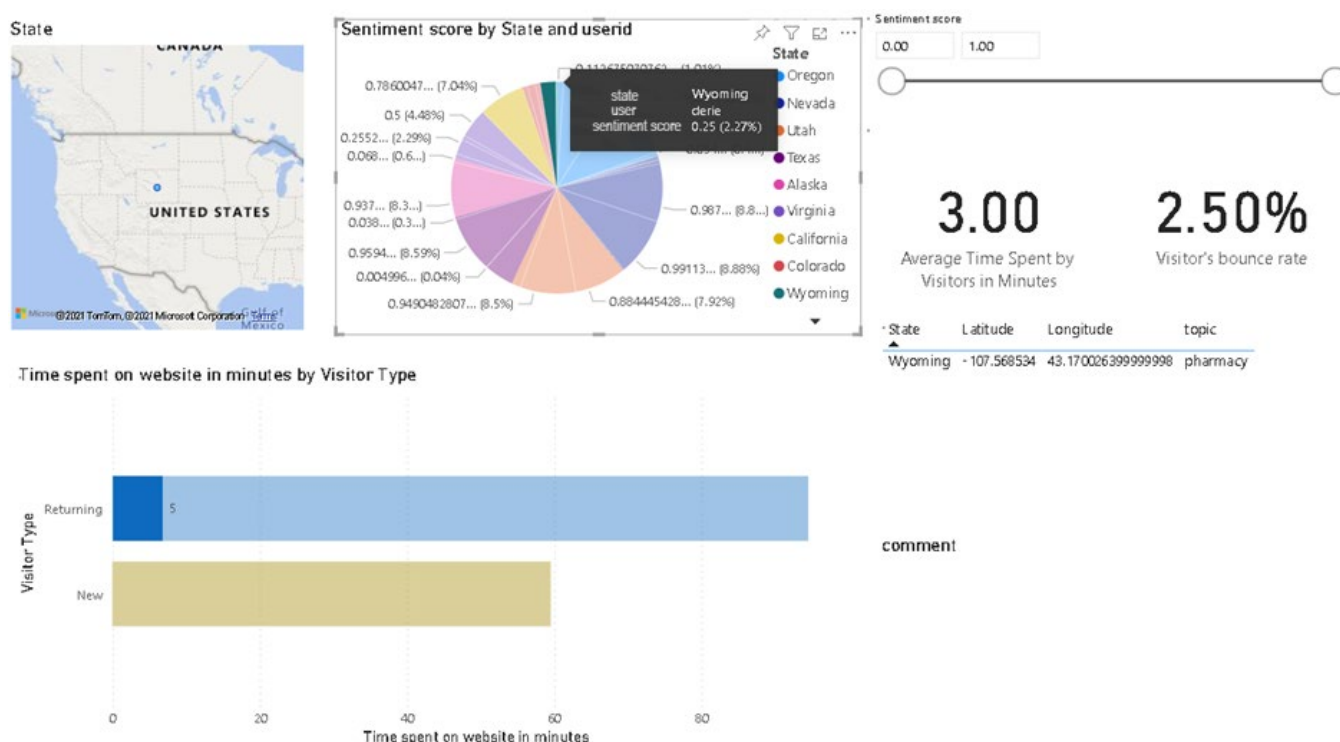


Figure 3. Dashboard Analysis for hearsay

6 Conclusion

In summary, valuable insights have been obtained regarding the challenges faced by numerous healthcare organizations in the industry, leading to effective solutions for operational issues within HHS. The significance of data integration for comprehensive analyses has become increasingly evident, driven by the growing volume of healthcare data and challenges in information management. Text analytics has played a pivotal role in providing a versatile solution, extracting unstructured text to categorize data through transformation and loading into the data infrastructure. Subsequently, the dashboard derived from this infrastructure has been instrumental in conducting web analytics, geospatial analysis, and social media analytics. The insights gleaned from these analyses have informed decision-making processes, aiding in addressing the challenges confronted by HHS.

Moreover, the incorporation of IoT devices enables the presentation of real-time analysis reports to healthcare professionals through stream analytics, enhancing their ability to attend to patients more efficiently. Looking ahead, as diverse data types continue to be collected, digitizing storage and analyses on cloud servers will prove essential in handling the data explosion within the healthcare sector. This approach promises a faster decision-making process for HHS and contributes to cost reduction in storage infrastructure.

BIBLIOGRAPHY

- [1]. G. Krishna and S. Howard, "How Yemen's Healthcare has been destroyed," *BMJ*, 2021. doi:10.1136/bmj.n3110.
- [2]. M. Chapman, J. Philip, and P. Komesaroff, "Systems of care and experience for Dying well," *Toward an Integrated Science of Wellbeing*, 2023. doi:10.1093/oso/9780197567579.003.0013.
- [3]. G. Rey, "The core galilean idea and some crucial data," *Representation of Language*, pp. 13–44, 2020. doi:10.1093/oso/9780198855637.003.0001.
- [4]. "Infrastructure for research data storage and preservation," *Exploring Research Data Management*, pp. 147–158, 2018. doi:10.29085/9781783302802.016.
- [5]. J. of Healthcare Engineering, "Retracted: College physical education and training in big data: A big data mining and analysis system," *Journal of Healthcare Engineering*, vol. 2023, pp. 1–1, 2023. doi:10.1155/2023/9808964.
- [6]. M. Soni and D. K. Singh, "A key exchange system for secure data coordination in Healthcare Systems," *Healthcare Analytics*, vol. 3, p. 100138, 2023. doi:10.1016/j.health.2023.100138.
- [7]. H. Yeo, "A machine learning based natural language question and answering system for healthcare data search using complex queries," *2018 IEEE International Conference on Big Data (Big Data)*, 2018. doi:10.1109/bigdata.2018.8622448.
- [8]. M. Hyder and G. M. Nafees, "Harnessing deep learning and natural language processing (NLP)," *ENGINEERING THE FUTURE: MACHINE LEARNING AND DATA SCIENCE IN PRACTICE*, 2023. doi:10.52458/9789388996747.nsp2023.eb.ch-02.
- [9]. A. Zago Jorgetto, *Patient Safety & Pharmacovigilance Program for enhancement of patient safety system in medical institutions*, 2023. doi:10.26226/m.638624aefdc1e80014da8d82.
- [10]. C. Schwemmer and S. Ziewiecki, *Social Media Sellout - the increasing role of product promotion on YouTube*, 2018. doi:10.31235/osf.io/tskjb.
- [11]. J. M. Kudari, "Data Analytics to predict, detect, and monitor chronic autoimmune diseases using machine learning algorithms," *Advances in Medical Technologies and Clinical Practice*, pp. 150–182, 2021. doi:10.4018/978-1-7998-7188-0.ch012.
- [12]. M. Robakowska et al., "Increased demand for services provided in hospital emergency departments," *Emergency Medical Service*, vol. 9, no. 1, pp. 43–47, 2022. doi:10.36740/emems202201107.
- [13]. A.-E. H. Abd-Elrahman, "Telecommunications service quality, customer satisfaction and customer loyalty in pandemic times," *Management Research Review*, vol. 46, no. 8, pp. 1112–1131, 2022. doi:10.1108/mrr-08-2021-0595.
- [14]. T. Iyamu, "The integration of social media with healthcare big data for services," *Advancing Big Data Analytics for Healthcare Service Delivery*, pp. 126–140, 2022. doi:10.4324/9781003251064-9.
- [15]. T. Ramathulasi, U. Kumaran, and K. Lokesh, "A survey on text-based topic summarization techniques," *Advanced Practical Approaches to Web Mining Techniques and Application*, pp. 1–13, 2022. doi:10.4018/978-1-7998-9426-1.ch001.
- [16]. S. Misra, S. K. Saha, and C. Mazumdar, "A survey on Data Stream Analytics," *Handbook of Big Data Analytics. Volume 1: Methodologies*, pp. 175–208, 2021. doi:10.1049/pbpc037f_ch6.
- [17]. E. Gavagsaz, "Efficient parallel processing of K-nearest neighbor queries by using a centroid-based and hierarchical clustering algorithm," *Artificial Intelligence Advances*, vol. 4, no. 1, pp. 26–41, 2022. doi:10.30564/aia.v4i1.4668.

- [18]. M. H. Tekieh and B. Raahemi, "Importance of data mining in Healthcare," Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, 2015. doi:10.1145/2808797.2809367.
- [19]. N. Menachemi and Collum, "Benefits and drawbacks of Electronic Health Record Systems," Risk Management and Healthcare Policy, p. 47, 2011. doi:10.2147/rmhp.s12985.
- [20]. L. Pereira, R. Rijo, C. Silva, and R. Martinho, "Text mining applied to electronic medical records," International Journal of E-Health and Medical Communications, vol. 6, no. 3, pp. 1–18, 2015. doi:10.4018/ijehmc.2015070101.
- [21]. N. Arfaoui and J. Akaichi, "A new approach to generate hospital data warehouse schema," Hospital Management and Emergency Medicine, pp. 1–26, 2020. doi:10.4018/978-1-7998-2451-0.ch001.