

From Offline to Online: Utilizing Sentiment and Web Analytics to Navigate Retail Transformation

¹Bayani Krisanto Agustin, ²Alon Dakila Ting, ³Althea Keenan Dolores

^{1,2,3}Department of Computer Science, University of the Philippines Diliman, P. Velasquez St., Quezon City, The PHILIPPINES

e-mail : ¹bayani66@dcs.upd.edu.ph; ²alon81@dcs.upd.edu.ph; ³althea49@dcs.upd.edu.ph

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



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Corresponding Autor: Bayani Krisanto Agustin

Abstract

The covid-19 pandemic has forced the way businesses run, including the offline apparel store that needs to shift their business to an online shopping platform. This also means that the volume of unstructured data that needs to be analyzed has increased significantly. This unstructured data should be analyzed accurately to help businesses with decision-making and solve their problems such as understanding customer satisfaction levels and evaluating marketing approaches. One way to utilize unstructured data is by using text analytics. The data like customer reviews from the online shopping platform and social media can be integrated and analyzed using a sentiment analysis approach in order to gain a better understanding of customer satisfaction levels. Furthermore, web analytics can also be utilized to evaluate the current marketing approach and how to maximize the marketing strategy for the business.

Keywords— Text Analytics, Social Media Analytics, Web Analytics, Sentiment Analysis, Online Shopping Platform

1 Introduction

Due to the Covid-19 pandemic, the restriction of movement has forced people to shift their shopping behavior from offline shopping to online shopping [1]. Initially, people would love to go to their favorite stores for shopping, but nowadays people are already used to opening their phones and checking on their favorite e-commerce to shop as it is more convenient. This change in shopping behavior has also forced businesses to adapt, those who can adapt successfully will survive while those who cannot, eventually will be left behind [2]. The same also applies to Mariana Store, which is a Phillipines apparel store that focuses on sports equipment and shoes. Mariana Store has been established in 2015 and has gotten a lot of fame in society. The restriction of movement and customer demand has forced Mariana Store to expand its business and have an online shopping platform. However, as the online shop platform grew, they started to be unsure whether they had done well in marketing their product and satisfying the customers. Thus, this paper intends to explore how data analytics can help Mariana Store to solve the business problems it encountered.

1.1 Analytics' Application on Online Shopping Platform

Nowadays, data can be found everywhere, either in the form of structured data or unstructured data. Although businesses prefer to use structured data as it is easier to analyze, unstructured data can provide a lot more information if handled properly [3]. For online shopping platforms, one of the most valuable unstructured data is customer reviews. For an online shopping platform like Mariana Store, customer reviews can give insight into how the customer feels about the products and services offered which can be used to help businesses in the decision-making process and understand customer satisfaction [4]. Additionally, these customer reviews also can impact customers' purchase intention.

[5] identified that text analytics can be used to generate actionable and meaningful insight from unstructured textual data like customer reviews. Furthermore, [6] stated that text mining can be used to discover interesting patterns hidden in unstructured textual data. One text mining technique that can be implemented for analyzing

unstructured data like customer reviews is sentiment analysis. Sentiment analysis refers to a method to understand the Mariana Store's customer opinion by analyzing the sentiment of a text [7]. Sentiment analysis can be used to analyze the data retrieved from social media which aims to understand customer opinions towards the products and services offered by the businesses. Social media analytics look at the content of a plethora of social media like Twitter, Instagram, and Facebook and analyzes it to gain insightful knowledge. Social media analytics as mentioned before can be used to analyze customers' sentiment toward businesses specifically through their posts or comments, furthermore, it can also help businesses to understand the current trends in society which can be further utilized to strategize new products [8].

Additionally, a deeper understanding of how well the online shopping platform can be gained by using web analytics. Web analytics is another part of data analytics in which the web data is gathered and analyzed in order to understand visitors' behavior, it allows businesses to optimize the performance of their website as it derives insight from visitors' behavior [9]. According to Huidobro et al. [9], web analytics help businesses evaluate their marketing strategy by analyzing the traffic sources and help to understand the visitor by clustering them based on their behaviors in order to target the correct marketing strategy for each cluster, in addition, businesses can also further optimize their website content based on the derived insight and reach the target audiences effectively.

1.2 Literature Review

Previous research conducted by Rachmawati and Alfitrassalam, [10] has successfully utilized sentiment analysis to explore customer satisfaction in healthcare e-commerce, two models were built for the research, the first is an explanatory model was used to explain the relationship between sentiment scores from customer reviews and customers satisfaction, while a predictive model was constructed to predict the customer's satisfaction level based on customer's sentiment scores. While Ozyurt et al. [11] use text mining methods namely the Latent Dirichlet Allocation (LDA) topic model and sentiment analysis to study the relationship between customers' reviews and customer's repurchase intention by measuring the consumer perceived value.

Another research by Denecke [12] has developed a lexicon-based sentiment analysis algorithm that focused on comprehensive analysis. Zulfiqar et al. [13] also further discussed how social media analytics can be used to measure the effectiveness of marketing strategies for businesses such as promotional events. Zhao [14] proposed a social media analytics framework that involved trend analysis and social media analytics, the research stressed on the use of both temporal and spatial data in gaining meaningful knowledge. Schmidt [15] has conducted research on how clickstream data and web analytics helps online shopping platforms to increase visitors' buying intentions, rebuying intention, evaluating marketing strategies, and giving visitors accurate product recommendations based on their previous activity history.

2 Problem Statement

2.1 Difficulty in understanding customer satisfaction level

Narrative, emotion, and Insight research [16] have stressed how organizations struggle to proceed customer's online reviews data into a meaningful insight that can help the organization enhance the decision-making process. Not only because of the difficulty in organizing unstructured data but also because of insufficient knowledge about handling textual data. Mariana Store's online shopping platform has provided an attached online review section for every product which allows the customers to share their thoughts on the products and services. It has received a bunch of reviews from customers regarding the quality of its products and services however, it only remains there to help other customers get a deeper overview of the products and services. It has not fully utilized these reviews for further analysis such as determining the customer satisfaction level. The higher-ups believed that the best strategy to boost brand awareness is by understanding the customer satisfaction level since the business focus is to fulfill the customers' needs. This paper aimed to utilize text mining techniques such as sentiment analysis to analyze the available reviews shared by the customers in Mariana Store online shopping platform, in addition, social media analytics will also be used to get the customers' posts and comments on the Mariana Store official account.

2.2 Difficulty in understanding the effectiveness of the marketing campaign

As the platform grows, the marketing team needs to plan a new set of strategies for next season's marketing campaign. Currently, the Mariana store has applied digital marketing through various social media to increase its brand awareness, however, the marketing team is unsure whether the marketing strategy that it used has effectively boosted sales or it is the brand fame that derives more sales for the store. This paper proposes to use web analytics in

order to further investigate the traffic sources for all the online shopping platform’s visitors. Additionally, the Mariana Store can also analyze its website performance through several matrixes derived from web analytics. By understanding the visitors’ behavior, Mariana Store can cluster the visitors and understand the needs of each customer, which traffic are they from, whether are they new or returning customers, and so on. This insight can help the marketing team to strategize the correct marketing strategy for each cluster and make sure the business does not spend useless money on ineffective marketing campaigns.

3 Text Data Retrieval Perspective

3.1 Data Integration

As customer reviews are regarded as unstructured data, a proper transformation needs to be done to structure the data so it can be stored as a column inside OLTP. This paper proposes a framework to integrate unstructured data which is customer reviews to further analyze it using the text analytics approach. Mariana Store will retrieve all the customer reviews from various sources, especially its website and social media platforms and then the data will be stored in the data storage.

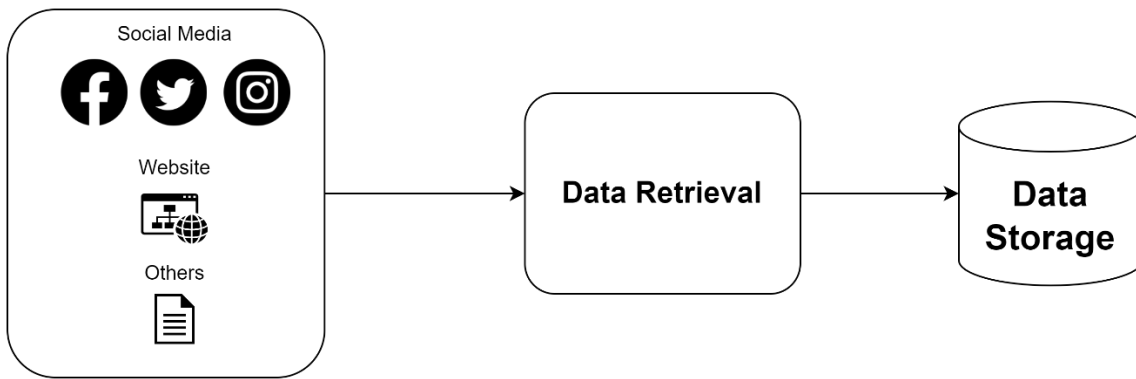


Figure 1. Initial Framework

Data integration is a crucial component that enables the use of text analytics to further analyze customer reviews data. Noverlita and Surbakti [17] highlighted the importance of data pre-processing for unstructured data before proceeding to do the analytics tasks. Noverlita and Surbakti [17] further described the pre-processing process as stop word removal, stemming, and important word identification. After the data has been structured it will then be combined with the other data from the website database, social media database, and others. Next, the data extracted from various sources will be joined together and further transformed into a similar format before it is loaded into the data warehouse. Alternatively, the streamed data from web analytics will go through the ELT process and be stored in the data lake, these data will be transformed when it is needed for the analytical tasks. Without data integration, it will be harder to generate the data insight effectively as a single table will have less impact compared to when it is combined with multiple tables. The whole data integration process is shown in Figure 2; it begins with extracting data from various sources all the way until the end goal which is in the form of a dashboard and report generated to be presented as the analytics results.

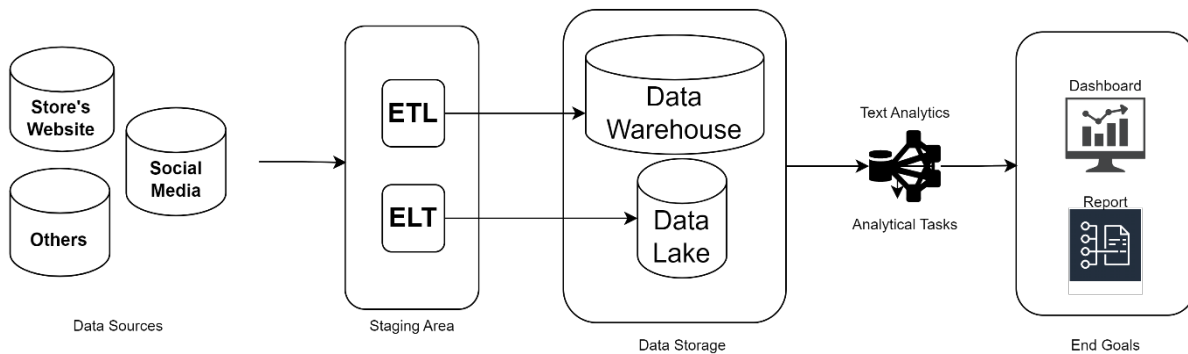


Figure 2. Integrated Framework

3.2 *Privacy Issues of Text Retrieval*

As described throughout this paper, textual data like customer reviews and social media postings can give impactful information for businesses in which they can further understand customers' attitudes toward their brand. However, it should be noted that this textual information might include personal information which if published might harm the writer. This personal information might be misused by some irresponsible individuals to commit serious offenses such as committing fraud. Each piece of personal information might be built by fraudsters to become the identifiers of someone and use this identity to commit crimes. Furthermore, there might be some people who accidentally disclose their sensitive information without understanding the consequences of it. Pande and Khamparia [18] stressed how the current privacy regulations increase users' worries about privacy. Thus, it is crucial for the organization to apply the appropriate method in preserving privacy concerns during text data retrieval. In relation to that, several studies have tried to propose privacy protection frameworks that can lessen users' worries about privacy issues. Zhan and Sheng [19] proposed to use double privacy-preserving text representation learning framework, while Roskó and Szöllősi [20] discussed the use of de-identification to protect users' privacy.

3.3 *Advantages and Disadvantages of Using Text Retrieval for Online Shopping Platform*

3.3.1 Advantages of Text Retrieval

As highlighted in section 1, text data can give a powerful insight if managed properly. This section will list some advantages of text data for online shopping platforms like the Mariana Store.

- Support Decision Making

The text data retrieved from customers' reviews can help the business to identify the excitement of a product and help the management decide whether a certain product is demanded by the user. Additionally, it can give the management insight into the customer's repurchase intention and decide on the inventory for that product to prevent out-of-stock situations.

- Improve services

Customer comments or feedback from Mariana Store's social media platform can help the management to understand the customer's needs and improve the services based on the information generated.

- Analyzing current trends

Text retrieval from social media allows businesses to understand the current trends in society. Analyzing trending topics related to the business domain can help Mariana Store to understand which kinds of products or brands are preferred by society.

3.3.1 Disadvantages of Text Retrieval

Unstructured data, as mentioned before, is useful, however it is important to note that it is harder to handle compared to structured data. This section will discuss the disadvantages of text retrieval for an online shopping platform like the Mariana Store.

- Complex data infrastructure

Without unstructured data, online shopping platforms can only use a simpler version of data infrastructure as the normal transactional data from the website can directly be loaded into the database as structured data without the need to pre-process the data. Unstructured data has given more workload to the online shopping platform data infrastructure as it requires the data to be transformed into a structured form before any analytical tasks can be done.

- Higher data infrastructure and maintenance cost

As the data infrastructure becomes more complex, the business also needs to put more resources into it including hiring a data specialist to handle the unstructured data. Additionally, the business also needs to spend a higher maintenance cost such as upgrading the hardware and software based on the newest technology.

- Privacy Issues

As highlighted in the previous section, the privacy issue is also one of the disadvantages of working with text retrieval, especially when dealing with social media text.

4 Results and Discussion

In order to handle and visualize the issues that the Mariana store is facing, a Social Media Sentiment Analysis and Web Analytics dashboard has been created.

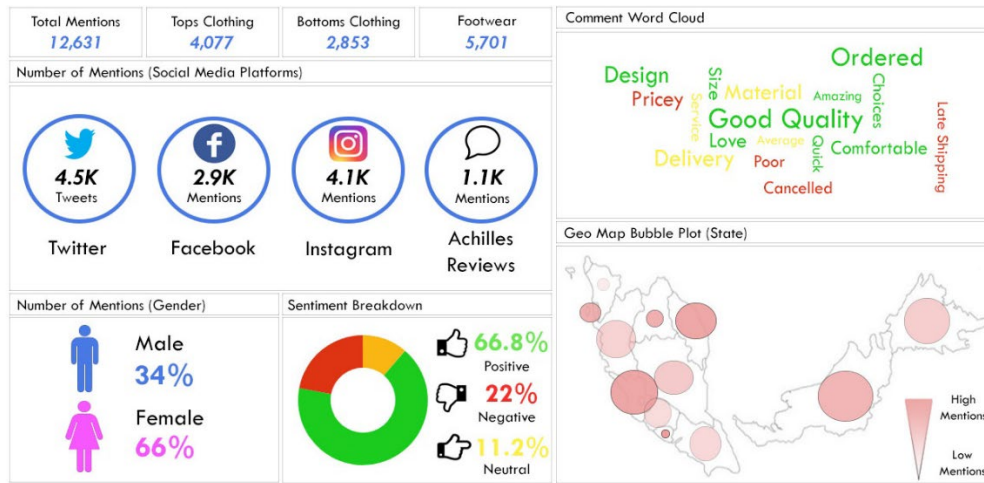


Figure 3. Social Media Sentiment Analysis Dashboard

Firstly, the bubble plot geo map in the dashboard is used to easily visualize and identify which states in the Philippines have been mentioning the Mariana store where the darker bubbles represent more mentions and the lighter bubbles in the plot represent less activity. Meanwhile, the donut chart shows a breakdown of the proportions of the sentiments by positive, negative, and neutral categories. Next, the number of mentions by social media platforms is used to determine how many people are tweeting or mentioning the Mariana store through Twitter, Facebook, Instagram, or Mariana’s own message board. Below, the number of mentions by gender is shown to identify which demographic has been more prevalent when mentioning the Mariana store. The total mentions and mentions for each clothing category such as the Tops Clothing, Bottoms Clothing, and Footwear categories are also displayed. These mentions are then shown in the word cloud to display the opinions and reviews of the customers regarding the Mariana store products so they can analyze and possibly make better decisions in improving and maintaining their services based on the review’s positive, negative, or neutral sentiments. Comments such as quality (positive sentiment), material (neutral sentiment), and canceled (negative sentiment) provide a more in-depth insight into understanding what the customers were happy, okay, or unhappy about and therefore help improve customer satisfaction levels.

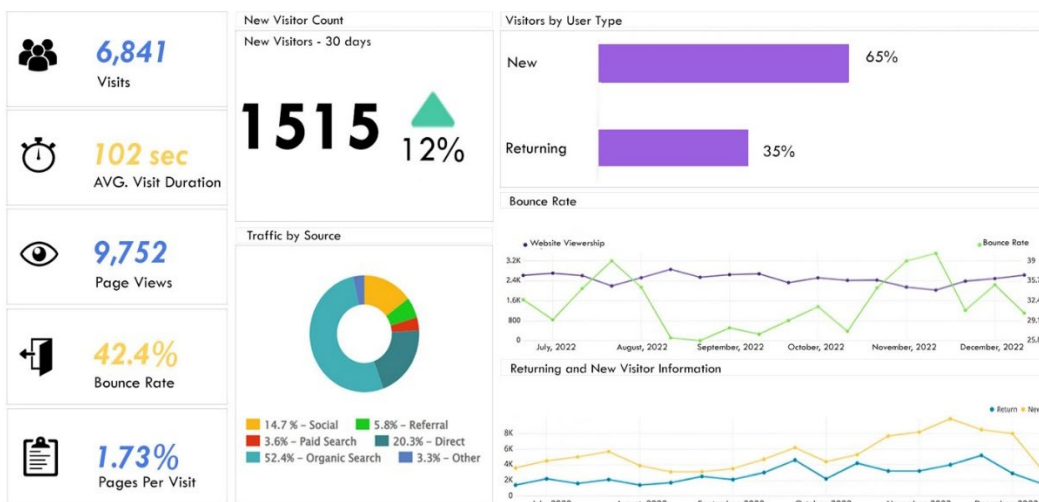


Figure 4. Web Analytics Dashboard

In the web analytics dashboard above, the number of visits, average visit duration, page views, bounce rate, and pages per visit are counted so the Mariana store can view how many people are visiting and viewing their website. The average visit duration shows how long the customer has been staying on the website before leaving, the dashboard also displays the bounce rate percentage of the customer and how many pages the customer opens during their visit. The counts highlighted in yellow indicate that the Mariana store should further delve into those aspects and improve upon them. With the new visitor count, it displays the increased percentage of visitors to the website for the past 30 days so the Mariana store can figure out whether their marketing strategies and campaigns have been improving their website's performance such as sales or brand fame. The horizontal bar chart shows the visitors by user type and displays the percentage of new and returning visitors. As indicated by the bar chart, there have been more new visitors than returning, indicating that the Mariana store's marketing strategy has been successful in bringing in new traffic. Furthermore, the donut chart shows the traffic by the source which indicates where visitors are arriving from. Next, the line charts show a more in-depth visualization of the bounce rate and returning visitors from the months of July 2022 to December 2022. The chart indicates that the viewership of the website has been steady while the bounce rate fluctuates with each month. The chart also indicates that there have been more new visitors than visitors returning every month. Therefore, as for decision-making, the Mariana store has been implementing the correct marketing strategy to bring in new traffic although they need to improve in making the visitors stay longer on the website. This is because there is no point in marketing to bring in new visitors if they are just going to leave soon after visiting without purchasing or viewing the products.

5 Challenges in Implementation

5.1 Data Security

By using text analytics and the ability of text mining, the data of the customers and their reviews were attained, the data will then be separated to their specific data type accordingly and stored in the data warehouse for analysis. The same method was also applied to social media analytics at all social media sites. The unstructured data collected like how long a user spends on a website, how many pages are visited and what page was visited were all stored in the data warehouse and ready to be analyzed. One of the main challenges we face is data security. Once the data has been created, the data needs to be secured. Most social media publishers are against it as it violates their users' terms of privacy.

5.2 Context-dependent Error

By having the customers unstructured data, rearranging and cleaning the data is a priority. Data from customers' comments were taken into the analysis process by implementing sentiment analysis. One of the challenges we faced during this phase was that customers' comments were using sarcasm to express their negative sentiment, and the analysis will detect the comments to be a positive sentiment hence it was quite a nuisance to filter out from the analyzed data.

5.3 Data Accuracy

The data stored by the customers in the data warehouse attained from the World Wide Web has undergone the processing phase, processed the data and turned it into information. Typically, the data stored will become metrics and the only issue we had been the data accuracy, the data can be inaccurate due to some customers using VPNs and ad-blockers to access the website. Ad-blockers usually trace JavaScript-based ads. Thus, using non-JavaScript for visitor tracking can counter it. Using all this software to enter our website will affect the website's bounce rate, time on the page, and where it was accessed.

6 Conclusion

Ever since Mariana Store launched its online platform store, it has been having difficulty understanding customer satisfaction levels and difficulty understanding the effectiveness of the marketing campaign. We attempt to solve the problem by performing analyses. Text analytics was the first implementation to generalize a solution as unstructured data is collected and classified by transforming it and loading it into the data warehouse to construct the data infrastructure. The data infrastructure was used to perform web analytics and social media analytics. The results gained from the analyses performed were shown on the dashboard and helped Mariana Store to make strategic business decision-making. Furthermore, using social media analytics helped enable Mariana Store to keep up with

the current ongoing trends. Sentiment analysis also helped improve the quality of the online store by analyzing the negative sentiments in customers' reviews and executing a business strategy.

BIBLIOGRAPHY

- [1]. T. E. Mofokeng, "Antecedents of trust and Customer Loyalty in online shopping: The moderating effects of online shopping experience and E-shopping spending," *Heliyon*, vol. 9, no. 5, May 2023. doi:10.1016/j.heliyon.2023.e16182
- [2]. D. Kramer, "National Labs are forced to adapt during coronavirus outbreak," *Physics Today*, vol. 2020, no. 03, Mar. 2020. doi:10.1063/pt.6.2.20200318a
- [3]. P. Macharia, A. Pérez-Navarro, I. Inwani, R. Nduati, and C. Carrion, "Developing an unstructured supplementary service data-based mobile phone app to provide adolescents with sexual reproductive health information: A human-centered design approach," *Appl. Sci.*, vol. 12, no. 15, Apr. 2022. doi:10.21203/rs.3.rs-1551057/v1
- [4]. "Customer satisfaction and perception towards service quality by using fuzzy multi criteria decision making model (FMCDM)," *Central European Management Journal*, 2022. doi:10.57030/23364890.cemj.30.4.240
- [5]. R. A. MAYNARD, "How researchers can make the evidence they generate more actionable," *Next Generation Evidence*, pp. 35–45, 2023. doi:10.5040/9780815743576.0007
- [6]. A. Charitopoulos, M. Rangoussi, D. Metafas, and D. Koulouriotis, "Text Mining Technologies applied to free-text answers of students in e-assessment," *Discover Computing*, vol. 28, no. 1, Jan. 2025. doi:10.1007/s10791-024-09496-9
- [7]. C. C. Aggarwal, "Opinion mining and sentiment analysis," *Machine Learning for Text*, pp. 491–514, 2022. doi:10.1007/978-3-030-96623-2_15
- [8]. C. Y. Ng, K. M. Law, and A. W. Ip, "Assessing public opinions of products through sentiment analysis," *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, pp. 1422–1440, Jun. 2022. doi:10.4018/978-1-6684-6303-1.ch073
- [9]. A. Huidobro, R. Monroy, and B. Cervantes, "A high-level representation of the navigation behavior of website visitors," *Applied Sciences*, vol. 12, no. 13, p. 6711, Jul. 2022. doi:10.3390/app12136711
- [10]. I. Rachmawati and E. Alfitrassalam, "Identify issues that influence customer satisfaction using sentiment analysis and topic modeling," *International Journal of Current Science Research and Review*, vol. 07, no. 09, Sep. 2024. doi:10.47191/ijcsrr/v7-i9-26
- [11]. O. Ozyurt, A. K. Kabakus, and A. Ayaz, "Exploring scholarly journal content: Topic modeling analysis using LDA algorithm," *Text Mining in Educational Research*, pp. 71–95, 2024. doi:10.1007/978-981-97-7858-4_5
- [12]. K. Denecke, "Lexicon-based medical sentiment analysis," *Sentiment Analysis in the Medical Domain*, pp. 67–69, 2023. doi:10.1007/978-3-031-30187-2_10
- [13]. S. Zulfiqar, A. Lakho, and K. Nizam, *Social Media Analytics: Application towards social media marketing*, Jul. 2022. doi:10.20944/preprints202207.0363.v1
- [14]. B. Zhao, "Social Media Analytics," *Geography*, Oct. 2023. doi:10.1093/obo/9780199874002-0275
- [15]. H. C. Schmidt, "Clickstream Analytics," *Encyclopedia of Big Data*, pp. 169–170, 2022. doi:10.1007/978-3-319-32010-6_36
- [16]. "Narrative, emotion, and Insight," *Choice Reviews Online*, vol. 49, no. 06, Feb. 2012. doi:10.5860/choice.49-3193
- [17]. J. Noverlita and H. Surbakti, "Streamlining stock price analysis: Hadoop ecosystem for Machine Learning Models and big data analytics," *International Journal of Information Technology and Computer Science*, vol. 15, no. 5, pp. 25–34, Oct. 2023. doi:10.5815/ijitcs.2023.05.03
- [18]. S. D. Pande and A. Khamparia, "Networks attack detection on 5G networks using data mining techniques," *Appl. Sci.*, vol. 14, no. 1, Mar. 2024. doi:10.1201/9781003470281
- [19]. H. Zhan and V. S. Sheng, "Privacy-preserving representation learning for text-attributed networks with simplicial complexes," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 13, pp. 16143–16144, Jun. 2023. doi:10.1609/aaai.v37i13.26932
- [20]. T. Roskó and G. J. Szöllösi, "Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy," *First Monday*, Jul. 2021. doi:10.5210/fm.v26i8.10616