International Journal of Information Technology and Computer Science Applications (IJITCSA)
p-ISSN: 2964-3139 e-ISSN: 2985-5330
Vol. 03, No. 02, page 78 - 85
Submitted 11/7/2025; Accepted 19/8/2025; Published 20/8/2025

A Desk Review of The Application of Data Analytic on Tesla Inc.

Mulya Firmansyah
School of Management, Asia e University, Subang Jaya, Selangor Darul Ehsan, MALAYSIA
e-mail : mulya.sah88@gmail.com

Corresponding Autor: Mulya Firmansyah

Abstract
Tesla Inc. is frequently regarded as a pioneer in the fields of big data analytics and artificial intelligence. They generate, collect, and analyse a large amount of data every day for better decision-making for their business and for their self-driving car. However, a little effort was put into marketing. Customer segmentation and customer retention are noticed to be the problems Tesla Inc. should take into consideration. Therefore, the data that is relevant to solve these problems is extracted from various websites and social media platforms by using text-scrapping techniques. Web analytics of Tesla's official website is studied to analyse the demographic and geographic details of the audience. Geospatial analysis is also carried out to further analyse the top 5 countries the audiences are coming from. Customers' reviews that were collected undergoes sentiment analysis to determine whether it is positive, neutral, or negative. Text analytics is done in the later stage by gathering all the visualisations into an interactive dashboard and coming up with a possible solution.

Keywords— Text scrapping, Text analytics, Web analytics, Sentiment analysis, Geospatial analysis, Tesla, Customer segmentation, Customer retention

1 Introduction

Tesla is well known as the leading electric car manufacturing company internationally. Not only limited to manufacturing, but Tesla inc. also designs, sells, and leases electric vehicles and energy generation and storage systems [1]. Since the Tesla cars that are produced are technologically advanced and have self-driving ability, a huge amount of data is needed to be processed. Each car is said to provide approximately 2 to 5 terabytes of data weekly [2]. With the installation of eight cameras, 12 ultrasonic sensors, and a forward-facing camera, the process of data collection becomes much easier and faster [1], [2].

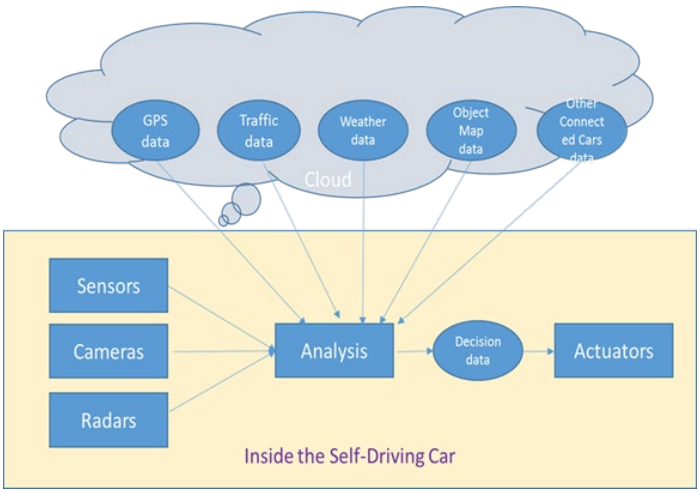

Figure 1. Data Flow Model of Self-driving Car

The data flow model above represents the infrastructure of Tesla company. The data collected from the hardware inside the self-driving car needs to interact with the data from the company cloud in order to make any decision. With no interaction, there is no possibility of the self-driving car existing as it would be just a normal regular car. There are varieties of data that need to be collected by the company so that the self-driving car can make decisions on its own. In order for the self-driving car to process emergency braking, object detection, self-parking, identifying traffic light and etc., they need to perform real-time analytic. Therefore, the data collected and loaded into the company's cloud should be GPS data, traffic data, weather data, other connected cars data and object mapping data on a real-time basis [3].

1.1 *Literature Review*

Text scrapping techniques can be used to extract insights/data from different data sources such as websites and social media platforms, as previous research has demonstrated. Holubiev and Simishko [4] presented a practical approach that makes use of text mining and web scraping to gather information about the audience visiting Tesla official website. He also stated that text mining method helps to analyse the trend in their dataset. It breaks down the potential targeted audiences of Tesla Inc based on their age distribution. On the other hand, Fardi and Motlagh [5] place an emphasis on customers' reviews as they play an important role in managing customer retention. With this data, sentiment analysis can be done to observe the satisfaction of the customers with buying Tesla's products. It helps to identify the customer who is likely to churn.

2 Problem Statement

2.1 *Customer Segmentation*

Although Tesla Inc. makes billions of dollars in profit, there are some problems faced by them. The two obvious problems would be the problem in customer segmentation and the problem in customer retention [6]. Tesla cars can be bought and leased by anyone no matter how young and how old they are. Since there is not any specific customer segmentation, it would be hard for the Tesla company to find their targeted customers and reach them. Having no target market and saying that the Tesla cars are for everyone shows that the Tesla products do not appeal to anyone in particular. Besides, targeting everyone might be costly and there is less chance that the product will be appealing to everyone.

However, Tesla Inc. can identify their potential customers through their official website traffic analysis. Here tesla company can observe what is the age range of the customers that are interested in their products.

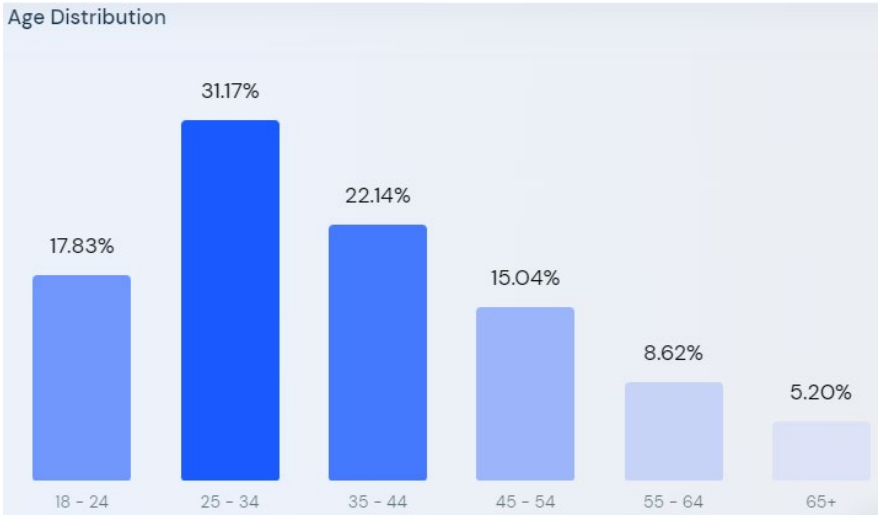

Figure 2. Bar chart of audiences' age distribution

Based on the traffic analysis on Tesla official website through similarweb.com, we can say that the majority of the page visitors are between the ages of 25 and 34. Therefore, this age group is more prevalent and seems to be the potential targeted customer for the Tesla Inc. Targeting users within this age range from all around the world might be costly and illogical as Tesla is only available in certain countries. Thus, Tesla company should identify from which countries their potential customer is coming.

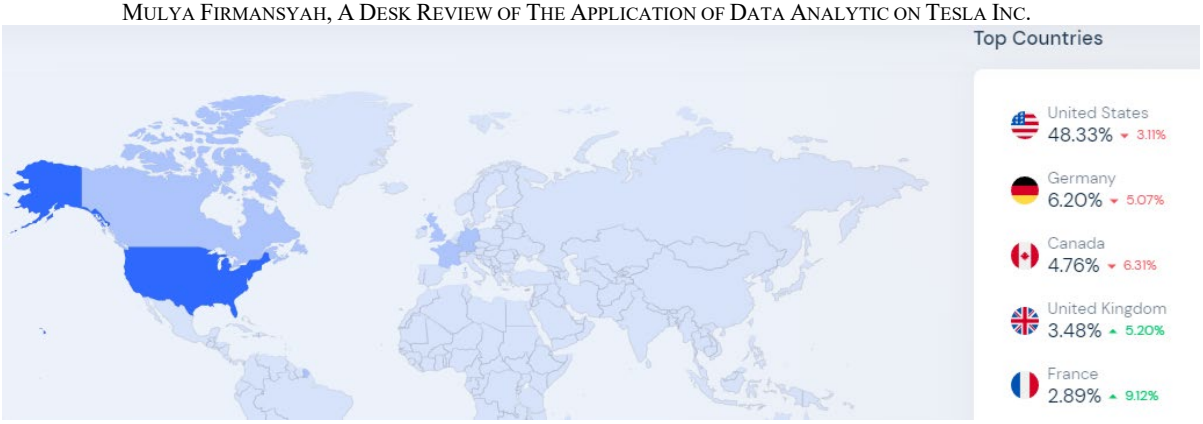

Figure 3. Geomap of top 5 countries by percentage of audiences' visit

As we can see, the geospatial analysis from the similarweb.com shows that the top 5 countries that fill the traffic in Tesla official website are United States, Germany, Canada, United Kingdom and France. By looking at the analysis above, Tesla Inc. can narrow down their target customer base to only targeting customers with the age range of 25 to 34 from US, Germany, Canada, UK and France countries. As the demographic and geographic segmentation is done, Tesla Inc. can use Facebook paid ads method to easily reach their targeted customers.

2.2 *Customer Retention*

Besides customer segmentation, customer retention (churn rate) also seems like a problem for Tesla Inc. The overall Network Promotor Score (NPS) of Tesla is 36. NPS is known as the customer loyalty metric where it would not only measure how likely the customers will purchase but also how likely they will recommend the product to their friends and family [7]. There are three main labels in NPS such as promoters, passives, and detractors. 61% of Tesla's customers are labelled as promoters. These customers value the product very much, remain as the loyal customers, and very likely to repurchase and make referrals. On the other hand, 25% of Tesla's customers are labeled as detractors. These customers are not satisfied with the product and highly likely to churn [6].

In order for Tesla Inc. to reduce the churn rate, they will need to focus more on the detractors. These customers sometimes teach us valuable lessons. There are many ways in which Tesla Inc. can reduce churn. One of them is by providing transparency to the customers. For example, if any of the issues occur during manufacturing, the respective customers should be informed in advance. Furthermore, Tesla company should take the feedback given by the customers into account. This feedback can help Tesla with product development and enhance customers' driving experience. In addition to that, Tesla can also apply the closed-loop feedback method to understand the feedback much better and make the customers feel heard. As a result, the detractors might feel that their feedback is being appreciated and there are chances that they will change their minds and be the promoters. Therefore, the churn rate is reduced [6], [8].

3 Data Retrieval

3.1 *Data Integration*

Data integration plays a vital role in assisting the company in making decisions. With data integration, Tesla Inc. can analyse the valuable data gathered from different data sources in just one integrated system. Following data integration, Tesla inc. can come up with the solution for the problems stated in task 2 which are customer segmentation and customer retention. In order to solve the customer segmentation problem, web traffic data from their official website is required. Third-party tracking websites such as similarweb.com could help Tesla to perform web analytic for their website. Then, Tesla can retrieve all the necessary web traffic data such as visitors' demographic, geographic and referring sites from their site [9].

For customer retention problems, Tesla requires data of customers' reviews [6], [8], [9]. These reviews can be obtained by performing social media analytics on social media platforms such as Facebook and Twitter. In addition, some review websites such as Trustpilot.com can also be used by Tesla to obtain customers' reviews. After deciding on which data to be used to solve the problems, Tesla Inc. can start ETL processing.

In the extraction phase, Tesla can perform web scrapping techniques from the respective data sources. For web traffic data, the data source would be similarweb.com and the visitors' demographic, geographic and referring sites data will be scrapped based on the "tesla.com" keyword. For the customer reviews data, the data source would be

Trustpilot.com, Facebook and Twitter. As for the Trustpilot.com, all the reviews will be scrapped based on the "tesla" keyword. The same keyword will be used to scrap the postings and comments from the Facebook and Twitter platforms. Once all the data is extracted, Tesla can now continue with the transformation process. The data with different formats, missing values, duplicated, incorrect, and incomplete will be cleaned. Lastly, Tesla can load this collected data into their data warehouse. Therefore, we can say that the process of data integration is done by integrating the data from similarweb.com, Trustpillot.com, Facebook, and Twitter [8].

With all this data being in one place, Tesla Inc. can easily analyse their problems through visualisation. By using visualisation tools, Tesla can create an interactive dashboard to observe the details of customers based on their demographic, geographic, which site they are coming from and also their review on Tesla products. In addition, the interactive dashboard allows Tesla to filter information and deeply analyse the trend and results of specific data variables.

For a more detailed visualisation, geospatial analysis can be performed by Tesla Inc. by creating a Geo map using the geographic data of the customers. Besides, Tesla can also carry out sentiment analysis on their customers' reviews by using machine learning tools such as Microsoft Azure. By doing this, Tesla can categorize good and bad reviews, hence, identify the customers who are likely to churn [6], [10].

3.2 *Privacy Issues*

In this modern age of technology, the World Wide Web has been developed into a tool of wonders which everybody uses. Naturally with the growth of the World Wide Web, the amount of online text data on it would increase as well. With the help of text retrieval, we can quickly get the data needed from the massive text data available online. While the tool is very useful in its own right, it also threatens the user's privacy. The danger can be categorized into 2 groups. The first one which the user queries itself are compromised, and others know what you are looking for. The second one is the compromise of the results of the user query. Both these privacy issues would cause serious concerns by the users as their rights to privacy are not respected [11].

In terms of Tesla Inc., the privacy of their customers must be protected. When retrieving text data from sources like Facebook and Twitter for sentiment analysis, a person's identity and personal data might be taken without consent (Managing data acquisition). So, to obtain text data without compromising anyone's privacy, Tesla Inc. must be aware of the restrictions on the data and the collection methods. With such restrictions, the text data retrieved should be censored of any personal identifiable information that might compromise the people's private information. Another way to protect the people's private information from being obtained is to only collect necessary data. By only getting what is needed, it is less risky. So, there is less data and less risk. Among all things, personal identifiable information can also be avoided unless necessary.

Privacy issues are important as it is our right to control who can have access to our private lives and activities. It allows us to protect our private information from sharing to the public. It also protects us from physical danger by having our real time location leaked. All these dangers should be prevented by properly encrypting our queries and the results so nothing can be used to endanger the users. This can also be done by censoring anything that might link or relate to the user.

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

For businesses, it is important to be constantly evolving and improving to compete against others for business. One way to do this is by using text data retrieval to aid in the process. It can help by identifying specific areas for improvement. This is done by using text analytics to analyse unstructured text and turning into useful business intelligence. Another benefit text data retrieval can bring is the ability to analyse text in any language or format. By knowing what the customers are saying on social media, business can stand to gain valuable insights into customer experiences in the business. Considering both points above, by using text data retrieval business can make better decisions and strategies to improve their weaknesses. By doing so business can outcompete with their competitions and gain more revenue.

However, text data retrieval is not a tool without its disadvantages. By using it when handling users' personal data, it might cause some areas of concern. The concern lies when the personal data is collected and used, privacy is lost which raises problems stated above. Not only that, by mishandling the personal data of users, it might cause huge sums of money in damages. Finally, the business reputation might also be affected if the business decides to mishandle the data by selling personal data to other companies.

3.4 Dashboard on Tesla Inc.

The dashboard shows six visualisations based on different data that can be observed. The first visualisation would be the bar chart of audience's age distribution. The age group of audience that visited the Tesla official website the most are between the age of 25 and 34 with a percentage of 31.17%. The age group with the least visits are the

audience with the age more than 65. Therefore, we could say that the 25 to 34 age group is likely to be a potential target customer.

Figure 4. Tesla Inc. Dashboard

The second visualisation would be the Geo map of top countries by percentage of audience's visit. As we can see from the chart, the audiences from the United States contributed to a higher website traffic with a percentage of 48.33%. Although France is one of the top 5 countries by visits, it has the percentage of only 2.89%. Hence, we could say that there are more potential customers in United States. In the third visualisation, the trend chart of bounce rate is displayed. Bounce rate is the important metric in web analytics to show how fast the audience switches to another page after landing on a particular page. The bounce rate of Tesla website is 36.52% which is not that high. This shows that the audience are interested in the products or design as they spend more time on the page they landed.

The fourth visualisation shows the top referral website. LinkedIn has the highest percentage of 14.09% followed by Twitter and YouTube with 12.7% and 6.22% respectively. Although google.com is a popular search engine that can lead to any website, it has only 3.84%. The percentage suggests that most of the audience who visited Tesla's website came directly from LinkedIn. The last two visualisations are based on the sentiment analysis of the customers' reviews. The pie chart shows the classification of the reviews based on their sentiment score. Out of 154 thousand reviews, there are 61% positive reviews, 14% neutral reviews, and 25% negative reviews. The classification of these reviews contributed to the net promoter score (NPS) of 36 which is considered good.

The last visualisation shows the word cloud of the most used words in the customers' review. The words such as tesla, convenient, affordable, awesome, and amazing have a larger frequency as it appears bigger is size. The negative words such as worst, bad, expensive and slow have a smaller frequency as it appears smaller in size. This shows that majority of the customers like Tesla products and their services.

4 Challenges in Implementation

The main challenge faced by Tesla Inc. is implementing text mining to find the appropriate patterns and trends to derive useful information from the massive number of words to meet and satisfy the customers [3], [7]. Through text mining, which uses artificial intelligence to extract and transform unstructured data into structured data for analysis, Tesla can then identify customer dissatisfaction and trends to improve customer experience. This process can be applied to various mediums, such as social media, online forums, surveys, and feedback [5], [7]. However, texts from mediums containing sarcasm, idioms, or multilingual data will result in inaccurate detection results in addition to noisy data during text mining. Hence, text mining alone is inadequate to identify customer dissatisfaction for improving the overall customer experience. In addition, word clouds are considered to be another challenge since they capture data based on word frequency. On the other hand, forum interpretations are expressed in sentences. Due

to this drawback, the word cloud result will likely provide an incorrect interpretation that does not lead to an actionable decision [8], [11].

5 BEST PRACTICES

5.1 *Data Collection*

The collection of data begins with web scraping operations on forums such as teslaownersonline.com and teslamotorsclub.com, which contains complaints, feedback, and reviews from users and owners.

5.2 *Web Scraping*

Web scraping is an automated and targeted extraction of data by script in a conventional method for retrieving large quantities of web material [8]. This is especially useful in this case when scraping for complaints, feedback, and reviews on forums containing information relating to Tesla Inc. with the aim of improving customer experience. As shown in Figure 5, the flow of the web scraping process is illustrated.

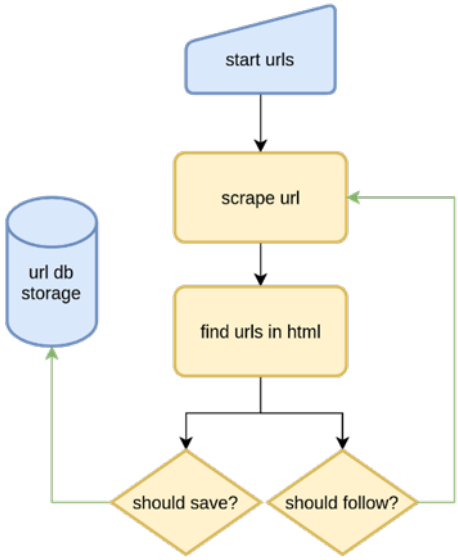

Figure 5. Web scrapping flow

5.3 *Data Cleaning*

Data Cleaning is the process of correcting or deleting inaccurate, improperly formatted, duplicate, or incomplete data from a dataset. The main purpose of this process is to eliminate incorrect values and inaccurate information that might result in poor decision making.

5.4 *Data Mining*

Data Mining is the process of finding significant patterns and trends, including analyzing large blocks of information. The purpose of this process is segmentation for clustering. K-Means based clustering is best suited for customer service segmentation. According to Tenya [8] and Vieri et al. [7], web forums of user posts that were analyzed using K-Means clustering revealed a classification of positive/negative user posts. In our case the same method can be applied for analysis.

5.5 *Data Visualization*

Data Visualization is essential these days conveying the message and helping Tesla Inc. make concise improvements, especially with the number of forums threads and replies increasingly adding one after another. Word cloud is one of the examples whereby frequency of words is determined by the boldness and the size of words as shown in Figure 6.

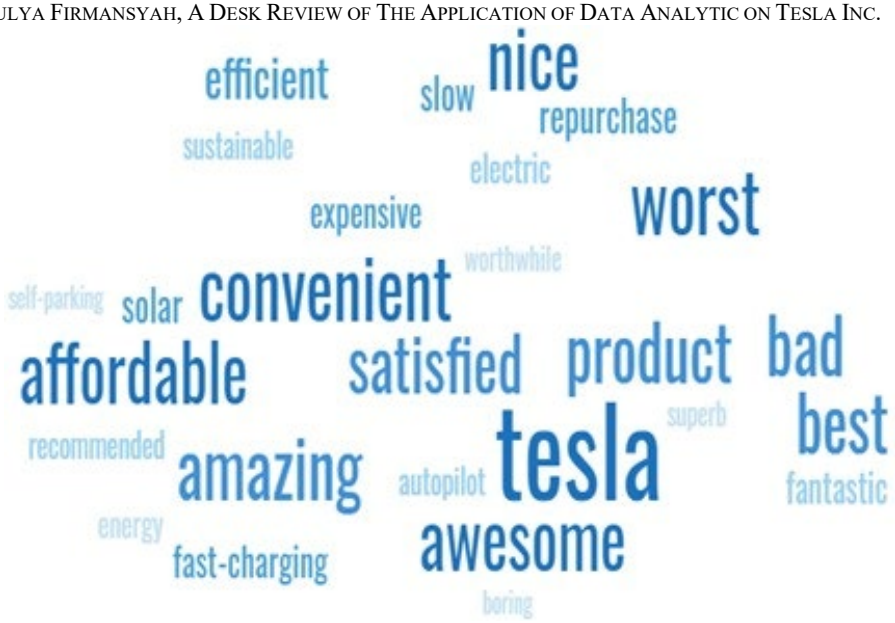

Figure 6. Word cloud showing keywords from customers' reviews

6 Conclusion

As a result, the present research indicates the practices of text analytics in text data retrieval with its concerns and the challenges implied for providing outcomes of word frequency by word cloud and pattern recognition by text mining. As modern day decisions are challenged by complexity and massive amounts of data, text analytics assists in analyzing the appropriate considerations that are usable for enhancing business decision making in terms of business operations and management. While managing a large amount of text data also results in the issue of a drop in data quality. Hence, using text analytics that has been preprocessed with data cleaning enhances machine decision support to lower the likelihood of misleading outcomes. With greater technological advancements, text analytics is a new artificial intelligence trend that is expanding quickly for obtaining essential business insights from derived text data.

BIBLIOGRAPHY

[1]. A. H. Fathima and K. Palanisamy, "Energy Storage Systems for energy management of renewables in Distributed Generation Systems," Energy Management of Distributed Generation Systems, Jul. 2016. doi:10.5772/62766.

[2]. A. Pál, Technical aspects of extracting millions of asteroids from tess: Data Processing and pipeline structures at the flow of 4 Terabytes/month., Sep. 2022. doi:10.5194/epsc2022-1226.

[3]. F. Rollo, C. Bachechi, and L. Po, "Semi real-time data cleaning of spatially correlated data in traffic sensor networks," Proceedings of the 18th International Conference on Web Information Systems and Technologies, pp. 83–94, 2022. doi:10.5220/0011588500003318.

[4]. V. Holubiev and V. Simishko, "Web scraping and text mining of Ukrainian news articles about ecology," 2021 11th International Conference on Advanced Computer Information Technologies (ACIT), pp. 672–675, Sep. 2021. doi:10.1109/acit52158.2021.9548450.

[5]. A. Fardi and A. S. Motlagh, "Elaborating the willingness to participate in social media with an emphasis on the intermediary role of the customers' perceived value," International journal of health sciences, pp. 6634–6651, Aug. 2022. doi:10.53730/ijhs.v6ns5.10312.

[6]. H. Surbakti and A. Taa, "Improving employees retention rate through knowledge management and business intelligence components," in Knowledge Management International Conference (KMICe), 2016, pp. 13–17.

[7]. J. Kristian Vieri, T. A. Munandar, and D. B. Srisulistiowati, "Exclusive Clustering Technique for customer segmentation in National Telecommunications Companies," International Journal of Information

Technology and Computer Science Applications, vol. 1, no. 1, pp. 51–57, Jan. 2023. doi:10.58776/ijitcsa.v1i1.19.
[8]. Y. Tenya, "A comprehensive exploration of text, web, social media, and geospatial analytics for informed decision making," International Journal of Information Technology and Computer Science Applications, vol. 2, no. 2, pp. 25–32, May 2024. doi:10.58776/ijitcsa.v2i2.148.
[9]. M. Santos Fernandez, "Optimizing retail strategy with IBM Retail Data Warehouse (RDW): A short review of data-driven approach to customer-centric decision-making," International Journal of Information Technology and Computer Science Applications, vol. 3, no. 1, pp. 33–39, Mar. 2025. doi:10.58776/ijitcsa.v3i1.178.
[10]. C. Kamusoko, "Machine Learning for Geospatial Data Analysis," Explainable Machine Learning for Geospatial Data Analysis, pp. 3–18, Nov. 2024. doi:10.1201/9781003398257-2.
[11]. R. Gong, "Transparent privacy is principled privacy," Harvard Data Science Review, no. Special Issue 2, Jun. 2022. doi:10.1162/99608f92.b5d3faaa.