SENTIMENT ANALYSIS OF INDONESIAN COMMUNITY TOWARDS ELECTRIC MOTORCYCLES ON TWITTER USING ORANGE DATA MINING

Zulham Sitorus[1], Maulian Saputra[2], Siti Nurhaliza Sofyan[3], Susilawati[4]
[1,2,3,4]*Master of Information Technology, Universitas Pembangunan Panca Budi*
Email: zulhamsitorus@dosen.pancabudi.ac.id

ABSTRACT

This study explores sentiment analysis of the Indonesian community towards electric motorcycles on Twitter using Orange Data Mining. In the context of the increasing popularity of electric vehicles, especially electric motorcycles, understanding public sentiment becomes crucial for various stakeholders. Twitter, as a leading social media platform, serves as a rich source of opinions and discussions on various topics, including electric motorcycles. This research utilizes Orange Data Mining with multilingual sentiment analysis techniques to analyze the sentiment of the Indonesian community regarding electric motorcycles. The results of sentiment analysis are visualized through box plots and scatter plots, aiming to classify Twitter users based on their emotional responses. The findings of this study provide valuable insights into the sentiment landscape surrounding electric motorcycles in Indonesia, benefiting policymakers, manufacturers, and marketers in understanding public perception and making informed decisions.

Keywords: Sentiment Analysis, Electric Motorcycles, Orange Data Mining, Multilingual Sentiment

ABSTRAK

Studi ini menggali analisis sentimen masyarakat Indonesia terhadap sepeda motor listrik di Twitter dengan menggunakan Orange Data Mining. Dalam konteks meningkatnya popularitas kendaraan listrik, khususnya sepeda motor listrik, pemahaman terhadap sentimen publik menjadi krusial bagi berbagai pihak terkait. Twitter, sebagai platform media sosial yang terkemuka, menjadi sumber pendapat dan diskusi yang melimpah tentang berbagai topik, termasuk sepeda motor listrik. Penelitian ini memanfaatkan Orange Data Mining dengan teknik multilingual sentiment untuk menganalisis sentimen masyarakat Indonesia terkait sepeda motor listrik. Hasil analisis sentimen divisualisasikan melalui box plot dan scatter plot, guna mengklasifikasikan pengguna Twitter berdasarkan tanggapan emosional mereka. Temuan dari penelitian ini memberikan wawasan berharga tentang lanskap sentimen seputar sepeda motor listrik di Indonesia, yang dapat memberikan manfaat bagi pembuat kebijakan, produsen, dan pemasar dalam memahami persepsi publik serta membuat keputusan yang berbasis informasi.

Kata Kunci: Analisis Sentimen, Sepeda Motor Listrik, Orange Data Mining, Multilingual Sentiment

1. INTRODUCTIONS
1.1. Background

Climate change has become a serious issue and a significant concern for several countries. The use of alternative energy is currently crucial in reducing carbon emissions and supporting environmentally friendly energy resilience for the present and future (Sudjoko 2021). Electric vehicles have emerged as a global trend as an alternative to fossil fuel-powered vehicles. Electric vehicles have become an increasingly important topic in the automotive industry in recent years (Alfarizi and Fitriani 2023). An electric vehicle is a type of vehicle that utilizes an electric motor as the primary source of power to drive the wheels and move (Agustian, Tukino and Nurapriani 2022). According to data from *Badan Pusat Statistik (BPS)* updated on February 29, 2024. The total number of motor vehicles in Indonesia in 2022 was 148.261.817 units. 85% of them are motorcycles, totalling 125.305.332 units. Motorcycles remain a popular mode of transportation for Indonesian society. In addition to being affordable, motorcycles are also easy to use for mobility purposes, making them a preferred choice as personal vehicles.

In line with the Sustainable Development Goals (SDGs) and one of Indonesia's priorities in strategic development, particularly related to the environment, such as clean energy and efforts to address climate change (Luqmania et al. 2022). Indonesia is committed to targeting a reduction in CO2 emissions by 29% - 41% by 2030. The use of electric motorcycles can be a solution to current environmental issues due to their use of environmentally friendly technology.

Sentiment analysis remains a part of opinion mining research (Minardi et al. 2023). Public sentiment towards electric motorcycles is a crucial factor influencing the adoption and acceptance of this technology. Analyzing public sentiment can provide valuable insights for manufacturers, policymakers, and researchers in understanding the perception and response of the community towards electric motorcycles.

In the digital era, social media serves as a platform for society to communicate and express opinions (Toy, Sari and Cholissodin 2021). Twitter has emerged as one of the primary platforms where people openly share their views, thoughts, and feelings. Twitter is also used by the general public to express their opinions on public topics and voice their complaints against a business or government institution (Andrian et al. 2023). Therefore, collecting and analyzing data from the Twitter social media platform can provide profound insights into public sentiment.

The use of the Orange data mining will facilitate this process by providing an intuitive and efficient environment for data analysis. The findings of this research are expected to provide valuable insights for various stakeholders in understanding public perceptions of electric motorcycles and promoting the adoption of environmentally friendly vehicles in Indonesia.

1.2. Literature Review

The development of electric vehicles in Indonesia has gained momentum again following the issuance of Presidential Regulation of the Republic of Indonesia Number 55 of 2019 concerning the acceleration program of battery-based electric motor vehicles (Battery Electric Vehicle). This progress can be seen in the production of several electric vehicles, particularly electric motorcycles, by local industries in 2020 (Raka Agung, Satya Kumara and Agus Surya Adi P 2022).

Sentiment analysis is the process of automatically extracting, processing, and understanding unstructured text data to gather information about the sentiment contained within a sentence, opinion, or statement (Arsi and Waluyo 2021). Sentiment analysis is the application of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from text sources (PUTRA, FADILAH and ENRI 2023). Sentiment analysis is one method used to classify objects into two categories: positive and negative sentiment (Utami 2022). The sentiment analysis approach aims to understand how people describe and express their feelings toward obesity conditions, whether positively, negatively, or neutrally (Mailo and Lazuardi 2021).

Data mining is the process of extracting valuable knowledge from a large and complex dataset (Sulianta 2024). One technique developed in data mining is how to explore existing data to build a model, and then use that model to recognize patterns in other data that are not stored in the database (Suryani et al. 2023). In the business world, sentiment analysis is used to apply the analysis of consumer comments on a product or service owned by a company or organization, the results of the analysis then become one of the bases for decision making (Hozairi, Anwari and Alim 2021).

Orange Data Mining is an open-source data mining application that has proven to assist researchers in analyzing their data (Rahmadini et al. 2023). Orange is an open-source software package released under the GPL and hosted on GitHub. Versions up to 3.0 included core components in C++ with wrappers in Python. Starting from version 3.0 and onwards, Orange utilizes common open-source Python libraries for scientific computing, such as numpy, scipy, and scikit-learn, while its graphical user interface operates within the Qt cross-platform framework. Default installation includes a variety of machine learning algorithms, preprocessing, and data visualization in 6 sets of widgets (data, transformation, visualization, model, evaluation, and unsupervised). Additional functionality is available as add-ons (text mining, image analysis, bioinformatics, etc.).

1.3. Research Methodology

The research method used is the experimental method by observing the variables of the studied objects. The experimental method aims to test the influence of one variable on another or to test the cause-and-effect relationship between variables. The research design steps are as follows:
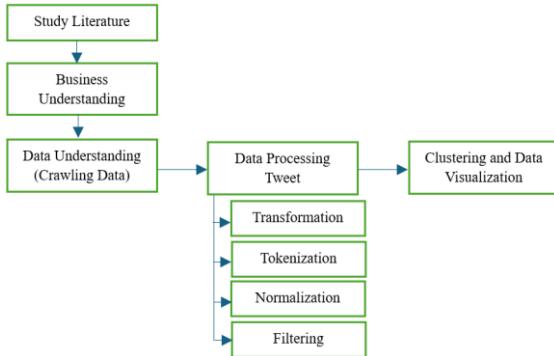

Figure 1. Research Methodology Flow

a.   Study Literature
     Previous research and literature studies define and cover aspects related to text mining.

b.   Business Understanding
     Analyzing the issues and facts surrounding electric motorcycles published by the government that are currently emerging in society.

c.   Data Understanding (Crawling Data)
     Utilizing comments provided by the community on the Twitter social media platform. Opinion retrieval is conducted using manual tweet collection methods, which involve entering keywords related to electric motorcycles into the Twitter search column.

d.   Data Processing Tweet
     Determining the class attribute and loading the dictionary involve matching base words with sentiment word dictionaries to determine the sentiment content (positive, neutral, negative). All tweet data is labeled according to classes, with three classes to be used in this study: positive class, negative class, and neutral class.

e.   Clustering and Data Visualization
     Clustering data text mining with Orange Data Mining involves visualizing Box Plot and Scatter Plot, which visualize the processed text mining data with the emotions of Twitter users.

2.   RESULT AND DISCUSSION
2.1. Research Scenario

The application of Orange Data Mining showcases the interface design of sentiment analysis widget integrated into the workflow, as illustrated in Figure 2 below:
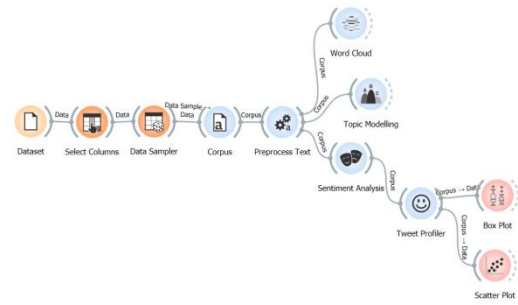

Figure 2. Sentiment Analysis Widget Data

The data crawled from the social media platform Twitter will be input and analyzed individually based on objects. Subsequently, it will be connected to the necessary widgets for research purposes, resulting in a widget design as shown in the figure above.

2.2. Data Crawling (Twitter)

In this study, the research data consists of comments from Indonesian society on Twitter regarding electric motorcycles from January 1, 2024, to May 29, 2024. The dataset for this research was obtained from a Python program written in Google Colab, as depicted in Figure 3 below.

```
1. filename = 'motor listrik.csv'
2. search_keyword = 'motor listrik
   lang:id until:2024-04-29 since:2024-
   01-01'
3. limit = 2000
5. !npx -y tweet-harvest@2.6.0 -o
   "{filename}" -s "{search_keyword}" -
   tab "LATEST" -l {limit} --token
   {twitter_auth_token}
```

Figure 3. Python Code for Crawling Data

The code above describes that data crawling is performed using the keyword 'motor listrik' with comments written in Indonesian language from 2024-01-01 to 2024-05-29. A total of 2000 tweets will be crawled, and the crawling results will be exported to a file named 'motor listrik.csv'.

From the data crawling results based on the above program, 2030 comments were obtained. Subsequently, the file is imported into Orange Data Mining as shown in Figure 4 below.
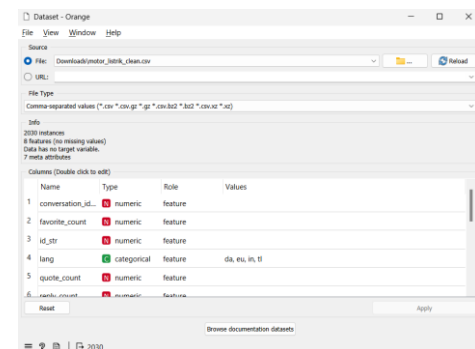

Figure 4. Import Dataset

The imported dataset will then have the necessary columns selected according to the requirements of sentiment analysis research. From the entire available dataset, a sampler will be taken comprising 70% of the total data, resulting in 1421 data samples to be processed by Orange Data Mining.

2.3. Preprocess Text

Before conducting text analysis, the text will first undergo preprocessing. This involves segmenting the text into smaller units (tokens), followed by transformation, tokenization, normalization, and filtering. Sequential steps in the analysis can be enabled or disabled within the Preprocess Text widget in Orange Data Mining. Figure 5 below shows the steps performed in the preprocess text widget in the Orange Data Mining application.
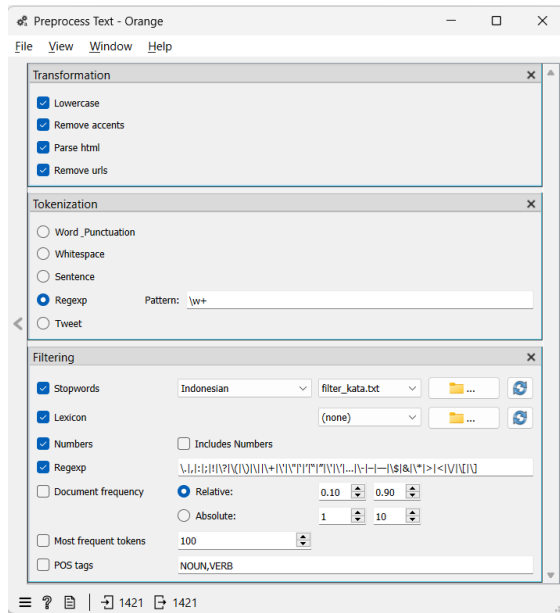

Figure 5. Preprocess Text

The steps carried out in the preprocess text in the Orange Data Mining application are as follows:

a. Transformation

The first step is transformation, which involves transforming the entire text into lowercase, removing accents contained within the text, identifying HTML tags, parsing HTML tags, and removing URLs from the text.

b. Tokenization

In this stage, sentences will be tokenized into words, preserving punctuation symbols.

c. Filtering

In this stage, a process of removing or preserving selected words will be conducted. During this process, words that are not relevant to sentiment analysis will be removed. All words to be removed have been written into a file named 'filter_kata.txt' using the stopword widget. Additionally, the lexicon widget is utilized to extract tokens from the lexicon dictionary. The number widget is employed to remove meaningless numbers, while the regexp

widget is used to eliminate tokens based on available regular expression patterns.

Once the preprocess text stage is completed, the text will be separated into individual words, and can observe the text distribution through a word cloud in Figure 6 below.
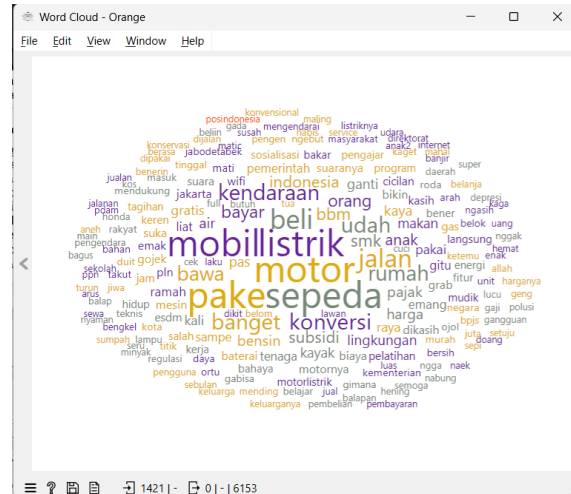

Figure 6. Word Cloud Electric Motorcycles

In the visualization above, a word cloud is displayed showing the frequency of the most commonly occurring words. The size of each word in the word cloud corresponds to its frequency of occurrence. The larger the word, the more frequent its occurrence. A word cloud serves as a visualization method to represent the results of text preprocessing. The visual settings and word variations enhance the attractiveness and comprehensibility of the visualization. The image depicts the preprocessing results of information previously comprised of comment lines on tweets about motorcycles from January to May 2024.

In addition to the word cloud, after preprocessing the text, there is a widget in Orange Data Mining to view topic modeling. Topic modeling in Orange Data Mining is a text analysis process aimed at identifying and extracting topics or patterns that emerge within a collection of documents. For more details, please refer to Figure 7 below.
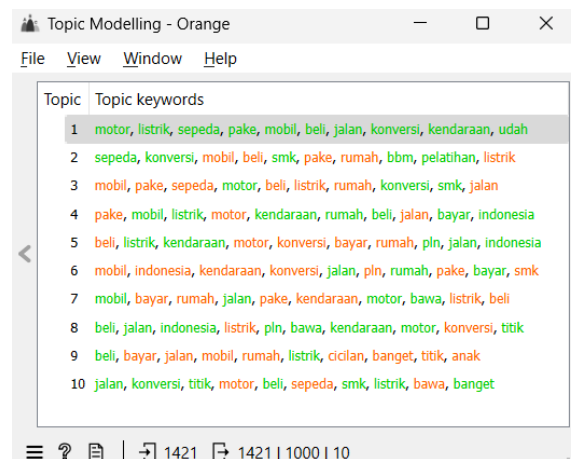

Figure 7. Topic Modeling Electric Motorcycles

2.4. Sentiment Analysis

The analysis process utilizes the MultiLingual Sentiment algorithm, which is capable of understanding opinions and viewpoints of users in various languages, in this case, using Bahasa Indonesia. According to (Olaleye et al. 2023), the MultiLingual sentiment analyzer has been shown to provide a significant level of invariance compared to traditional sentiment analysis systems, such as the Vader variant, thus enhancing the accuracy and diversity of sentiment analysis. As depicted in Figure 8 below.
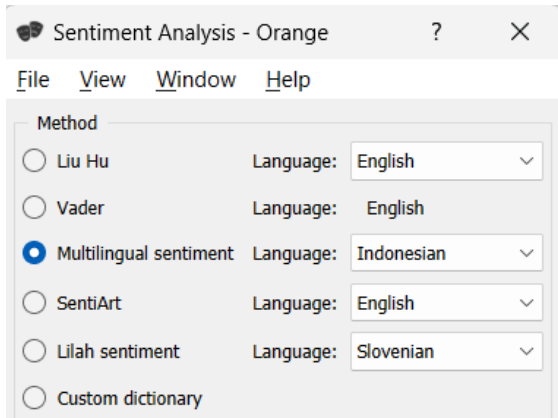

Figure 8. MultiLingual Sentiment Analysis

The MultiLingual sentiment approach enables a broader and more inclusive sentiment analysis, allowing text processing in various languages to understand the opinions, emotions, or sentiments contained within the text. Therefore, MultiLingual sentiment becomes crucial in the context of globalization and linguistic diversity in sentiment analysis and cross-cultural opinion understanding.

2.5. Tweet Profiler

Tweet Profiler is one of the features in the Orange Data Mining platform that enables this research to analyze sentiment from tweets or other text documents. By using Tweet Profiler, sentiment data can be retrieved from the available dataset through the server for each given tweet, and sentiment analysis can be conducted using various emotion classification methods provided, such as Ekman, Plutchik, and Profile of Mood States (POMS). Additionally, this feature allows for the utilization of specific attributes for analysis, such as content attributes, and performing emotion classification with multi-class options. Tweet Profiler is a valuable tool in text analysis and sentiment understanding in Orange Data Mining.

In this study, a dataset of 1421 tweets about motorcycles was utilized. The data was extracted using a widget from Orange Data Mining with Corpus and connected to Tweet Profiler using Ekman emotion. As depicted in Figure 9 below.
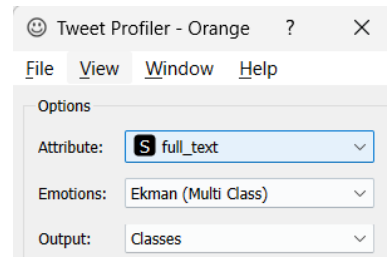

Figure 9. Tweet Profiler

2.6. Data Visualization

After performing tweet profiling in the Orange Data Mining widget, the next step is to connect the corpus to visualize the data and observe the results of sentiment analysis research using Box Plot and Scatter Plot.

In the Box Plot data visualization, a diagram is displayed showing the results of 6 emotions: joy, surprise, fear, disgust, sadness, and anger. From these 6 emotions, it can be observed that joy, surprise, and fear are the dominant emotional responses shown by Twitter users with the keyword 'motor listrik'. For further details, please refer to Figure 10 below.
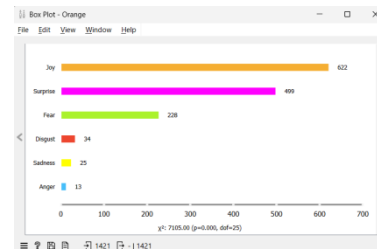

Figure 10. Box Plot Emotion

From the visualization above, it can be seen that the emotional responses exhibited by Twitter users with the keyword 'electric motorcycle' are as follows: the emotion 'joy' totals 622 or approximately 44%, the emotion 'surprise' totals 499 or approximately 35%, and the emotion 'fear' totals 228 or approximately 16%.

Data visualization can also be observed through Scatter Plots to visualize patterns or relationships between two variables, such as positive correlation, negative correlation, or no correlation at all. In a scatter plot, each point represents one observation, where one axis indicates the value of one variable and the other axis indicates the value of another variable. In this study, the variables used are emotion and sentiment variables.
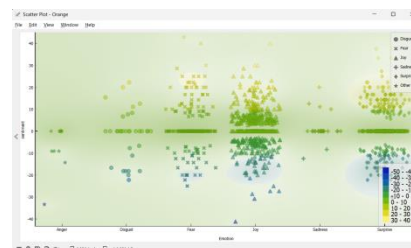

Figure 11. Scatter Plot

3. CONCLUSION

Based on the analysis outlined above, it can be concluded that:

a. The Multilingual sentiment algorithm can be utilized for sentiment analysis on social media platforms such as Twitter, to comprehend users opinions and viewpoints across various languages. It has been proven to provide a significant level of invariance compared to traditional sentiment analysis systems, thus enhancing the accuracy and diversity of sentiment analysis.

b. The analysis method utilizing tweet profiling enables the determination of the mood or emotions of Twitter users regarding trending topics in Indonesia, particularly concerning Electric Motorcycles.

c. By employing box plot and scatter plot visualizations, we can determine the classification of Twitter users with the visualization of emotions that have been input into each corpus within Orange Data Mining.

REFERENCES

Agustian, A., Tukino & F. Nurapriani (2022) PENERAPAN ANALISIS SENTIMEN DAN NAIVE BAYES TERHADAP OPINI PENGGUNAAN KENDARAAN LISTRIK DI TWITTER. *Jurnal TIKA (Teknik Informatika Aceh),* 7, 243-249.

Alfarizi, S. & E. Fitriani (2023) Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization. *Indonesian Journal on Software Engineering,* 9, 19-27.

Andrian, B. W., F. A. T. Tobing, I. Z. Pane & A. Kusnaidi (2023) Implementation Of Naïve Bayes Algorithm In Sentiment Analysis Of Twitter Social Media Users Regarding Their Interest To Pay The Tax. *International Journal of Science, Technology & Management,* 4, 1733-1742.

Arsi, P. & R. Waluyo (2021) Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* 8.

Hozairi, Anwari & S. Alim (2021) ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR, DECISION TREE AND NAIVE BAYES MODELS. *Jurnal Ilmiah NERO,* 6, 133-144.

Luqmania, D., A. Sunani, A. Septiani, F. A. Dwi Riyanto, M. B. Santoso & S. T. Raharjo (2022) MAS KLIMIS (MASYARAKAT PEDULI IKLIM YANG HARMONIS) KENDARAAN PT PJB UP GRESIK DALAM MEWUJUDKAN TUJUAN SUSTAINABLE DEVELOPMENT GOALS (SDGs). *Share : Social Work Journal,* 12, 45-56.

Mailo, F. F. & L. Lazuardi (2021) Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. *Journal of Information Systems for Public Health,* 6, 44 - 51.

Minardi, M., R. Lasepa, S. Riyadi, S. Ramadhan & D. D. Saputra (2023) Sentiment Analysis Terhadap Perspektif Warganet Atas Tragedi Kanjuruhan Malang di Twitter Menggunakan Naïve Bayes Classifier. *Jurnal Informatika,* 10, 45-53.

PUTRA, F. A. R., F. F. FADILAH & U. ENRI (2023) ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS IMDB MENGGUNAKAN METODE NAIVE BAYES. *Majalah Ilmiah UNIKOM,* 21, 87-94.

Rahmadini, E. E. L. Lubis, A. Priansyah, Yolanda & T. Meutia (2023) PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *JURNAL MAHASISWA AKUNTANSI SAMUDRA (JMAS),* 4, 223-235.

Raka Agung, I. G. A. P., I. N. Satya Kumara & I. P. Agus Surya Adi P (2022) Status Perkembangan Sepeda Listrik Dan Motor Listrik Di Indonesia. *Jurnal SPEKTRUM,* 8.

Sudjoko, C. (2021) Strategi Pemanfaatan Kendaraan Listrik Berkelanjutan Sebagai Solusi Untuk Mengurangi Emisi Karbon. *Jurnal Paradigma,* 2, 54-68.

Sulianta, F. 2024. *Buku Dasar Data Mining from A to Z.*

Suryani, M. F. Fayyad, D. T. Savra, V. Kurniawan & B. H. Estanto (2023) Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm. *Institute of Research and Publication Indonesia (IRPI),* 1, 1-9.

Toy, K. V. S., Y. A. Sari & I. Cholissodin (2021) Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 5, 5068-5074.

Utami, H. (2022) Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network. *Indonesian Journal of Applied Statistics,* 5, 31-38.