KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
VOL. 21 (2) (2024), 1-11, p-ISSN: 693-7554, e-ISSN:2654-3990

Sentiment Analysis Of Electric Car Product Trends In Indonesia Using BM25 And K-Nearest Neighbor Method

Ariel Alfaro[1*] , **Iffatul Mardiyah**[2]

[1.2] Information Systems Management Study Program, Department of Business Information Systems, Universitas Gunadarma, Jakarta

Abstract

The global and Indonesian shift towards electric vehicles (EVs) is driven by efforts to reduce emissions and promote sustainable energy. Social media, especially Twitter, functions as an important measuring tool regarding public sentiment towards electric vehicles in Indonesia, so that it can influence policy making. This research uses the BM25 and K-Nearest Neighbor (KNN) methods to analyze sentiment, which aims to improve EV adoption strategies. Conducted in 2023, this research applies data mining, specifically Knowledge Discovery and Data Mining (KDD), analyzing primary and secondary data descriptively and quantitatively starting with data collection from Twitter, followed by data crawling and initial text processing. Next, labeling, term frequency (TF) and inverse document frequency (IDF) calculations were carried out using the BM25 and KNN methods, with an Evaluation and Validation Diagram that visualized the process. The findings show that negative sentiment dominates at 48% (4800 data), followed by 34% (3400 data) neutral sentiment and 18% (1800 data) positive sentiment. The balanced distribution of sentiment highlights the diverse perceptions of society. BM25 and KNN pre-processing methods effectively reduce overfitting and underfitting, especially in negative and neutral sentiments. Accuracy testing without BM25 resulted in 58.6% to 60.25%, while integrating BM25 with KNN increased accuracy by 12.5% to 71% to 72.75%. Understanding sentiment provides a basis for decision making and policy development, as well as providing insight into public perceptions of electric vehicles in Indonesia. Implications include leveraging positive sentiment for marketing, adjusting strategies, refining pricing, addressing infrastructure and reliability issues, and collaborating with governments to increase adoption of electric vehicles in society.

Keywords : *trend sentiment analysis. electric car, Indonesia, BM25 and K-Nearest Neighbor*

1. Introduction

Currently, the use of alternative energy is very necessary to reduce carbon emissions and support the sustainability of green energy now and in the future [24]. The use of vehicles in Indonesia is currently increasing, in line with the increase in population and public economic opinion, both two-wheeled and four-wheeled vehicles. In supporting energy security, especially in the transportation sector, the government is currently supporting the development of electric vehicles to support the achievement of clean and environmentally friendly energy [1]. Currently, electric vehicles are an alternative solution that is being developed to support cleaner and environmentally friendly energy and reduce pollution and exhaust emissions due to the use of fuel oil in motorized vehicles [11]. The Indonesian government has also issued policies and incentives to encourage the adoption of electric vehicles to achieve sustainable development targets [17]. However, currently continuous energy consumption can cause various threats such as the energy crisis and pollution in the form of carbon dioxide emissions which are the cause of climate change. The more electric vehicles are developed, the more air pollution will be reduced, meaning that electric vehicles are one solution in anticipating the impact. energy crisis.

* **Corresponding author** . *E-mail address: alfaro.aril98 @ gmail .com*
Received: 20 December 2023, Accepted: 26 July 2024 and available online 31 July 2024
DOI: https://doi.org/10.33751/komputasi.v21i2.5260

Sentiment analysis to analyze responses and public views regarding the presence of electric car products by scraping tweet data on Twitter can facilitate [15] . via the K-Nearest Neighbor (KNN) method is applied in the case or specific context is a popular text classification method, so it can provide accurate results, data Text mining use method document Which is used cartwheels BM25, as a weighting and classification method , is an efficient method and provides accuracy in document sorting [19]. that, based on classification results research objects in the field Intelligent Computing uses BM25 and K-Nearest Neighbor. The research test process produces an average score test the best score on each test [2]. classify the gender of Twitter users by using the BM25 and K-Nearest Neighbor (KNN) methods can determine the accurate size of the results.

Based on the problems above, the results obtained show that the BM25 and KNN methods are used to classify Twitter users. To produce the highest accuracy, Indonesian language sports news is carried out and classified using BM25 and K-Nearest Neighbor (KNN) to get accurate results and is a simple but effective method for text classification [21]. The K-Nearest Neighbor (KNN) method is easily influenced by a single sample, because it is part of a learning method, K-Nearest Neighbor (KNN) is a classification method that is widely used in text data mining techniques [18]. In recent years, the electric vehicle trend has become a major highlight in the global automotive industry, including in Indonesia, where the transition to electric vehicles is considered an important step in reducing greenhouse gas emissions in overcoming the problem of climate change. The purpose of this writing is to find out, sentiment analysis of electric car product trends in Indonesia using BM25 and K-Nearest Neighbor (KNN) to classify sentiment related to electric car products in Indonesia so this research needs to be carried out.

2. Methods

The method for sentiment analysis research on electric car product trends in Indonesia uses BM25 and K-Nearest Neighbor (KNN) using *data mining* and *Knowledge Discovery and Data Mining* (KDD) processes [6]. As a set of processes, *data mining* in *text mining* refers to the process of examining and analyzing large amounts of unstructured text data using software that can identify concepts, patterns, themes, keywords and other attributes of the data [4]. The steps in using the data base are interactive, the user is directly involved as seen in Figure 1.

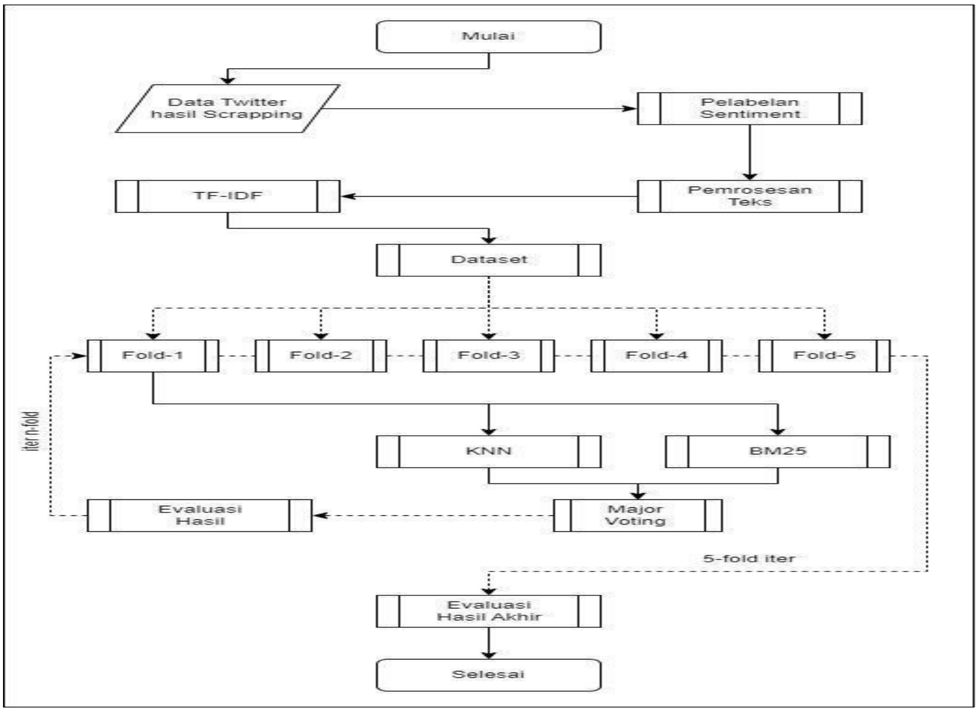

Figure 1. Flow Diagram Analysis Sentiment Product Trends Car Electricity

Figure 1 illustrates the complexity of the process in sentiment analysis for electric car products using the BM25 and KNN models. Process stages from the stage of collecting Twitter social media data, crawling the data media social Twitter next processing initial stage data known as Text preprocessing These include the normalization of HTML encoding, the presence of numbering in sentences, mentions of Twitter account users, the use of hashtags, site URLs, emoticons, punctuation, inconsistencies in text form (lowercase and uppercase writing), the presence of conjunctions, and words that are not has an important meaning in a sentence (called a stopword). Text reprocessing process which involves various stages this process aims to clean and tidy the data so that it can be used more effectively in subsequent data analysis. Data from the twitter dataset becomes more consistent by carrying out text preprocessing, free from noise, and ready to be analyzed in more depth. Then data Which has collected labeling is carried out, the labeling uses the help of external packaging, where the process uses the RoBERTa method apply preprocessing stage It is very important to ensure that the data used in this process completely ready for analysis. This process prepares data to calculate term frequency (TF), which measures how often a word appears in a document. Inverse document frequency (IDF) counting is performed to evaluate the importance of a word in a data set. The method used here involves the use of two main approaches, namely BM25 and k-Nearest Neighbor (KNN) [22].

The dots, dots, dots on the diagram reflect the results of interactions in the K-Fold method involving interactions from K-Fold1 to K-Fold5. Each literacy can show different data patterns. It is important to note that the literature that does not have a broken line shows clear data and the flow of sentiment analysis continues without interruption according to the pattern described in [7]. Thus, this diagram provides a visual depiction of the sentiment analysis process regarding electric car product trends [21]. The data collected was used to conduct experiments through tweets and comments on Twitter social media. The data comes from social media such as Twitter and online news, tweets totaling 10,000 lines. The results of data collection consist of the user's date and time columns and the content of the tweet, as seen in the example of table 1 of tweet data. Each tweet is given an identifier or index between 0 and 10,000. This identifier is used to track the continuity of Indonesian language tweets and process them further.

2.1. Data Analysis Stages

The data analysis stage is the data collection stage that will be carried out in the research, data is collected by crawling *Twitter social media data* using *the Twitter* API, Python programming language using the Tweepy library.

2.2. Crawling Data

Crawling data used uses *tweet data* taken from the social media network *Twitter* . Data was taken by crawling *Twitter data* using the *Twitter API* . To get the Twitter *API* namely by registering a developer's *Twitter account* , after that you get access to be able to retrieve the data needed to be processed in data processing.

2.3. Data Processing Begins

Text mining is mining carried out by computers to obtain something new, something previously unknown, or to retrieve implicitly implied information that automatically comes from information taken from various textual data sources by analyzing some or all of the text that is not structured, text mining tries to relate one text to another using rule certain. Usually text raw own property dimensions tall, the data contain noise, and structure the text is very poor, by because of that, *text mining* must through a number of stages which known with *pre-processing*, when processing the raw data. *Text preprocessing* is a very important step in the text data classification process. *Text preprocessing* is to remove noise, unify word forms and reduce word loudness, Organized research carried out to present information and solve problems and by using experimental methods and steps can divide into Some of the parts include data collection used to conduct experiments was collected through *tweets* and comments on social media [26]. *Twitter* for This data comes from social media such as Twitter and online news, *tweets* totaling 10,000 lines. The data collection results consist of date and time column users and *tweet data*, the *tweet data* is given an identifier or index between 0-10,000 tweets, as an identifier, the identifier can be used to track the continuity of Indonesian language *tweets and can be processed further, the Twitter* data seen in Table 1.

Table 1. Tweet Data

Indexes	Date	rawContent
0	2023-05-26 23:25:01+00:00	Katanya Punya APBN trilyunan, Hutang s/d Rp,7,8 T Infrastruktur...hebat (wow)..!? Jl.Tol, ribuan Km Proyek IKN KCJB Renc subsidi motor& mobil listrik… Trs ...bgmn ceritanya?
1	2023-05-2623:24:01+00:00	@mohmahfudmd kita hanya memperkaya dan memperkuat ekonomi negara lain. Orang pintar di indonesia tidak disupport utk membuat produk sendiri,hanya sebagai pekerja,seperti tidak ada gunanya. harus nya minimal sebahagian ada yg buatan sendiri. sudah bisa membuat mobil listrik,tidak diizinkan
2	2023-05-2623:22:43+0	@muchlis_ar Yg dipermasalahkan itu subsidi MOBIL LISTRIK

	0:00	
3	2023-05-2623:22:43+00:00	@luqmantaufiq_ @muchlis_ar Betul itu. Yg punya perusahaan motor/mobil listrik orang sekitat penguasa. Dibodohin kita ini
4	2023-05-2623:08:01+00:00	[POPULER OTOMOTIF] Beli Solar Wajib Pakai QR Code | Mobil Listrik Murah di INAPA 2023 | Bus Sleeper Premier PO Sempati Star https://t.co/nzS4UShtQf

2.4. Process Channel Testing Training And Testing

The training and testing flow, as seen in Figure 2, reflects a a highly structured and systematic process in training classification models.

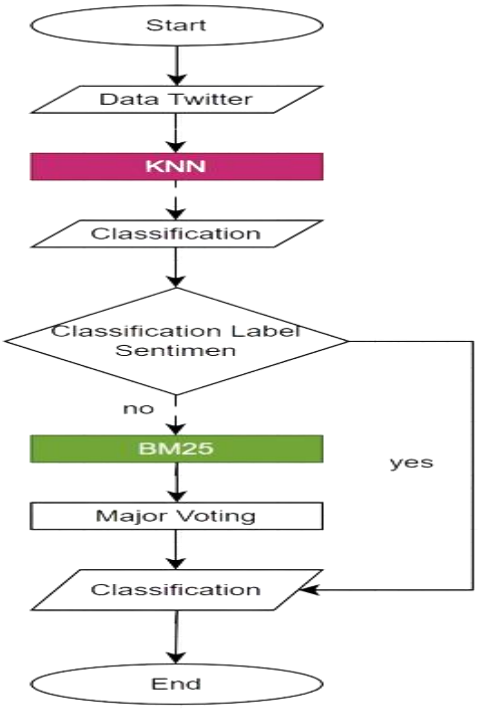

Figure 2. Channel Testing Training and Testing

The initial stage of the training and testing flow is usage K-Nearest Neighbors (KNN) algorithm to classify data. In In this context, the data that has been processed is assigned a label based on its characteristics. This stage is very important, because it becomes the basis for the steps the next step in the training flow. Once these labels are assigned, the next step is to match the classification results with the label that was generated in the previous 2nd stage. If results sentiment Which given by model KNN The same with label Which has is specified, then the next action is to immediately display the output that sentiment. However, if there is a discrepancy between the sentiment results and the existing labels, the next step is to involve an additional process called BM25. This BM25 process aims to more accurately classify complex or ambiguous data. This can happen when the KNN model cannot provide clear sentiment results. Next, data that has been classified using BM25 is further processed through document ranking using the BM25 method.

3. Results and Discussion

The results of the analysis show that the main focus in sentiment analysis in the use of the Twitter social media platform related to electric car products is effective and efficient. After evaluating and analyzing sentiment which is large and complex, especially data on public sentiment towards electric car products available on the social media platform Twitter has a real influence on people's needs as electric vehicle users.

Results of sentiment analysis and classification approach using the K-NN method (K-Nearest Neighbors), and the BM25 score calculation is proven to provide good accuracy results in the tweet classification process as seen in Table 2.

Table 2. Twitter dataset

Indexes	Date	rawContent
0	2023-05-26 23:25:24+00:00	Katanya\nPunya APBN trilyunan, \nHutang s/d Rp.7,8 T
1	2023-05-26 23:24:01+00:00	@mohmahfudmd kita hanya memperkaya dan memperk...
2	2023-05-26 23:22:43+00:00	@muchlis_ar Yg dipermasalahkan itu subsidi MOB...
3	2023-05-26 23:16:00+00:00	@luqmantaufiq_ @muchlis_ar Betul itu. Yg punya...
4	2023-05-26 23:08:01+00:00	[POPULER OTOMOTIF] Beli Solar Wajib Pakai QR C...

Table 2 shows that the results of the artistic analysis can be concluded that from the training data used in testing obtained by retrieving data from the Twitter social media platform using Google Colab Research software, it can provide accurate results. Where the total amount of tweet data that was collected was 10,000 tweets, which had been selected and selected to select 8,000 tweets which were used as training data (training data) 2000 as (testing data) to produce positive or negative data based on sentiment.

3.1. Text preprocessing results

The results of text preprocessing normalize HTML encoding, the presence of numbering in sentences, mentions of Twitter account users, the use of hashtags, site URLs, emoticons, punctuation, inconsistencies in text form (lowercase and uppercase writing), the presence of conjunctions, and words that does not have an important meaning in a sentence (called a stopword). The text preprocessing results depict a column labeled "Clean Content," which represents tweet data that has gone through a cleaning process using text preprocessing tools. This cleaning process is very important to ensure that the data used in sentiment analysis is free from interference or noise which can produce inaccurate results seen in Figure 3.

	Date	rawContent	Sentiment	Confidence Score	cleanContent
0	2023-05-26 23:25:24+00:00	He said\nHave a trillions in APBN,\nDebt up to IDR 7.8 T	positive	0.974231	APBN trillions in debt to SD RPT infrastructure hen...
1	2023-05-26 23:24:01+00:00	@mohmahfudmd we only enrich and strengthen...	negative	0.995624	Rich, strong economy, smart people in Indonesia...
2	2023-05-26 23:22:43+00:00	@muchlis_ar What is at issue is the MOB subsidy...	neutral	0.757518	Which subsidizes electric cars
3	2023-05-26 23:16:00+00:00	@luqmantaufiq_ @muchlis_ar That's right. Those who have...	negative	0.999349	People in the electric motorbike business are powerful...
4	2023-05-26 23:08:01+00:00	[POPULER AUTOMOTIVE] Buying Solar Must Use QR C...	neutral	0.768084	Popular automotive buying diesel must use QR code...

Figure 3 . Text Preprocessing Results

Figure 3 shows the composition of sentiment magnitudes that have been collected from tweet data that has been processed to better understand user responses and feelings towards a topic or entity discussed in the tweet. The results of the Sentiment Data Composition can help in revealing very valuable insights in understanding opinions and trends developing in social media conversations, negative, neutral and positive, as seen in Figure 3.

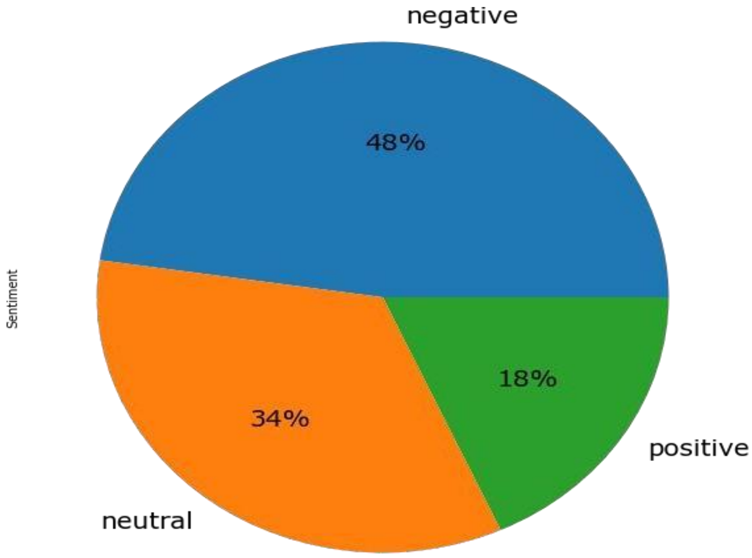

Figure 3 . Sentiment Data Composition Results

In Figure 3, the results of the analysis of the composition of sentiment data are shown which reveal a significant picture regarding the distribution of the data sentiment in a dataset. Where negative sentiment contributed 48% (4800 data), while positive sentiment contributed 18% (1800 data), and neutral sentiment contributed 34% (3400 data). The results of this analysis can reflect that the dataset has a fairly balanced sentiment distribution, which indicates that there is no significant sentiment dominance. In other words, no party dominates or has a sentiment that is much greater than the others. A positive indication in the context of sentiment analysis, as diversity in the dataset can enable more representative results and reduce bias in sentiment assessment. A balanced data composition such as sentiment analysis can be more reliable and informative for Twitter users to use.

3.2. Proses Menggunakan KNN dan BM25

The modeling process uses KNN and the BM25 process, the K-Nearest Neighbors (KNN) algorithm is used to classify data in a series of structured steps. The process begins with the use of Term Frequency-Inverse Document Frequency (TF-IDF) to create a text representation vector on the training data. Next, the K value is set to 5 to determine the number of nearest neighbors in the classification. The TF-IDF method is also applied to the test data, then the distance between the test data vector and the training data is calculated and sorted to find the nearest neighbors. Sentiment labels are determined based on the majority of positive, negative and neutral neighbors. If there is a difference between the actual and predicted sentiment, the BM25 algorithm is used to obtain 25 similar documents, and the sentiment prediction on these documents determines the majority sentiment label, thereby ensuring high accuracy. This process also includes tokenization steps, lower case reduction, and stop word removal for analysis consistency. In cases of classification uncertainty, voting is used to determine the dominant category. This systematic process integrates KNN, BM25, to produce accurate and reliable classification output.

3.3. Model Evaluation and Validation

Test results using the Best Match 25 (BM25) and K-Nearest Neighbor (KNN) method models provide a more in-depth view of algorithm performance, in the context of the desired [8]. The evaluation or interpretation process is a crucial stage in model evaluation and validation, where the results of the classification are tested to ensure the level of accuracy. The testing process is carried out by comparing classification results using accuracy parameters as a measure of the actual value. An important evaluation instrument, Confusion Matrix is used to calculate accuracy, providing a clear picture of the extent to which the model can differentiate different classes. In addition, dividing the data into training data and testing data via the K-Fold method (in this context, with 5 scenarios).

3.4. Confusion Matrix

Analysis of training and testing results, the resulting decisions need to be evaluated carefully and one of the evaluation methods used is the confusion matrix, which provides an assessment of classification performance based on the accuracy of predictions for actual objects. Confusion matrix provides very valuable information by separating the results of Accuracy, Recall F-1 Score.

Confusion matrices make it possible to gain a deep understanding of the effectiveness of classification models, identify possible error patterns, and optimize the overall performance of the classification system. By detailing actual and predicted information, the confusion matrix becomes a very useful tool in measuring and improving the accuracy and reliability of a classification system.

3.5. Confusion Matrix Training results are not with the BM25 Model

Exploration of the model without using BM25 found results that provide an in-depth view of the model's ability to perform classification in the training and testing phases. So that it can provide a clear picture of accuracy, recall and F-1 score at each iteration in the matrix (confusion matrix), Monfusion Matrix Training is not with the BM25 Model, as seen in the table. 3

Table 3. Confusion Matrix Training is not with Model BM25

.

	Training		
Description	**Accuracy**	**Recall**	**F-1 Score**
Iteration 1	0.7505	0.7505	0.7519
Iteration 2	0.7444	0.7444	0.7433
Iteration 3	0.7594	0.7594	0.7585
Iteration 4	0.7511	0.7511	0.7526
Iteration 5	0.7494	0.7494	0.7506

Source: processed data (2023)

On Table. 3 , show that, phase training in model shows variability in performance at each iteration, although with a relatively stable level of accuracy. Recall and F-1 score results tend to be in line with accuracy levels and Iteration 3 stands out with high levels of accuracy, recall and F-1 scores, indicating the model's capacity to identify complex patterns in the training data. However, the fluctuations observed in other iterations require special attention to the stability of the performance of the .onfusion Matrix Training model not with the BM25 Model, as can be seen in Table 5.

Table. 4 . Confusion Matrix Training is not with Model BM25

	Testing		
Description	**Accuracy**	**Recall**	**F-1 Score**
Iteration 1	0.6025	0.6025	0.6008
Iteration 2	0.614	0.614	0.6175
Iteration 3	0.586	0.586	0.5879
Iteration 4	0.6025	0.6025	0.6014
Iteration 5	0.5965	0.5965	0.6002

Source: processed data (2023)

In Table 4 , it shows that the model faces greater challenges and a lower level of accuracy in testing compared to training, indicating the potential for overfitting or difficulty in generalizing the patterns that have been learned. The 2nd iteration showed significant improvements, highlighting the model's ability to adapt to never-before-seen data. The overall model without using BM25 shows varying results, with a tendency for better performance on training data than on test data. The increase in performance from several iterations shows that the model can utilize learning from training data well. However, there are certain challenges and they lie in the ability to process data effectively in applying new data learning. However, further analysis is needed to understand the factors that influence the differences in performance between the training and testing phases, as well as to identify potential data improvements to increase the generality of the model that has been carried out effectively.

3.6. Confusion Matrix Training with BM25 and K-NearestNeighbor Models

The purpose of the confusion matrix is to evaluate model performance using the BM25 and K-Nearest Neighbor (k-NN) algorithms in the training and testing phases. In-depth analysis is carried out to interpret the results of the confusion matrix, which includes accuracy, recall and F-1 score for each iteration, Confusion Matrix Training with the BM25+ K-Nearest Neighbor Model, shown in Table. 5 .

Table 5. Confusion Matrix Training with the BM25+ K-Nearest Neighbor Model

	Training		
Description	**Accuracy**	**Recall**	**F-1 Score**
Iteration 1	0.8185	0.8185	0.8181
Iteration 2	0.8167	0.8167	0.8137
Iteration 3	0.8196	0.8196	0.8174
Iteration 4	0.82	0.82	0.8199
Iteration 5	0.8089	0.8089	0.8093

Data source in if (2023)

Table 5 shows that the training phase of the BM25 k-NN model has high consistency in performance with accuracy values ranging from 0.8089 to 0.82 in 5 (five) different iterations. Apart from that, the recall value and F-1 score also reached a very good level, ranging from 0.8089- 0.8200. Meanwhile, the significant increase in the 3rd iteration indicates the model's ability to understand complex patterns in the training data. Confusion Matrix Testing with the BM25 k-Nearest Neighbor Model can be seen in Table 6.

Table. 6. Confusion Matrix Training with the BM25+ K-Nearest Neighbor Model

	Training		
Description	**Accuracy**	**Recall**	**F-1 Score**
Iteration 1	0.71	0.71	0.7047
Iteration 2	0.726	0.726	0.7283
Iteration 3	0.726	0.726	0.7251
Iteration 4	0.7275	0.7275	0.719
Iteration 5	0.7265	0.7265	0.7195

Data source in if (2023)

Table.7, shows that, when dealing with test data on. The Confusion Matrix model still maintains solid performance. Even though there are slight fluctuations between iterations, the accuracy, recall and F-1 scores remain in a satisfactory range. While the 2nd Iteration shows a marked improvement in the F-1 score, highlighting the model's ability to generalize on unknown data. The Confusion Matrix for test data provides further insight into the performance of the model and its ability to recognize both positive and negative classes reflected in high recall values, while the accuracy and F-1 scores provide a holistic picture of its predictive ability, a holistic picture of the results of using the method and enable it and predict accuracy comparison results as seen in Figure 4.

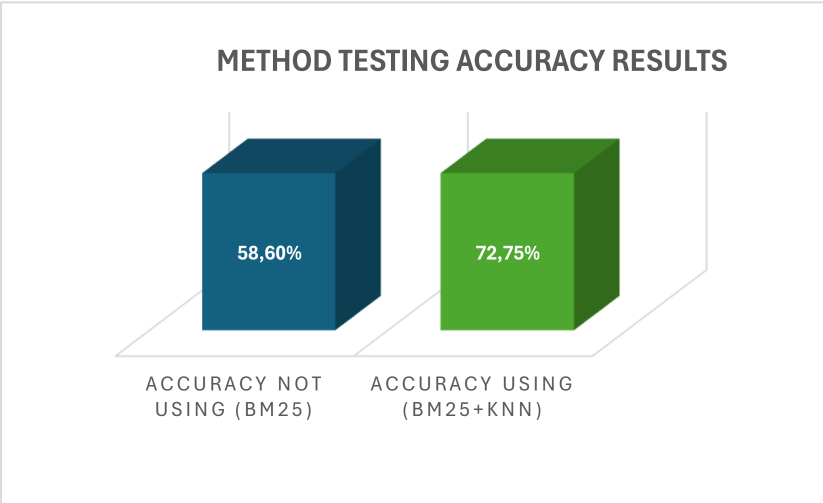

Figure 4. Accuracy comparison results

Figure 4 shows that the BM25 model with k-NN shows impressive performance in both the training and testing phases. The accuracy of not using the testing method and also not using the Best Match 25 (BM25) process reached 58.6-60.25% and the accuracy using the Best Match 25 (BM25) and K-Nearest

Neighbor (KNN) process was 71-72 .75%. From these results, accuracy increases by 12.5%, this increase in accuracy results in a better prediction level.

A high level of consistency in testing shows the model's ability to overcome the complexity of the data and provide consistently solid results. Provides confidence that the model can reliably classify never-before-seen data. Further attention to matrix interpretation can help understand specific aspects where the model may need improvement for more realistic situations. The sentiment distribution results for electric car brands are seen in Figure 5 and the sentiment iteration wordcloud results are seen in Figure 6 .

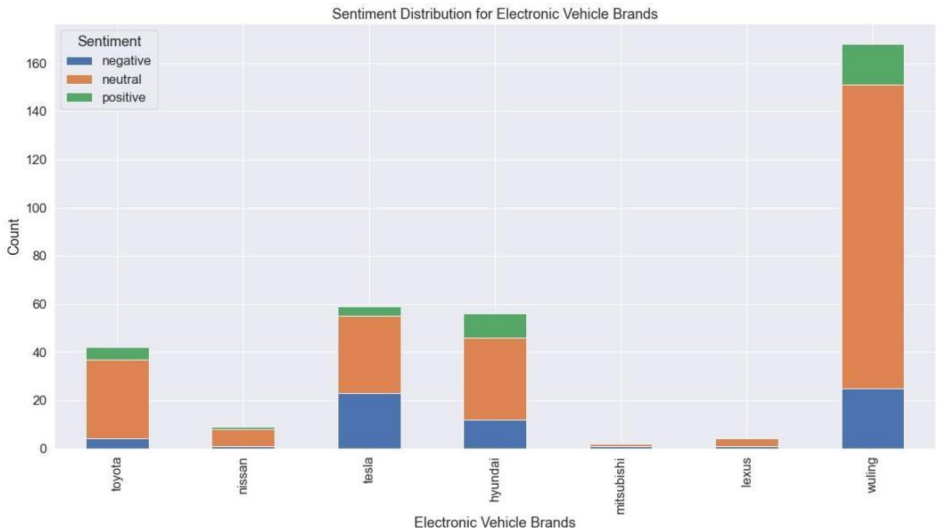

Figure 5 . Electric Car Brands Sentiment Distribution Results

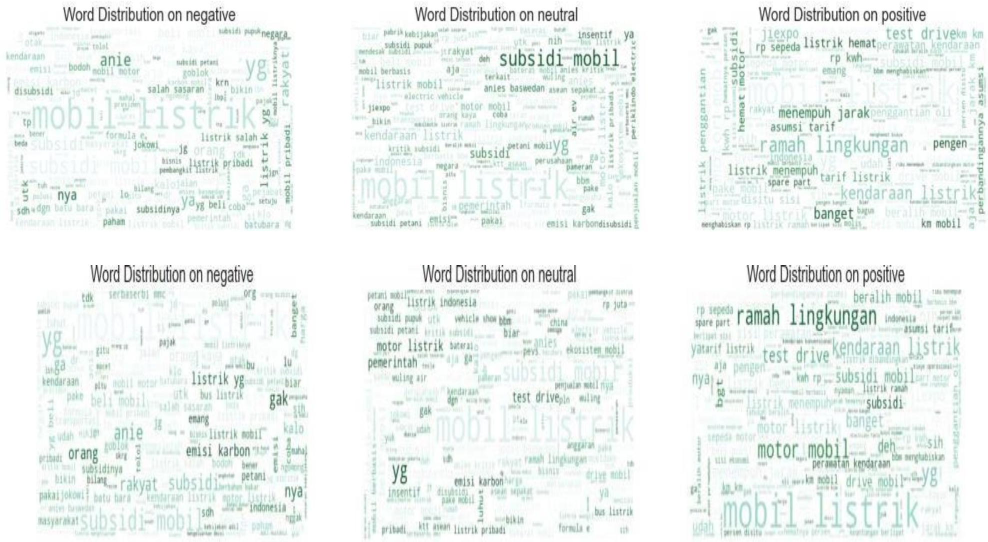

Figure 6 . Wordcloud results of Sentiment iteration

4. Conclusion

Based on the research results, it can be concluded that testing social media user data, especially Twitter, identified trends and sentiments related to electric car products on the Twitter social media platform. The results provide a comprehensive picture of how people respond to electric vehicles, that negative sentiment contributes 48% while positive sentiment contributes 18% and neutral sentiment accounts for 34%. Reflecting that, the dataset has a fairly balanced sentiment distribution, and indicates there is no significant sentiment dominance. That means, nothing the party that dominates or has a much greater magnitude of sentiment. The use of the BM25 method in pre-processing tweets in the training and testing phases has a significant impact in reducing the level of overfitting and underfitting, especially for negative and neutral sentiment. Tweet classification to be more accurate and optimal The BM25 model with k-NN shows impressive performance in both the training and testing phases. Accuracy not using the testing method not using the Best Match 25 (BM25) process reached 58.6 - 60.25% and accuracy using the Best Match 25 (BM25) and K-Nearest Neighbor (KNN) processes 71 - 72.75%. These results increase accuracy by 12.5%, with this increase in accuracy the prediction level is good.

References

[1] Anwar, TM, Riandhita and d., Permana, A. 2023. Analysis of Indonesian public sentiment towards electric vehicle products using VADER. Journal of Informatics Engineering and Information Systems, 10(1), 783-792

[2] Ayudhitama, Annisa Putri, and Utomo Pujianto. 2020. "Analysis of 4 Algorithms in liver classification using rapidminer." Polinema Informatics Journal 6(2): 1–9.

[3] Binawan and D. Hendra. 2019. Classification of thesis abstract documents based on research focus in the field of intelligent computing using Bm25 and K- Nearest Neighbor. Bachelor's thesis, Brawijaya University. Skipsi p.1-69, November 2022, SKR/FILKOM/2019/121/051902291, (internet in aplot, 15 December 2023) available http://repository.ub.ac.id/id/eprint/169247/

[4] Berry, W. and Kogan, J. (2013) 'Michael W. Berry and Jacob Kogan (eds.): Text mining: applications and theory', Information Retrieval, 14(2), pp. 208– 211. doi: 10.1007/s10791-010-9153-5.

[5] Binawan, D. Hendra, Indriati, I., and Adikara, PP (2019). Classification of Thesis Abstract Documents Based on Research Focus in the Field of Intelligent Computing Using BM25 and K-Nearest Neighbor. Journal of Information Technology and Computer Science Development, 3(3), 2640– 2645.

[6] Han, et al., 2022. Data mining: concepts and techniques Florin, G., 2011. Data Mining Concepts, Models and Techniques. Volume 12 ed. [online] Poland: Springer. Available at: <Springer.com>.

[7] Jena, R., 2019. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. Industrial Marketing Management, 90(2), 605-616

[8] Johar T, Asahar. Yanosma, D. Anggriani, K. (2016). Implementation of the K-Nearest Neighbor (KNN) and Simple Additive Weighting (SAW) Methods in Making Selection Decisions for Accepting Paskibraka Members. Bengkulu: Informatics Engineering Study Program, Faculty of Engineering, Bengkulu University. ISSN: 2355-5920. Jena, R., 2019. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. Industrial Marketing Management, 90(2), 605-616.

[9] Ministry of Environment and Forestry. (2019). 2018 Statistics of the Directorate General of Climate Change Control. Directorate General of Change Control (internet on plot date, 15 December 2023) available Climate. http://ditjenppi.menlhk.go.id

[10] Ministry of Environment and Forestry. (2020). Strategic Plan for 2020-2024. Directorate General of Pollution and Environmental Damage Control. (internet on plot, 15 December 2023) available https://www.menlhk.go.id/. JHan, et al., 2022. Data mining: concepts and techniques Florin, G., 2011. Data Mining Concept, Models and Techniques. Volume 12 ed. [online] Poland: Springer. Available at: <Springer.com>.

[11] Indah., Rumiasih., Carlos., Firmansyah. (2019). Analysis of Determining Battery Capacity and Charging in Electric Cars. Elektra Journal,4(2),29-37

[12] (IEA) International Energy Agency. (2021). Global EV Outlook 2021. "Accelerating ambitions despite the pandemic". Published April 2021. (internet in plot date, 11 December 2023) from https://www.iea.org/.J

[13] (IEA) International Energy Agency. (2019). Global EV Outlook 2019. "China's CATL staInternational Energy Agency. (2019). Global EV Outlook 2019. "China's CATL begins mass production of high-nickel batteries: chairman". (internet in plot date, 15 December 2023) available https://www.iea.org/.

[14] Isman, A., & Dagdeviren, E. 2018. "Diffusion of Twitter in Turkey". The Turkish Online Journal of Educational Technology, 17(4), 1-7

[15] Laurensz, B., and E. Sediyono. 2021. Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic. National Journal of Electrical Engineering and Information Technology, 10(2), 118-123.

[16] Lubis, AR, Lubis, M., & Khowarizmi, A.-. (2020). Optimization of distance formula K-Nearest Neighbor method. Bulletin of Electrical Engineering and Informatics, 9(1), 326–338. https://doi.org/10.11591/eei.v9i1.1464

[17] Nisa Lily Choirun and Anita Susanti. 2018. Strategy for implementing electric cars in Surabaya as smart mobility. Journal; applied transportation publication media 1(2):213-225 https://journal.unesa.ac.id/index.php/mitran/article/view/26193

[18] Mentari, ND, Fauzi, MA and Muflikhah, L. (2018) 'Sentiment Analysis of the 2013 Curriculum on Twitter Social Media Using the k-Nearest Neighbor Method and Feature Selection Query Expansion Ranking', Development of Information Technology and Computer Science, 2(8), pp. 2739– 2743.

[19] Pardede, J., Husada, MG and Riansyah, R., 2018. Implementation and Comparison of the Okapi BM25 and PLSA Methods in Information Retrieval Applications.

[20] Pratama, Y., Murdiansyah, T., D., Lhaksmana, M., K. 2023. Analysis of electric vehicle sentiment on Twitter social media using the logistic regression algorithm and principal component analysis. Budidarma Media Informatics Journal, 7(1), 529-535.

[21] Purnamawati, A., Winnarto, M., N., Mailasari., P., Y. 2023. Sentiment analysis of the TikTok application using the BM25 Method and Improved K-NN Chi-Square Features. Journal of Computational and Informatics, 7(1), 97-105.

[22] Sakariana, MID, Indriati, I., & Dewi, C. (2020). Sentiment Analysis of the Move of the Indonesian Capital City Using BM25 Term Weighting and K-Nearest Neighbor Weighted Neighbor Classification. Journal of Information Technology and Computer Science Development, 4(3), 748–755.

[23] Santoso, A., Nugroho, A., and Sunge, AS 2022. Sentiment Analysis About Electric Cars Using Support Vector Machine Methods and Feature Selection Particle Swarm Optimization. Journal of Practical Computer Science, 2(1), 24-31. https://doi.org/10.37366/jpcs.v2i1.1084.

[24] Sudjoko Cakrawati. 2020. Strategy for using sustainable electric vehicles as a solution to reduce carbon emissions Paradigm Journal: Multidisciplinary Journal of Indonesian Postgraduate Students, 2(2):54-68 https://doi.org/10.22146/jpmmpi.v2i2.70354

[25] Suharno, CF, et al., 2017. Classification of Indonesian Text in Online Samsat Complaint Documents Using K-Nearest Neighbors and Chi-Square Methods. Journal of Information Technology and Computer Science Development 1(10): 1000-1007.

[26] Zakia, AS and Indriati, M., 2020. Gender Classification of Twitter Users using the BM25 and K-Nearest Neighbor (KNN) Methods. Journal of Information Technology and Computer Science Development e-4(10): 3331-