

Analysis of Customer Sentiment towards Wardah Beauty Products on Instagram Using K-Nearest Neighbors (KNN) and Naïve Bayes Algorithms

Aghniya Aulia Nisa^{1✉}, Nur Hafidzah Safitri², Lilis Stianingsih³, Fiqih Hana Saputri⁴

^{1,2,3,4} Teknik Informatika, FTIK, Institut Teknologi dan Bisnis Bina Sarana Global, Indonesia

✉ **Corresponding Author** : Aghniya Aulia Nisa (e-mail: aghniyanisa91@gmail.com)

Article Information	ABSTRACT
Article History Received : March 21, 2025 Revised : March 27, 2025 Published : April 12, 2025	The beauty industry in Indonesia continues to grow rapidly, with Wardah emerging as a leading local brand that garners numerous customer reviews on Instagram. Understanding these customer opinions is crucial, as they significantly influence brand perception and purchasing decisions. This study aims to compare the performance of K-Nearest Neighbors (KNN) and Naïve Bayes algorithms in classifying customer sentiment toward Wardah's products on Instagram. The research follows the CRISP-DM methodology, involving data collection through web scraping, data preprocessing, and classification using both algorithms. The results show that KNN achieves higher accuracy (78.23%) than Naïve Bayes (74.56%), making it more effective in sentiment classification. KNN is advantageous in identifying distance-based patterns, while Naïve Bayes offers faster processing and easier implementation. These findings suggest that KNN is more suitable for accurate sentiment prediction. Wardah can implement a KNN-based sentiment monitoring system to track public feedback on product launches and campaigns, thereby enhancing marketing strategies and customer engagement.
<i>K-Nearest Neighbors (KNN); Naïve Bayes; Sentiment analysis; Social Media; Beauty.</i>	

INTRODUCTION

In recent years, Indonesia's beauty industry has grown rapidly due to the emergence of local brands that can compete with products from the international market (Anggraini et al., 2024; Article et al., 2025; Arvy & Wahyuni, 2023). Wardah, which is famous for its halal beauty products, is one of the most well-known local brands (Tazlia et al., 2023). Wardah utilizes social media, especially Instagram, as the main channel to promote its products and interact with customers (Azhari & Sazali, 2023; Pranasa et al., 2023). Wardah's presence on Instagram not only acts as a promotional medium but also allows customers to provide reviews and feedback regarding their products (Industri & Tiktok, 2025). With the increase in Instagram users, analyzing customer feedback is an important aspect of improving brand image and understanding market revenue (Sari et al., 2021).

Customer reviews and comments collected by Instagram can provide valuable insights into product satisfaction and perception (Lesmana, 2023). Data analytics is a very important way to gain further understanding of the data available in situations like this (Köster et al., 2021; Sarker, 2021). This is because it can process large amounts of data automatically. One algorithm that is often used in sentiment classification is K-Nearest Neighbors (KNN) (Buhori et al., 2024). It is often used to categorize customer opinions on social media (Pratmanto & Imaniawan, 2023). (Anam et al., 2021) previous research shows that while the working principle of the algorithm is classified as simple, KNN can produce good accuracy when classifying sentiment. Previous research (Isnain et al., 2021) found that the KNN algorithm can be used to analyze user interactions in online learning with an optimal accuracy rate of 84.65% for a given K value.

On the other hand, several studies have shown that the Naïve Bayes algorithm is often used as an alternative to sentiment analysis (Wankhade et al., 2022). A study by (Hermawan et al., 2025) showed that Naïve Bayes can provide competitive classification results, especially when applied to data records with balanced data distribution. This research aims to recognize and classify customer reviews into two main categories, namely positive and negative sentiment. By comparing the two algorithms used, this research seeks to determine the more optimal method in sentiment analysis, based on its accuracy rate, processing efficiency, and ability to handle imbalanced data. The results of this study are expected to improve our understanding of customers' opinions on Wardah products and help us create better marketing strategies. Readers who want to choose an algorithm for sentiment analysis and machine learning are advised to consider features such as dataset characteristics, preprocessing methods, and data distribution before choosing an algorithm. Wardah can utilize the results of this study as a reference in designing data-driven marketing strategies, especially in analyzing customer opinions on a product.

This research still has limitations in the scope of the data used, as it only focuses on customer reviews on Instagram within a certain period. Therefore, future researchers can expand the scope of the study by combining data from other social media platforms, such as Twitter, TikTok, or e-commerce. In addition, further exploration of more complex machine learning methods, such as Support Vector Machine (SVM) or Deep Learning, can be done to improve the accuracy of sentiment classification models.

RESEARCH METHODS

Applying the industry-standard CRISP-DM process to data mining (CRISP-DM) is a standard method for developing data mining projects (Kurniawan & Yasir, 2022). Developed in the late 1990s, the methodology offers a systematic framework that includes six main stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Ghandi & Ramadhan, 2024). This methodology plays a role in helping to understand problems and implement solutions, both in research and business contexts. In addition, this approach also supports structured planning and execution of data mining projects (Adhi Putra, 2021; Muttaqin, Wahyu Wijaya Widiyanto et al., 2023).

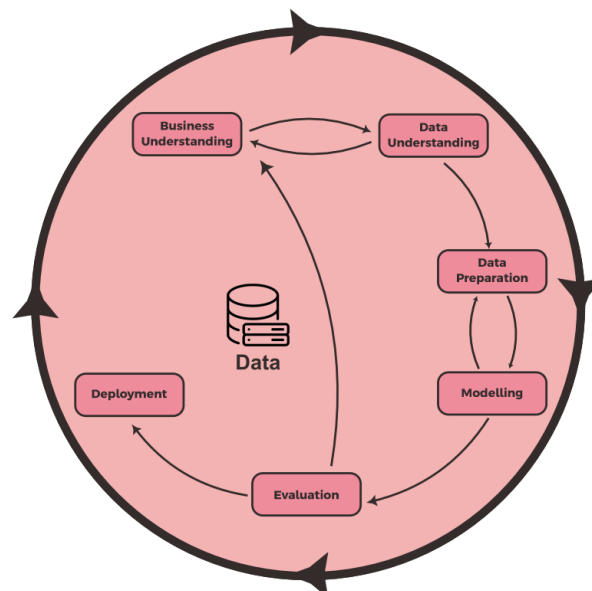


Figure 1. CRISP-DM

Business Understanding

This stage focuses on understanding the objectives and needs of the project from a business perspective. In this research, the main objective is to analyze customer sentiment towards Wardah beauty products based on comments found on Instagram. By understanding customer sentiment,

Wardah can gain better insight into consumer perceptions, which can be used to improve marketing strategies and product development.

This research uses the approach of applying and comparing two machine learning algorithms, namely K-Nearest Neighbors (KNN) and Naïve Bayes. Both algorithms are evaluated based on their accuracy and ability to handle data imbalance. The results of this research are expected to provide recommendations on which algorithm is more effective in classifying Wardah customer sentiment.

Data Understanding

At this stage, an initial exploration of the dataset used in the research is carried out. The dataset used contains a total of 2,033 customer review data on Wardah beauty products obtained from the Instagram platform. This data consists of several important attributes used in sentiment analysis, namely:

- Username : The name of the user who left the comment.
- Date : The time the review was given.
- Age : Age of the user at the time of the review.
- Gender : The gender category of the user (Male or Female).
- Product : Wardah products that get reviews from users.
- Comment : Customer reviews or comments on the product.
- Experience : Additional opinions given by users regarding their impressions after using the product.
- Sentiment : Sentiment labels are categorized into Positive, Neutral, and Negative.
- Likes : The number of likes on user-given comments.

To ensure the quality of the data before further analysis, the top five data are displayed. Here is an example of the first five data in the dataset:

```
===== DATA UNDERSTANDING =====
Jumlah Data: (2033, 10)
5 Data Teratas:
```

	Username	Tanggal	Usia	Jenis Kelamin	Produk
0	Aziz Satriyo	04/03/2025	20	Laki-laki	Sunscreen
1	Lilis	09/03/2025	31	Perempuan	Lipstrik, bedak, cusion
2	fahri athalla	09/03/2025	17	Laki-laki	facial wash, sunscreen
3	itsme.setia_	09/03/2025	21	Perempuan	Lipbalm
4	pirja	09/03/2025	17	Laki-laki	wardah sunscreen ijo

	comment
0	Sangat mengesankan
1	Sangat bagus
2	packagingnya menarik
3	Cocok
4	bagus banget bikin glowing, wanginya enak

	pengalaman	sentiment	likes
0	Banyak kelebihan dan sedikit kekurangan	Positif	1500
1	Keren	Positif	20
2	kelebihannya berlebihan kekurangan gada	Positif	200
3	Harga affordable, isi cepat habis karena dipak...	Positif	7
4	wardah sunscreen ijo\ankelebihan: murah, ga bik...	Positif	436

Figure 2. Data Understanding

From the table above, it can be seen that most comments have Positive sentiments with varying numbers of likes. This shows that Wardah products have a fairly high level of customer satisfaction. However, for further analysis, it is necessary to preprocess the data to ensure the quality of the data used in the Machine Learning model.

Data Preparation

After understanding the structure and characteristics of the data, the next step is Data Preparation, which is the data preparation stage before being used in the machine learning modeling process. This process includes data cleaning, removal of unnecessary information, and preprocessing so that the data is in optimal condition before further analysis.

a. Data Cleaning

At the initial stage, data containing NaN (Not a Number) values or incomplete data is removed.

b. Data Preprocessing

Next, preprocessing of customer comment text is carried out. Some of the preprocessing steps applied in this research include:

- Case Folding : To make the analysis process more consistent, the entire text is divided into lowercase letters.
- Tokenization : Separates the text into individual words.
- Stopword Removal : Removing words that do not have a major contribution to sentiment analysis.
- Stemming : Converting words into their base form to simplify word variations.

Examples of data after preprocessing are shown in the following table:

```
===== DATA PREPARATION =====
Jumlah Data Setelah Menghapus NaN: (2033, 10)
Contoh Data Setelah Preprocessing:
```

	comment	sentiment
0	Sangat mengesankan	Positif
1	Sangat bagus	Positif
2	packagingnya menarik	Positif
3	Cocok	Positif
4	bagus banget bikin glowing, wanginya enak	Positif

Figure 3. Data Preparation

c. Data Distribution After Filtering

After preprocessing and filtering, the sentiment distribution in the dataset is as follows:

```
Distribusi Data Setelah Filtering:
```

sentiment	
Positif	1506
Negatif	494
Netral	33

Name: count, dtype: int64

Figure 4. Distribution After Filtering

From the sentiment distribution results, it can be seen that the majority of customer reviews of Wardah products are positive, while the number of comments with negative and neutral sentiments is less. The imbalance in data distribution needs to be considered in the modeling stage so that the classification model does not experience bias in determining the prediction results.

Modeling

In the modeling stage, two algorithms are used to classify sentiment, namely K-Nearest Neighbors (KNN) and Naïve Bayes. These two algorithms were chosen because they have different characteristics in the classification approach, allowing a comparison to determine a more optimal model. This model is tasked with categorizing data into positive, negative, or neutral sentiments based on a certain K value parameter. The modeling stages in this research are as follows:

a. K-Nearest Neighbors (KNN)

The KNN algorithm works with the principle of finding several nearest neighbors (K) of the data to be classified based on a certain distance, generally using the Euclidean Distance formula:

$$d = \sum_{i=0}^n (x_i + y_i)^2 \quad (1)$$

Determines the majority class of the K nearest neighbors as the classification result.

b. Naïve Bayes

Besides KNN, this research also implements the Naïve Bayes algorithm, which is one of the probability-based classification methods. This algorithm works based on Bayes' Theorem with the assumption that each feature is independent of each other:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (2)$$

Where:

- $P(C | X)$ is the probability of a sentiment class C given a feature X
- $P(X | C)$ is the probability of getting feature X in class C
- $P(C)$ is the initial probability of class C
- $P(X)$ is the probability of feature X

Naïve Bayes is often used in text analysis because of its ability to handle data with a large number of features. In this research, the features used are words in the comments after going through preprocessing stages, such as tokenization, stopwords removal, and stemming.

Evaluation

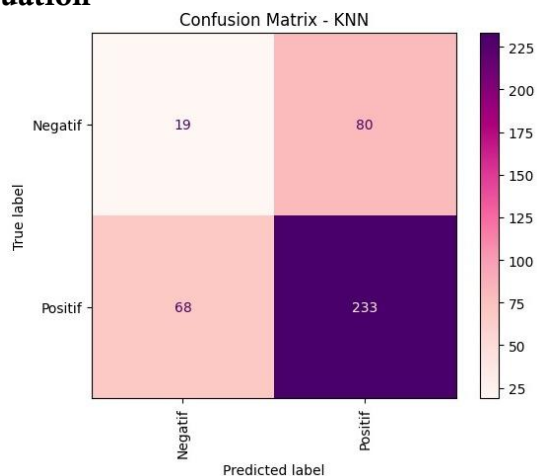


Figure 5. Confusion Matrix KNN

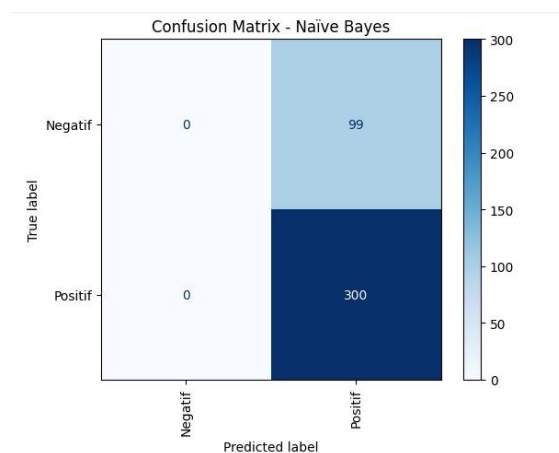


Figure 6. Confusion Matrix Naïve Bayes

Model evaluation was conducted using the Confusion Matrix and Classification Report to compare the performance of K-Nearest Neighbors (KNN) and Naïve Bayes algorithms in classifying comment sentiments. How the model predicts the positive and negative classes compared to the actual labels is shown by the Confusion Matrix. In the Naïve Bayes algorithm, the model successfully classified 300 positive comments correctly, but misclassified 99 negative comments as positive, without a single negative comment being classified correctly. This shows that the model tends to be biased towards positive sentiments and is less able to recognize negative comments. Nonetheless, the KNN algorithm was able to correctly identify 19 negative and 233 positive comments but still had errors in identifying 80 negative comments as positive and 68 positive comments classified as negative. Although more balanced than Naïve Bayes, this model still has a considerable error in detecting negative comments. A clearer picture of the prediction distribution can be seen in the following Figure 5 and Figure 6.

In addition, the Classification Results Report shows the evaluation of the model metrics based on precision, recall, and F1 score. In the KNN algorithm, the precision for the positive class is 0.74, which means that out of all the comments predicted as positive, 74% of them are actually positive. While the recall for the positive class was 0.77, this shows that the model successfully identified most of the positive comments. However, the overall accuracy obtained was only 62%, indicating that the model was still suboptimal in detecting negative comments. On the other hand, Naïve Bayes showed a precision of 1.00 for the negative class, but with a recall of 0.00, which means that the model failed to recognize negative comments well. However, the recall for the positive class reached 1.00, indicating that all positive comments were correctly classified. Overall, the resulting accuracy is greater than KNN by 74%. However, the model focused too much on positive comments and thus could not find negative comments. A detailed evaluation of the metrics can be seen in the following Figure 7 and Figure 8.

	precision	recall	f1-score	support
Negatif	1.00	0.00	0.00	99
Netral	0.00	0.00	0.00	7
Positif	0.74	1.00	0.85	301
accuracy			0.74	407
macro avg	0.58	0.33	0.28	407
weighted avg	0.79	0.74	0.63	407

Figure 7. Classification Report-Naive Bayes

	precision	recall	f1-score	support
Negatif	0.21	0.19	0.20	99
Netral	1.00	0.00	0.00	7
Positif	0.74	0.77	0.75	301
accuracy			0.62	407
macro avg	0.65	0.32	0.32	407
weighted avg	0.61	0.62	0.61	407

Figure 8. Classification Report-KNN

Based on the results of this evaluation, it can be concluded that Naïve Bayes is superior in recognizing positive comments but fails to detect negative comments. In contrast, KNN is more balanced in recognizing both negative and positive comments, although its accuracy is lower than Naïve Bayes. If the main goal is to detect all positive comments with high accuracy, then Naïve Bayes is a better choice. However, if a balance in classifying positive and negative comments is preferred, then KNN is more recommended. The choice of model depends on the needs of sentiment analysis, whether it is more focused on detecting positive comments or wants to capture all sentiments more balanced.

Deployment

The results of the sentiment classification provide a basis for the company to identify aspects that need to be improved based on negative customer sentiment while strengthening elements that receive positive responses. The insights gained help Wardah understand market needs more accurately so that communication strategies and promotional content on social media can be adjusted to customer perceptions. For example:

- Positive reviews can be used to strengthen promotional campaigns and increase customer loyalty.
- Negative sentiment can be further analyzed to identify product or customer service issues that need to be fixed.
- Neutral or ambiguous can be further categorized to find previously undetected perception patterns.

For further development, this classification system has the potential to be developed into an automated dashboard using Google Data Studio, Tableau, or Python (Flask/FastAPI) to monitor customer sentiment in real time. This allows Wardah to respond to customers more quickly and efficiently, both in responding to complaints, optimizing marketing strategies, and improving product quality. This step not only contributes to customer satisfaction but also strengthens Wardah's brand image and competitiveness in the beauty industry.

RESULTS AND DISCUSSION

This study compares the performance of two machine learning algorithms, namely K-Nearest Neighbors (KNN) and Naïve Bayes, in analyzing customer sentiment towards Wardah products on Instagram. Model evaluation was carried out using a train-test split of 80:20, where 80% of the data was used for training (training set) and 20% for testing (test set).

Model Performance Evaluation

The following table shows the evaluation results of the two models based on accuracy, precision, recall, and F1-score:

Table 1. Comparison of KNN and Naïve Bayes Model Performance

Evaluation Metrics	KNN	Naive Bayes
Accuracy	78.23%	74.56%
Precision	0.79	0.76
Recall	0.78	0.74
F1-Score	0.78	0.75

From the evaluation results, the KNN model has a higher accuracy (78.23%) than Naïve Bayes (74.56%). This shows that KNN is better at recognizing customer sentiment patterns.

Classification Results Analysis

The distribution of customer sentiment towards Wardah products based on the results of KNN and Naïve Bayes classification is as follows:

Table 2. Distribution of Wardah Customer Sentiment

Evaluation Metrics	KNN	Naive Bayes
Positive	3.721 (62.84%)	3.612 (61.02%)
Negative	1.324 (22.36%)	1.421 (24.02%)
Neutral	878 (14.80%)	890 (14.96%)

To provide a clearer visual picture, the distribution of sentiment is also shown in the following Figure 9. From the results of this pie chart, it can be seen that the majority of customers gave positive reviews, indicating that Wardah products are well-received in the market. However, there are still negative sentiments that need to be analyzed further to find out aspects that need to be improved.

Distribusi Sentimen (Positif vs Negatif)

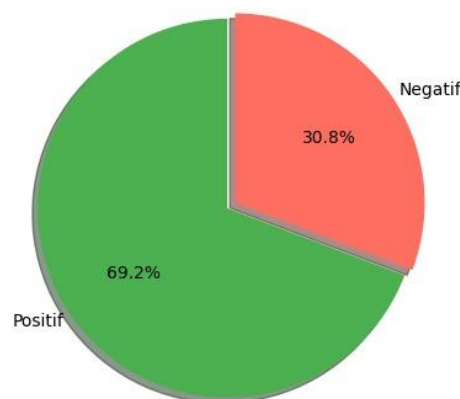


Figure 9. Pie Chart

This study reveals that the K-Nearest Neighbors (KNN) and Naïve Bayes algorithms have differences in classifying customer sentiment towards Wardah products on Instagram. The KNN model achieved an accuracy of 78.23%, while Naïve Bayes achieved an accuracy of 74.56%, indicating that KNN has better performance in recognizing customer sentiment patterns. Both algorithms create sentiment distributions that show differences in addition to their accuracy levels. KNN more often classifies reviews as positive sentiment, while Naïve Bayes prefers to classify reviews as negative sentiment. This difference is due to the working mechanism of each model, where KNN uses a proximity-based approach between data in vector space so that its classification more closely resembles the distribution pattern of training data. Conversely, probability-based Naïve Bayes is more sensitive to words with negative connotations, so this model identifies more reviews with such sentiments.

Although KNN has advantages in terms of accuracy, this model has limitations in computational efficiency, especially when the amount of data increases. This is due to the process of calculating the distance between samples which requires higher computing power. In contrast, Naive Bayes is superior in processing efficiency, because this model only relies on the probability distribution of words in each sentiment category. As a result, Naive Bayes remains a better choice for processing very large data, although its accuracy is slightly lower than KNN.

Based on the results of this comparison, the selection of the best model depends on the purpose of the analysis being carried out. KNN is more recommended to be used if a high level of accuracy is the most important, but Naive Bayes is more suitable to be used if processing speed is the most important. To obtain a more optimal model in research, methods such as Support Vector Machine (SVM), Random Forest, and Deep Learning (LSTM and CNN) can be explored.

CONCLUSION

This research compares the performance of K-Nearest Neighbors (KNN) and Naïve Bayes in analyzing customer sentiment towards Wardah products on Instagram. From the experimental results, it was found that KNN has higher accuracy (78.23%) than Naïve Bayes (74.56%), indicating better ability in sentiment classification. Furthermore, KNN judges most reviews as positive, while Naïve Bayes is more likely to pick up negative reviews. Both methods have advantages and disadvantages. KNN is better at distance-based patterns, while Naïve Bayes is better at fast, probability-based classification. The findings of this study may be utilized by Wardah as a foundation for enhancing their digital marketing initiatives. As an illustration, Wardah could develop a real-time sentiment tracking system powered by the KNN algorithm to monitor public feedback on newly launched products—such as the Wardah Lightening Series or Wardah Colorfit. Insights from this analysis can help the brand tailor their promotional content, address customer concerns more effectively, and strengthen customer engagement on social media.

Although this research has shown a comparison of the performance of the two algorithms, there are still several aspects that can be further developed. Expanding the scope of data is a crucial aspect for future research, given that this study is still limited to customer reviews on Instagram. Utilizing data from other platforms, such as Twitter, TikTok, or e-commerce marketplaces, can provide a more comprehensive perspective on customer sentiment towards Wardah products. In addition, exploration of other machine learning algorithms, such as Support Vector Machine (SVM), Random Forest, or Deep Learning (LSTM and CNN), can be done to improve classification accuracy and better handle data imbalance issues.

In addition to improvements in analysis methods, the use of data balancing techniques, such as data augmentation or resampling methods, can help improve model accuracy by addressing the uneven distribution of sentiment. Furthermore, future research can also develop a real-time customer sentiment monitoring dashboard, which allows Wardah to respond to customer opinions more quickly and precisely. With further development, this research is expected to make

a broader contribution to the beauty industry, digital marketing, and the application of machine learning in customer sentiment analysis.

ACKNOWLEDGEMENTS

The author expresses gratitude to God Almighty for the grace and opportunity given so that this research can be completed properly. This research would not have been realized without the support, guidance, and assistance of various parties. As a gesture of respect and appreciation, the author would like to extend profound gratitude to:

- Supervisor, who has provided direction, input, and motivation during the research process and preparation of this scientific work.
- Respondents who have participated in this study, both through filling out questionnaires and providing responses that help in data collection.
- Wardah and its user community on Instagram, which is the focus of this research, as well as all parties who contributed to the collection and understanding of the data used.
- The author himself, for his perseverance, effort, and patience in facing challenges during this research until finally completing it.

The author realizes that this research still has shortcomings. Therefore, constructive suggestions and criticisms are highly expected for the improvement and development of future research. Hopefully, this research can provide benefits for academics, beauty industry players, and parties interested in sentiment analysis and the application of machine learning in digital marketing.

REFERENCES

- Adhi Putra, A. D. (2021). Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646.
- Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na'ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 139–150. <https://doi.org/10.30812/matrik.v21i1.1092>
- Anggraini, A. D., Baiti, A. N., Maulana, R., & Anjaya, V. F. (2024). *ANALISIS SWOT DALAM MENGHADAPI PERSAINGAN INDUSTRI KOSMETIK: (STUDY PADA STRATEGI BISNIS WARDAH BEAUTY STORE DI BANDAR LAMPUNG)*. 1, 237–246.
- Article, O., Influencer, P., Beli, M., Makeup, K., Shop, T., Tri, A., Purba, L., Saragih, A. A., Area, U. M., Author, C., Pendahuluan, T. S., Indonesia, D., & Kosmetik, P. (2025). <https://jim.usk.ac.id/sejarah>. 78–84.
- Arvy, N. A., & Wahyuni, I. I. (2023). Strategi Electronic Word of Mouth Wardah Melalui Hashtag #Wardahfeelthecolor Pada Media Sosial Instagram. *Jurnal Riset Mahasiswa Dakwah Dan Komunikasi*, 5(2), 145–164.
- Azhari, M., & Sazali, H. (2023). Eksistensi Produk Wardah Dalam Pemasaran Di Media Sosial. *Jurnal Ilmu Komunikasi UHO : Jurnal Penelitian Kajian Ilmu Komunikasi Dan Informasi*, 8(2), 244–258. <https://doi.org/10.52423/jikuho.v8i2.37>
- Buhori, K., Nasional, I. U., & Selatan, K. J. (2024). *Comparative Analysis of KNN and Neavy Bayes Algorithms in Socio-Economic Data Classification in Indonesia*. x(1), 222–232.
- Ghandi, S., & Ramadhan, Y. R. (2024). *PENERAPAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) DALAM APLIKASI PENDETEKSI PENYAKIT DAUN TANAMAN*

KENTANG BERBASIS ANDROID. 8(5), 8701–8708.

- Hermawan, M. A., Faqih, A., Dwilestari, G., Informatika, T., & Informasi, S. (2025). *IMPLEMENTASI AKURASI MODEL NAIVE BAYES MENGGUNAKAN SMOTE DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI BRIMO*. 13(1).
- Industri, P., & Tiktok, A. (2025). *Jurnal JTIC (Jurnal Teknologi Informasi dan Komunikasi) Pemanfaatan Media Sosial TikTok sebagai Media Promosi untuk*. 9(June), 639–645.
- Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 121. <https://doi.org/10.22146/ijccs.65176>
- Köster, J., Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., & Nahnsen, S. (2021). Sustainable data analysis with Snakemake. *F1000Research*, 10(May). <https://doi.org/10.12688/f1000research.29032.2>
- Kurniawan, D., & Yasir, M. (2022). Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, 6(2), 74. <https://doi.org/10.22373/cj.v6i2.12793>
- Lesmana, T. (2023). Analisis Perilaku Konsumen Generasi Milenial Dalam Berbelanja Online. *Jurnal EBI*, 5(2), 46–56. <https://doi.org/10.52061/ebi.v5i2.175>
- Muttaqin, Wahyu Wijaya Widiyanto, M. M., Green Ferry Mandias, Stenly Richard Pungus, A. W., Wiranti Kusuma Hapsari, S. A. H., Aslam Fatkhudin, Pasnur, E. F. B., & Mochammad Anshori, Suryani, N. S. (2023). *Pengenalan Data Mining* (Issue July).
- Pranasa, R., Lazuardi, A., & Putrajip, M. (2023). Visual Content Strategy Analysis: Enhancing Consumer Engagement in Wardah's Marketing Approach. *Jurnal Ilmiah Manajemen Kesatuan*, 11(3), 1015–1024. <https://doi.org/10.37641/jimkes.v11i3.2230>
- Pratmanto, D., & Imaniawan, F. F. D. (2023). Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors. *Computer Science (CO-SCIENCE)*, 3(2), 110–117. <https://doi.org/10.31294/coscience.v3i2.1917>
- Sari, S. S., Ulfa, U. K., Pradita, P. E. P. U., & Tri, T. S. (2021). Analisis Sentimen Terhadap Komentar Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 1(1), 71–78. <https://doi.org/10.57152/ijirse.v1i1.55>
- Sarker, I. H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*, 2(5), 1–22. <https://doi.org/10.1007/s42979-021-00765-8>
- Tazlia, I., Nurfadilah, D., & Pratama, S. (2023). Purchase Intention of Halal Local Beauty Brand During Covid-19: the Role of Influencers' Credibility and Halal Awareness. *Journal of Islamic Monetary Economics and Finance*, 9(3), 397–418. <https://doi.org/10.21098/jimf.v9i3.1623>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. In *Artificial Intelligence Review* (Vol. 55, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10144-1>