Vol. 14, No. 5 (2025): 1905 - 1912 https://doi.org/10.23960/jtepl.v14i5.1905-1912

JURNAL TEKNIK PERTANIAN LAMPUNG

ISSN 2302-559X (print) / 2549-0818 (online)
Journal homepage : https://jurnal.fp.unila.ac.id/index.php/JTP

Comparison of Machine Learning Models for Classifying Consumer Sentiment of Coffee Shops on Social Media X

Agung Putra Pamungkas[1,✉], Adam Mahendra[1], Ibnu Wahid Fakhrudin Aziz[1]

[1] Department of Agroindustrial Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, INDONESIA.

Article History:

Received : 28 June 2025
Revised : 23 August 2025
Accepted : 28 August 2025

Keywords:

Logistic regression,
Naïve bayes,
Sentiment analysis,
Social media,
Support vector machine.

Corresponding Author:
✉ agungputra.pamungkas@ugm.ac.id
(Agung Putra Pamungkas)

ABSTRACT

With the intense competition in the coffee shop industry, understanding consumer opinions has become crucial for businesses. This study analyzes consumer sentiment toward the Janji Jiwa and Kopi Kenangan brands using tweet data from platform X. Sentiments were classified into positive, neutral, and negative categories using three algorithms: Logistic Regression (LR), Naïve Bayes (NB), and Support Vector Machine (SVM). The performance of these algorithms, in terms of accuracy and predictive capability, was evaluated using the TF-IDF method for text representation. The evaluation results show that LR achieved the highest accuracy at 79%, followed by SVM (78%) and NB (75%). Additionally, LR recorded consistent and balanced scores across the precision, recall, and F1-score metrics. These findings indicate that LR and SVM are more effective for multiclass sentiment classification in social media contexts.

1. INTRODUCTION

The coffee shop industry in Indonesia has grown rapidly. This growth reflects changes in lifestyle and consumer preferences. According to Toffin Insight (2020), the number of coffee shops almost tripled between 2016 and 2019, reaching 2,950 outlets in 2019. Coffee shops serve not only as places to buy coffee but also as social spaces where people meet, work, and relax. Among the many brands, Janji Jiwa and Kopi Kenangan stand out as the most successful. Studies indicates that these brands have strong brand value, positive customer experiences, and solid market positions. This is better than other local competitors (Amalia *et al*., 2022; Hidayah *et al*., 2024).

Market share reflects the level of competitiveness of a business (Shadeni & Erinos, 2022). To remain competitive, coffee shops must understand customer needs and preferences. Meeting customer needs is crucial for customers satisfaction, which in turn supports business sustainability (Mardikaningsih, 2021).

Social media platforms such as X (formerly Twitter) provide real-time data on consumer opinions and sentiments. People can freely share their opinions, making X useful for understanding public opinion (Solihin *et al*., 2021). Sentiment analysis helps extract insights from such data. It shows what consumers think about brands by analyzing text content (Alfajri *et al*., 2019; Hasri & Alita, 2022). Sentiment analysis uses natural language processing to find patterns in unstructured text (Priyanto & Ma'arif, 2018; Siringoringo & Jamaluddin, 2019). This approach is useful for understanding brand perception in competitive markets.

Machine learning helps automate sentiment analysis. Algorithms like Naïve Bayes, Logistic Regression, and Support Vector Machine are commonly used as they perform well in categorizing opinions as positive, neutral, or negative (Daqiqil, 2021). Each algorithm has strengths, and it is important to compare them to find the best one. This study

examines public perceptions of Janji Jiwa and Kopi Kenangan based on tweets from X. It compares the accuracy and ability of three machine learning models. The goal is to contribute to existing knowledge and provide practical insights for businesses. It shows how traditional machine learning models work for sentiment analysis. The results can help coffee shops make better decisions, improve marketing, and build better customer relationships.

2. MATERIALS AND METHODS

2.1. Research Object

The data for this study was collected from tweets on the social media platform X (formerly Twitter). These tweets reflect consumer opinions about the coffee shop brands Janji Jiwa and Kopi Kenangan. As illustrated in Figure 1, the research process consisted of several steps, ranging from data collection to model testing. A Python script on Google Colab was used to scrape the tweets. After collecting the tweets, we manually labeled them into three groups: positive, negative, and neutral. A total of 1,230 tweets were selected, with a balanced distribution across the three groups to ensure fairness in testing. The labeled tweets were then cleaned and normalized to prepare the data. Next, we extracted features using the Term Frequency-Inverse Document Frequency (TF-IDF) method. Finally, sentiment classification was performed using three machine learning algorithms: Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM).

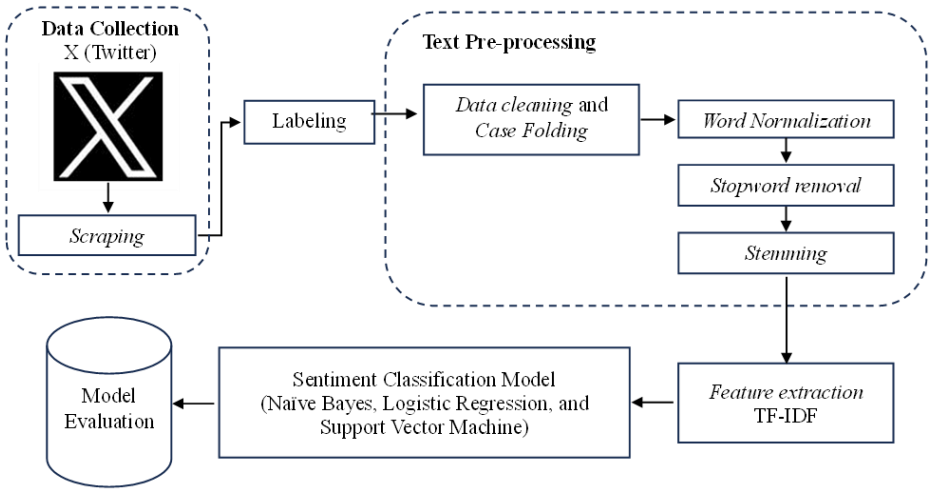

Figure 1. Research workflow

2.2. Data Pre-processing

Data pre-processing is an essential step prior to analyzing raw data. It makes the text easier to interpret and improves model accuracy. In this study, the pre-processing steps included data cleaning, converting text to lowercase, normalizing words, removing stopwords, and stemming. Case folding and cleaning were performed to ensure text consistency. Converting text to lowercase standardized the data, as social media posts often use capital letters unevenly. Cleaning involved removing URLs, numbers, and punctuation to retain only the main content. This process was implemented using Python's "re" module.

Next, word normalization was applied to replace slang or abbreviations. Normalization changed these informal words into standard forms. For this purpose, we used a normalization list developed by Ksnugraha, consisting of 3,720 pairs of non-standard and standard words, which helps improve model accuracy.

We then removed stopwords. These are common words with little meaning, such as "dan," "yang," or "adalah." Removing them reduces dataset size and improves performance. The process used the NLTK library, which provides a list of Indonesian stopwords.

Finally, we performed stemming to simplify words to their root form. This step was carried out using the Sastrawi library with *StemmerFactory*, a Python tool designed for the Indonesian language. By removing affixes, stemming helps make the text more consistent for analysis.

2.3. Feature Extraction (TF-IDF)

We used TF-IDF to transform text into numerical representations. It counts how often a word appears in a document. It also considers how common the word is across all documents. Words that appear frequently across many documents are assigned lower weight. TF-IDF combines these two pieces to create a feature vector. This allows machine learning algorithms to process the data. TF-IDF was chosen because it is simple and effective approach for capturing word importance in a dataset (Ramos, 2003; Manning *et al*., 2008).

2.4. Classification Method

2.4.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a common algorithm used in data analysis. It creates a classification model by learning from labeled data. SVM finds the best hyperplane that separates data points from different classes, such as positive and negative. It does this in a high-dimensional space, which helps in making accurate decisions. The basic form of this hyperplane is shown in Eq. (1):

$$x_i\, w + b = 0 \tag{1}$$

where w represents the normal vector of the plane and b is a constant. The margin or distance between the boundary planes is expressed in Eq. (2):

$$\frac{2}{||w||} \tag{2}$$

when both boundaries are represented as inequalities, they can be written as shown in Eq. (3):

$$y_i(x_i w + b) - 1 \geq 0 \tag{3}$$

The task of finding the optimal separating hyperplane is formulated as an optimization problem. The best hyperplane, defined as the one with the maximum margin, was obtained by minimizing the objective function given in Eq. (4):

$$\min \frac{1}{2} ||\mathrm{w}||^2 \tag{4}$$

2.4.2. Naïve Bayes (NB)

Naïve Bayes (NB) is a method that uses Bayes' Theorem for classification. It predicts outcomes based on probabilities from previous data. It is often used in text classification, especially with features like TF-IDF. In this study, we used Naïve Bayes to create a sentiment classifier. The model was trained on labeled data and then used to identify sentiment in a test set. The basic formula for Naïve Bayes is shown in Eq. (5):

$$P_{(j|t)} = \frac{P_{(j)}\, P_{(t|j)}}{P_{(t)}} \tag{5}$$

where $P_{(j|t)}$ denotes the probability of term t appearing in category j, $P_{(t|j)}$ is the probability of term t belonging to category j, $P_{(j)}$ is the probability of a document belonging to category j, and $P_{(t)}$ is the probability of term t appearing.

2.4.3. Logistic Regression (LR)

Logistic Regression (LR) is a method that estimates the chance that a data point belongs to a specific class. It uses the sigmoid function to convert a linear combination of features into a probability between 0 and 1. This method works well with categories and is often used in text sentiment analysis. The prediction formula for LR is shown in Eq. (6):

$$P(y = 1|x) = \frac{1}{1+\mathrm{e}^{-(\beta_0+\sum_{i-1}^{n}\beta_i x_i)}} \tag{6}$$

where x is the input feature vector, β_i represents the weight (coefficient) for each feature x_i, and β_0 is the intercept (bias) term. The parameters β are estimated by maximizing the likelihood function, typically using optimization algorithms such as *gradient descent* or *liblinear* solver, depending on the dataset size and characteristics.

2.5. Model Evaluation

2.5.1. Confusion Matrix

A confusion matrix is a tool for evaluating the performance of classification models. It presents the number of correct and incorrect predictions for each class. This allows assessment of how well the model distinguishes different sentiments (Han & Kamber, 2001). Table 1 displays the confusion matrix used in this study.

Table 1. Confusion matrix

Actual	Prediction True	Prediction False
True	True Positive (TP)	False Negative (FN)
False	False Positive (FP)	True Negative (TN)

Based on the confusion matrix, we evaluated the model using four metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions, while precision indicates how reliable positive predictions are. Recall assesses how well the model identifies actual positive cases. The F1-score combines precision and recall into a single value, which is especially useful when some classes are less common. These metrics are defined in Equations (7) to (10).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

where *TP* denotes True Positive, *TN* denotes True Negative, *FP* denotes False Positive, and *FN* denotes False Negative.

3. RESULTS AND DISCUSSION

This study compares three machine learning algorithms: Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM). The dataset was collected from social media platform X, formerly Twitter. It consists of tweets posted in 2022 about two Indonesian coffee shop brands: Janji Jiwa and Kopi Kenangan. We used sentiment analysis to classify opinions into three groups: positive, negative, and neutral. Positive sentiment reflects satisfaction with aspects such as taste, price, or service. Negative sentiment indicates dissatisfaction or criticism, while neutral sentiment consists of factual or descriptive statements without emotional tone (Liu, 2012; Medhat *et al*., 2014).

Many factors influence consumer perception of a brand, including price, location, product quality, and promotional efforts. Social media activity and online trends also shape these perceptions (Pang & Lee, 2008). Therefore, sentiment analysis is essential for understanding customer reactions in today's digital marketplace.

First, we conducted experiments using different train-test splits: 50:50, 60:40, 70:30, 80:20, and 90:10. The purpose was to examine the effect of training data size on model accuracy. Each model was trained on the training set and evaluated on the test set. As shown in Figure 2, a larger proportion of training data generally improved accuracy. For example, LR's accuracy increased from 67% with a 50:50 split to 73% with a 90:10 split, while SVM's accuracy rose from 68% to 74%. These results indicates that both models perform better with larger training data.

By contrast, the NB model did not improve as consistently. Its accuracy improved only slightly, from 66% to 69%, as the training data expanded. This limitation arises from NB's assumption of feature independence, which reduces its

ability to capture complex text patterns. Based on these results, a 90:10 split was selected for all subsequent tests to achieved the best performance.

3.1. Hyperparameter Tuning

After determining the optimal train-test split, we applied grid search to identify the best model parameters. Table 2 presents these parameters, along with their corresponding accuracy and macro F1-scores.

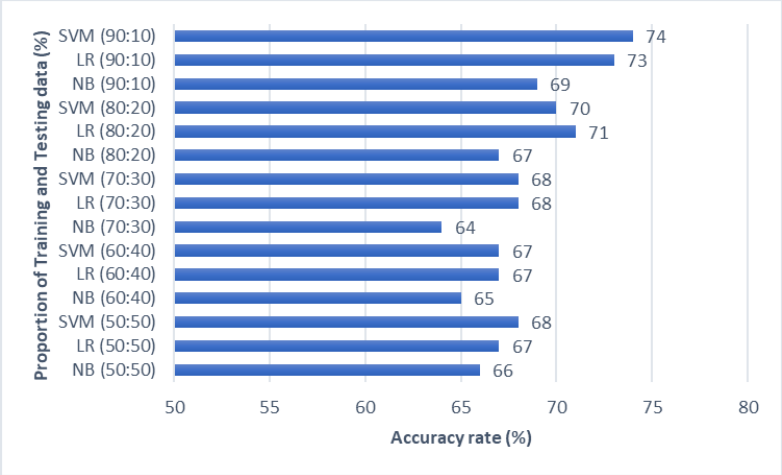

Figure 2. Accuracy results based on the difference between the proportion of training and testing data

Table 2. Hyperparameter tuning results for NB, LR, and SVM

Model	Best Parameters	Accuracy (%)	Macro F1-Score (%)
NB	α = 2.0	74.8	74.4
LR	C = 10, max_iter = 1000, solver = 'liblinear'	78.9	78.8
SVM	C = 1, max_iter = 1000	78.0	77.9

3.2. Confusion Matrix

We used a confusion matrix to evaluate each model. The matrix presents the classification performance on 123 data points, representing 10% of the dataset. In addition, we also calculated key evaluation metrics: accuracy, precision, recall, and F1-score, using both macro and weighted averages. Table 3 summarizes these results. This approach provides insights into how well each model classifies different sentiment categories, particularly in a balanced dataset. The confusion matrices of the three models (NB, LR, and SVM) are shown in Figure 3

The NB model achieved 75% accuracy. Its macro precision was 77%, recall was 75%, and F1-score was 74%. Its performance was fair. It did not perform as well as the other two models. NB assumes features are independent. This limits its ability to understand the meaning of text.

Table 3. Performance comparison of machine learning models in sentiment classification

Evaluation Metrics	Algorithm		
	NB	LR	SVM
Accuracy (%)	75	79	78
Macro Precision (%)	77	79	78
Macro Recall (%)	75	79	78
Macro F1-score (%)	74	79	78
Weighted Precision (%)	77	79	78
Weighted Recall (%)	75	79	78
Weighted F1-score (%)	74	79	78

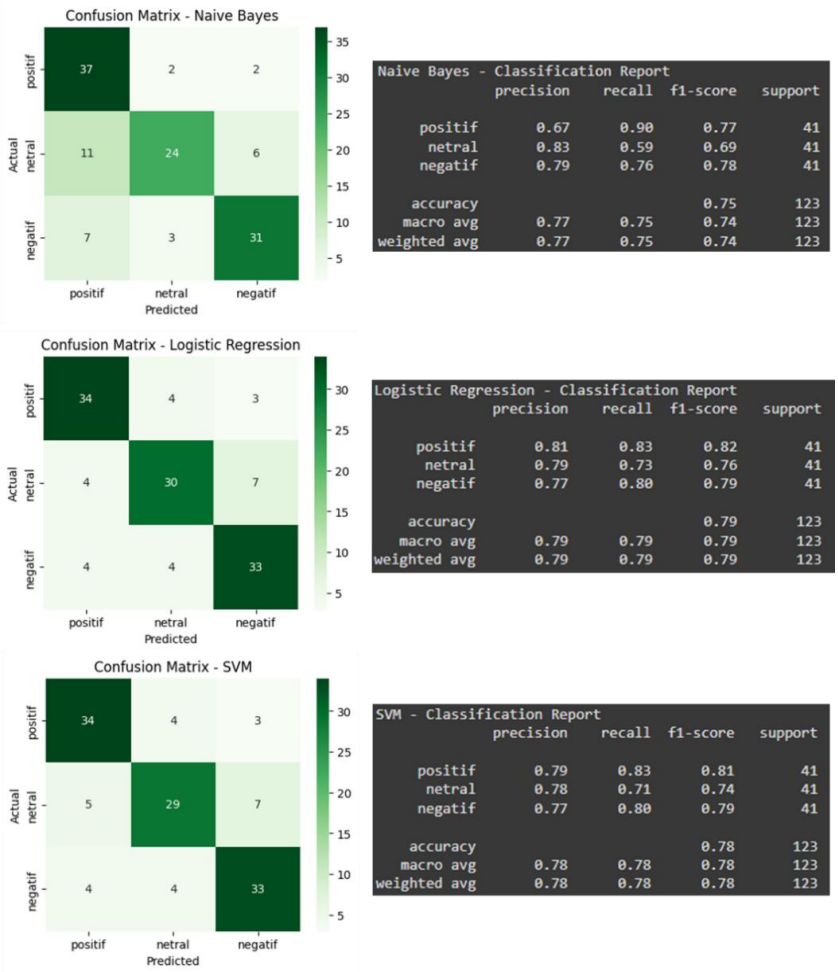

Figure 3. Confusion matrices of the NB, LR, and SVM models

The LR model showed the best results. It had 79% accuracy. Its macro and weighted precision, recall, and F1-scores were also 79%. This means LR is accurate and consistent. It was the most reliable model for this study.

The SVM model also performed well. It had 78% accuracy. Its macro and weighted metrics were 78% as well. SVM was slightly less accurate than LR. It was stable and reliable. It can separate different classes well, even with complex data. These results match earlier studies by Maulana *et al*. (2021) and Mahendra *et al*. (2023). Those studies also found SVM is better than NB in sentiment classification. This shows SVM is good at capturing small differences in text patterns.

Overall, simple models like LR and SVM work well for multi-class sentiment classification. They give consistent results on different metrics. These models are also faster to run than more complex methods. Henderi & Siddique (2024) found that LR and SVM work well on large consumer review datasets. They are reliable choices for sentiment analysis.

LR balances accuracy and stability. It has high accuracy and fewer errors in positive sentiment classification than SVM, according to Budianto *et al*. (2024). LR is also easier to interpret. Its coefficients show which words influence sentiment. This is useful for marketing and business work.

Classical models like LR and SVM often perform and or better than deep learning models. This is true when using simple text features such as TF-IDF or *bag of words*. Recent studies show that traditional algorithms are still good and faster options for sentiment analysis. Kamruzzaman & Kim (2024) support this. These results mean that LR and SVM are good choices for practical sentiment analysis tasks. They are reliable, efficient, and easy to understand, especially for work with balanced datasets.

4. CONCLUSIONS

The evaluation of three sentiment classification models: Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM) indicates that performance is directly tied to an algorithm's ability to capture the contextual relationships in textual data. LR achieved the best results overall, with consistently high accuracy and balanced evaluation metrics. It was followed closely by SVM, which also demonstrated robust and stable performance. In contrast, NB proved to be less effective, likely due to its reliance on the assumption of feature independence, which limits its capacity to handle complex, nuanced semantics. These results highlight a key consideration for model selection: beyond just simplicity and computational efficiency, flexibility and balance are essential for effective sentiment classification. LR, in particular, emerged as the most reliable approach in this study. From a practical standpoint, businesses can use these findings to refine their marketing strategies and improve customer engagement by leveraging insights from consumer sentiment. For future research, expanding the dataset and exploring more advanced methods, such as deep learning models, are recommended to further enhance classification performance.

REFERENCES

Alfajri, M.F., Adhiazni, V., & Aini, Q. (2019). Pemanfaatan social media analytics pada Instagram dalam peningkatan efektivitas pemasaran. *Interaksi: Jurnal Ilmu Komunikasi*, ***8***(1), 34–42. https://doi.org/10.14710/interaksi.8.1.34-42

Amalia, F.N., Ariadi, B.Y., & Widyastuti, D.E. (2022). Comparative analysis of brand equity of Janji Jiwa Coffee Shop with Kenangan Coffee Shop in Malang City. *AGRIMOR*, ***7***(2), 45-53. https://doi.org/10.32938/ag.v7i2.1589

Anugrah, D.P., & Suparwito, H. (2022). Analisis sentimen bantuan langsung tunai COVID-19 menggunakan algoritma support vector machine. *Seminar Nasional Sanata Dharma Berbagi 2022*. Accessed on September 19, 2025 from: https://e-conf.usd.ac.id/index.php/usdb/usdb2022/paper/view/1730

Budianto, A.G., Rusilawati, Suryo, A.T.E., Cahyono, G.R., Zulkarnain, A.F., & Martunus. (2024). Perbandingan performa algoritma support vector machine (SVM) dan logistic regression untuk analisis sentimen pengguna aplikasi retail di android. *Jurnal Sains dan Informatika*, ***10***(2). https://doi.org/10.34128/jsi.v10i2.911

Daqiqil, I. (2021). *Machine Learning*: *Teori, Studi Kasus, dan Implementasi Menggunakan Python.* Pekanbaru: UR Press. Diakses dari https://zenodo.org/records/5113507

Han, J.W., & Kamber, M. (2001). *Data Mining: Concepts and Techniques.* San Francisco: Morgan Kaufman Publisher. Diakses dari https://hanj.cs.illinois.edu/bk2/toc.pdf

Hasri, C.F., & Alita, D. (2022). Penerapan metode naive bayes classifier dan support vector machine pada analisis sentimen terhadap dampak virus corona di twitter. *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, ***3***(2), 145-160. Diakses dari https://jim.teknokrat.ac.id/index.php/informatika/article/view/2026

Henderi & Siddique, Q. (2024). Comparative analysis of sentiment classification techniques on flipkart product reviews: A study using logistic regression, SVC, random forest, and gradient boosting. *Journal of Digital Market and Digital Currency*, ***1***(1), 21-42. https://doi.org/10.47738/jdmdc.v1i1.4

Hidayah, N., Nur, A., Supira, W., Fauziah, E., Ratna, S., Fuzi, Y., Ramadita, N., Rifki, M., Haidar, M., Martin, S., & Huda, M. (2024). Analisis perbandingan customer experience dan harga pada coffee shop "Kopi Kenangan", "Janji Jiwa" dan "Starbucks" di Cikarang. *Journal of Management and Creative Business*, ***2***(3), 255-267. https://doi.org/10.30640/jmcbus.v2i3.2871

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression* (Vol. 398). John Wiley & Sons. Diakses dari https://dl.icdst.org/pdfs/files4/7751d268eb7358d3ca5bd88968d9227a.pdf

Kamruzzaman, M. & Kim, G.L. (2023). Efficient sentiment analysis: a resource-aware evaluation of feature extraction techniques, ensembling, and deep learning models. *Proceeding of the 11th International Workshop on Natural Language Processing for Social Media*, 9-20. https://aclanthology.org/2023.socialnlp-1.2.pdf

Liu, B. (2012). *Sentiment analysis and opinion mining. Morgan & Claypool Publishers*. Diakses dari https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

Mardikaningsih, R. (2021). Pencapaian kepuasan pelanggan pada jasa pengiriman barang melalui harga, ekuitas merek, dan kualitas pelayanan. *Jurnal Baruna Horizon, **4**(1)*, 64-73. https://doi.org/10.52310/jbhorizon.v4i1.58

Mahendra, A., Pamungkas, A.P., & Aziz, I.W.F. (2023). Analisis sentimen dari media sosial twitter terhadap Kopi Janji Jiwa dan Kopi Kenangan menggunakan Metode Machine Learning. [*Undergraduated Thesis*]. Universitas Gadjah Mada.

Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. Diakses dari https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

Maulana, A.N., Pamungkas, A.P., & Falah, M.A.F. (2021). Sentimen Ulasan untuk Peningkatan Layanan dan Pengalaman Pelanggan Menggunakan Bahasa Pemograman Python (Studi di Toko Online Produk Makanan Pendamping ASI) [*Undergraduated Thesis*]. Universitas Gadjah Mada.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, ***5***(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, ***2***(1–2), 1–135. https://doi.org/10.1561/1500000011

Priyanto, A., & Ma'arif, M.R. (2018). Implementasi web scraping dan text mining untuk akuisisi dan kategorisasi indormasi dari internet (Studi kasus: Tutorial hidroponik). *Indonesian Journal of Information Systems (IJIS)*, ***1***(1), 25-33. https://doi.org/10.24002/ijis.v1i1.1664

Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning,* 1-4. Diakses dari https://www.scribd.com/document/339224468/TF-IDF-to-determine-word-relevance-in-document-queries-pdf

Shadeni, E.A., & Erinos. (2022). Pengaruh market share dan intellectual capital terhadap kinerja keuangan perbankan syariah di Indonesia. *Jurnal Eksplorasi Akuntansi (JEA)*, ***4***(2), 363-376. https://doi.org/10.24036/jea.v4i2.531

Siringoringo, R., & Jamaluddin. (2019). Text mining dan klasterisasi sentimen pada ulasan produk toko online. *Jurnal Penelitian Teknik Informatika UNPRI*, ***2***(2), 314-319. https://doi.org/10.34012/jutikomp.v2i1.456

Solihin, F., Awaliyah, S., & Shofa, A.M.A. (2021). Pemanfaatan Twitter sebagai media penyebaran informasi oleh dinas komunikasi dan informatika. *Jurnal Pendidikan Ilmu Pengetahuan Sosial (JPIPS), **1**(13)*, 52-58. https://e-journal.upr.ac.id/index.php/JP-IPS/article/view/2813

Toffin Indonesia. (2020, November 12). *Toffin Indonesia Merilis Riset "2020 Brewing in Indonesia"*. Diambil kembali dari TOFFIN INSIGHT: Diakses dari https://insight.toffin.id/toffin-stories/toffin-indonesia-merilis-riset-2020-brewing-in-indonesia/

Wang, J. (2025). Comparative analysis of machine learning and deep learning models for text emotion classification in federated learning. *Applied and Computational Engineering, **155***, 220-227. https://doi.org/10.54254/2755-2721/2025.GL23568

Zhafira, D.F., Rahayudi, B., & Indriati. (2021). Analisis sentimen kebijakan kampus merdeka menggunakan naive bayes dan pembobotan TF-IDF berdasarkan komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi (JUST-SI)*, ***2***(1), 55-63. https://doi.org/10.25126/justsi.v2i1.24