Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Sentiment Analysis of Twitter Discussions About Lampung Robusta Coffee: A Comparative Study of Machine Learning Algorithms with SVM as The Optimal Model

Yodhi Yuniarthe[1*], Admi Syarif[2], Imam Marzuki Sofie[3] and Fatimah Fahurian[4]

[1]Department of Informatics, Faculty of Computer, Indonesia Mitra University
[2]Department of Computer Science-Faculty of Mathematics and Natural Sciences, Lampung University
[3] Department of Informatics, Faculty of Science and Technology, UIN Syarif Hidayatullah Jakarta
[4]Department of Information System, Faculty of Computer, Indonesia Mitra University
[1,4]JL.ZA Pagar Alam Gedong Meneng Rajabasa, Bandar Lampung, Indonesia
[2]Jl. Sumantri Brojonegoro No. 1, Bandar Lampung, Lampung, Indonesia
[3]Jl. Ir H. Juanda No.95, Ciputat, Kec. Ciputat Team., Banten, Indonesia

Article:
Accepted: June 26, 2025
Revised: April 17, 2025
Issued: October 30, 2025

***Correspondence Address:**
yodhi@umitra.ac.id

ABSTRACT

Lampung Robusta coffee is an important commodity in Indonesia, particularly in terms of local economic potential and global recognition. However, public perception of this product on social media, particularly Twitter, remains underexplored. This study addresses the need for a deeper understanding of consumer sentiment towards Lampung Robusta coffee, which could inform branding and marketing strategies. To approach this issue, we used five supervised machine learning algorithms-KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression-to perform sentiment classification on a dataset of tweets containing relevant keywords. The dataset was pre-processed using standard natural language processing techniques, including tokenization, stopword removal, and TF-IDF feature extraction. The SVM achieved the best performance on the unbalanced dataset for all metrics, with high and consistent accuracy and F1 scores. Logistic regression followed closely with similarly strong and stable results. Therefore, SVM is recommended as the final model. These results suggest that machine learning approaches can effectively classify sentiment in social media discussions about regional agricultural products and that random forest may provide the most robust performance in this context.

Keywords : *lampung robusta coffee; sentiment analysis; twitter; machine learning.*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

1. INTRODUCTION

The province of Lampung is the second-largest coffee-producing region in Indonesia [1]. Lampung Robusta coffee is one of the well-known coffee varieties in Indonesia, particularly in the Lampung region. This coffee has a distinctive taste with a bitter and strong aroma, making it popular in both domestic and international markets [2].

Indonesia, as one of the largest coffee producers in the world, has a growing coffee consumption. According to the local Central Bureau of Statistics, coffee production experienced fluctuations from 2020 to 2022. In 2020, coffee production was 762.38 thousand tons, increasing to 786.19 thousand tons in 2021, representing a 3.12 percent increase. However, in 2022, coffee production decreased to 774.96 thousand tons, a decline of 1.43 percent [3]–[5].

The analysis of sentiment on social networks, such as Twitter or Facebook, has become a powerful means of learning about the users' opinions and has a wide range of applications. However, the efficiency and accuracy of sentiment analysis is being hindered by the challenges encountered in natural language processing (NLP) [6]. Sentiment analysis, also known as opinion mining, is a process of automatically understanding, extracting, and processing textual data. This is done to obtain the sentiment information contained within an opinion. Sentiment analysis is conducted to observe opinions or the tendencies of opinions on a particular topic or issue expressed by a group of people. These opinion tendencies can be either positive or negative [7]. Sentiment analysis and opinion mining are terms that are interchangeably used to refer to a field of study that concludes a product or organization can be affected by people's views, emo- tions, and attitudes [8]. By empowering sentiment analysis on e-commerce platforms to make informed judgements and expand their service offerings, these results have substantial practical consequences and will give them confidence in their strategies [9]. A comparison of sentiment analysis methods that utilise machine learning has been demonstrated to illuminate the merits of various strategies, thereby propelling the field towards enhanced precision and dependability in sentiment analysis systems [10].

This research becomes relevant amidst the increasing reliance of industry players on AI-based data analysis to understand consumer behavior. On the other hand, the lack of literature discussing sentiment analysis of local products, especially Lampung Robusta Coffee, opens a research gap that needs to be filled to make a real contribution in the field of information technology and digital agribusiness. Research related to sentiment analysis of coffee has grown rapidly in recent years, particularly with the increasing use of social media as a valuable data source. According to a study by Samoggia et al. (2020), consumer perceptions of coffee's health attributes were explored using Twitter data. This research focused on how consumers view the health benefits of coffee, utilizing content analysis and sentiment analysis to identify these perceptions. The study found that most tweets tended to be neutral or slightly positive regarding the impact of coffee on health [11].

In recent years, social media platforms such as Twitter have become popular among fans to discuss and share their opinions about the matches [12]. Twitter or X is one of the most popular social media platforms. Twitter users are free to post and express anything, including their opinions, which may consist of facts, suggestions, information, and criticism towards something [13]. This research differs from previous studies as it focuses on sentiment analysis of Lampung Robusta coffee, specifically on the Twitter platform, using a comprehensive machine learning approach. SVM is often used to find the one with the best global attributes [14].Therefore, this study aims to compare the performance of various algorithms in classifying sentiment from Twitter data related to Lampung Robusta coffee using different machine learning algorithms such as Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM). In this study, five machine learning algorithms - Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbors (KNN)—were selected for comparative analysis in sentiment classification. These algorithms were chosen on the basis of their popularity and evidence of effectiveness in previous sentiment analysis research.

This research makes several contributions to the field of sentiment analysis and regional product promotion. Firstly, this research focuses specifically on Twitter data pertaining to Lampung Robusta Coffee, a topic that has received scant attention in the context of computational analysis. This study makes a novel contribution by applying supervised machine learning techniques to analyze public sentiment towards Lampung Robusta coffee. This is a geographically specific agricultural product from Indonesia that has received limited exploration in computational and social media research. This study thus seeks to establish a connection between the promotion of local agricultural products and the utilization of modern machine learning applications. Secondly, the comparative evaluation of five distinct machine learning algorithms provides insights into their relative performance on real-world, noisy, and domain-specific social media data. This study provides a substantial dataset and performance benchmarks that will act as a valuable reference point for future research in the fields of sentiment classification and regional product branding through digital platforms. This approach demonstrates the potential of machine learning to inform data-driven marketing and branding strategies for local agricultural products, thereby facilitating the integration of AI technologies into regional economic development.

2. METHODS

This study aims to compare sentiment analysis of Lampung Robusta Coffee on the Twitter platform using five machine learning algorithms: Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM). The research methodology is divided into several stages: data collection, data preprocessing, modeling, model evaluation, and result interpretation.

2.1 Data Collection

Data was collected from Twitter using the Twitter API, which allows for retrieving tweets based on specific keywords [15]. The keywords used in this study include 'Lampung Robusta Coffee,' 'Lampung Coffee,' and 'Robusta Lampung.' Data collection was conducted over a specified period, such as one month, to obtain a sufficiently large and representative sample of tweets. The collected tweets include the tweet text, posting date, username, and other relevant metadata.

2.2 Data Prepocessing

Data preprocessing is an important step in the sentiment analysis process, as it helps standardize the text data and remove any irrelevant or noisy elements [16]. The raw data obtained from Twitter needs to be further processed to be ready for sentiment analysis. The preliminary stage in the preprocessing is the cleansing of the data. The process of data cleaning has been defined as the improvement of data sets through the replacement, deletion, or modification of irrelevant or valid data [17].

The preprocessing steps include [18]:

a. Cleansing: Removing symbols, punctuation, URLs, usernames, and other irrelevant elements from the tweet text.
b. Tokenization: Breaking the tweet text into individual words or tokens.
c. Lowercasing: Converting all text to lowercase to reduce variations of the same word.
d. Stopword Removal: Removing common words (stopwords) that do not provide significant information, such as 'and,' 'in,' 'that,' etc.
e. Stemming: Reducing words to their base form to simplify analysis, for example, converting 'liking' to 'like.'
f. Filtering: Removing irrelevant tweets, such as spam or tweets not related to Lampung Robusta Coffee.

2.3 Modeling

After the data is processed, the next step is modeling using machine learning algorithms. In this stage, the cleaned data is divided into training data and testing data with a certain ratio, for example, 80% training data and 20% testing data. Below is a brief explanation of each algorithm used [19]:

a. Decision Tree (DT): This algorithm builds a predictive model in the form of a branching decision tree, where each node represents a feature attribute, and the branches represent the outcomes of decisions based on those attributes.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

b. K-Nearest Neighbors (KNN): This algorithm classifies a tweet based on the majority of its k-nearest neighbors in the feature space. KNN is a simple yet effective algorithm for classification with a relatively small amount of data.

c. Logistic Regression (LR): Although initially developed for regression, LR is often used for binary classification. This algorithm predicts the likelihood of a tweet having a positive or negative sentiment based on a logistic function of the input features.

d. Naive Bayes (NB): This probabilistic algorithm is based on Bayes' Theorem, which assumes that text features are independent. Naive Bayes is commonly used in text classification tasks, including sentiment analysis.

e. Support Vector Machine (SVM): This algorithm seeks the best hyperplane that separates the data into two distinct classes. SVM is highly effective in text classification tasks with high-dimensional data.

2.4 Model Evaluation

After the models are trained, evaluation is performed on the testing data to measure the performance of each algorithm. The evaluation metrics used include [20] [21]:

a. Accuracy: The percentage of correct predictions out of the total predictions.

b. Precision: The proportion of true positive predictions out of all positive predictions.

c. Recall: The proportion of correctly predicted positive instances out of all actual positive instances.

d. F1-Score: The harmonic mean of precision and recall, providing a balance between the two. The f1-score obtained from each model could represent in the form of a table based on the sentiments [22].

In addition, in some experiments, researchers also use data balancing techniques such as class weighting and SMOTE (Synthetic Minority Over-sampling Technique) to address the issue of class imbalance in sentiment data.

2.5 Interpretation of Results

The results of each algorithm are compared based on the aforementioned metrics. This analysis helps determine which algorithm is best suited for sentiment analysis of Lampung Robusta Coffee on Twitter. The interpretation of results also includes an analysis of the differences in algorithm performance under unbalanced data conditions, with class weight balancing, and after applying the SMOTE technique. interpretation of results also includes an analysis of the differences in algorithm performance under unbalanced data conditions, with class weight balancing, and after applying the SMOTE technique interpretation of results also includes an analysis of the differences in algorithm performance under unbalanced data conditions, with class weight balancing, and after applying the SMOTE technique.

3. RESULTS AND DISCUSSION

The following are the results and discussion of this study, which include both training and test data.

Figure 1. *accuracy of unbalanced (Original) results across all algorithms (KNN, Naive Bayes, SVM, Decision* Tree, and Logistic Regression)

Figure 1 displays the performance of various classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, Logistic Regression) on an imbalanced dataset with different ratios of minority and majority classes (ranging from 10%:90% to 50%:50%).

The metrics used are Accuracy, Precision, Recall, and F1-Score for both training (Train) and testing (Test) data. Overall, Logistic Regression and Decision Tree demonstrate the best performance across most ratios, with consistently higher accuracy and F1-Score compared to other algorithms, especially on the test dataset. This table displays various splits of training and testing data, ranging from 10%:90% to 50%:50%. Each algorithm is evaluated using the metrics of Accuracy, Precision, Recall, and F1-Score. The

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

results show that Logistic Regression (LR) has the most consistent and superior performance across all metrics, while SVM and Decision Tree also exhibit strong results, particularly in F1-Score. Conversely, KNN and Naive Bayes tend to have lower accuracy and precision, especially on the test data. Overall, LR stands out as the best-performing algorithm across different data split scenarios. SVM excelled consistently across all data ratios, with the highest accuracy (up to 80.61%).

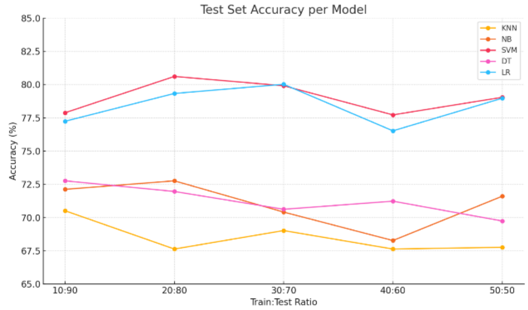

Figure 2. *visualisation to compare the accuracy and F1-Score on test data (test set) of each algorithm across various training:test data ratios*

Based on the results of evaluating five classification algorithms-K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR)-with a class balancing approach through class weight. SVM showed the highest accuracy and F1-Score on the test data, especially on the 20:80 and 30:70 ratios, with accuracy reaching 80.61% and F1-Score 80.62%. Logistic Regression matched SVM's performance in many ratios, and was more stable overall with a maximum accuracy of 80.02% and F1-Score of 80.01%. The best model for this classification case is Support Vector Machine (SVM) if you want maximum accuracy. Logistic Regression (LR) if you want a simple, fast, and still highly accurate model.

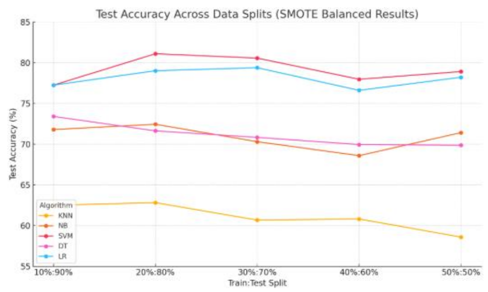

Figure 3. *shows a line graph showing the testing accuracy of various algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) on five different data sharing scenarios with the SMOTE method.*

This figure presents the use of the SMOTE technique to address imbalanced data. The tests were conducted with various training and testing data splits, ranging from 10%:90% to 50%:50%. Each algorithm was evaluated based on Accuracy, Precision, Recall, and F1-Score.

The results indicate that Logistic Regression (LR) consistently performs the best, particularly in terms of Accuracy and Precision. SVM also shows high results, especially in F1-Score. In contrast, when using test data, KNN tends to have lower performance compared to other algorithms, particularly in Recall and F1-Score. Overall, the use of SMOTE generally enhances the performance of algorithms in handling data imbalance.

The evaluation results of the five machine learning algorithms for sentiment analysis of Lampung Robusta Coffee on Twitter can be assessed using various performance metrics. Based on the results obtained:

a. KNN shows an accuracy of 76.32% under unbalanced data conditions and increases to 77.93% when using class weight balancing. However, its performance drops to 69.02% when using the SMOTE technique.
b. Naive Bayes (NB) consistently shows high performance with accuracy above 90% across all data settings, whether unbalanced, class weight, or SMOTE.
c. Support Vector Machine (SVM) achieves the best performance with an accuracy of 96.86% under unbalanced data conditions and shows a slight increase when using class weight balancing and SMOTE.
d. Decision Tree (DT) reaches an accuracy of 84.41% under unbalanced conditions, and its performance remains relatively stable around 83%-84% in other settings.
e. Logistic Regression (LR) also demonstrates strong performance with an accuracy of 93.91% under unbalanced conditions and increases to 95.38% when using class weight balancing and SMOTE.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

3.1 Data Train

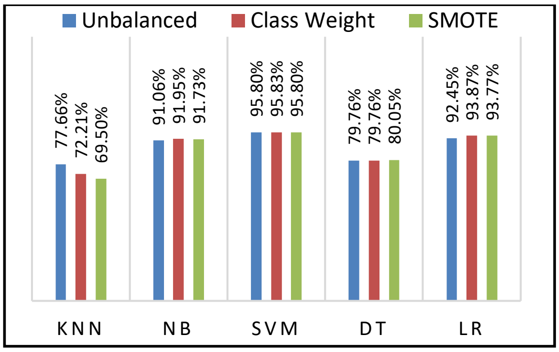

Figure 4. *Accuracy of the algorithm in comparison graph of 10%:90%*

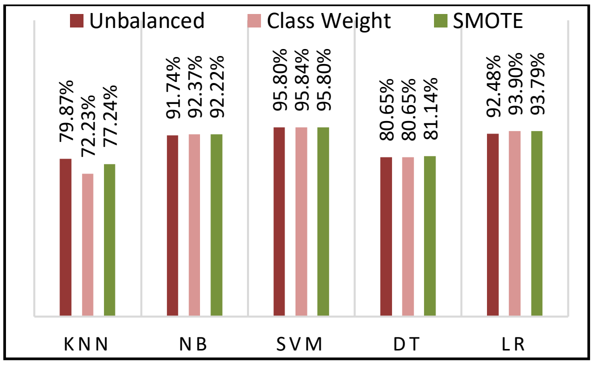

Figure 5. *Precision of the algorithm in comparison graph of 10%:90%*

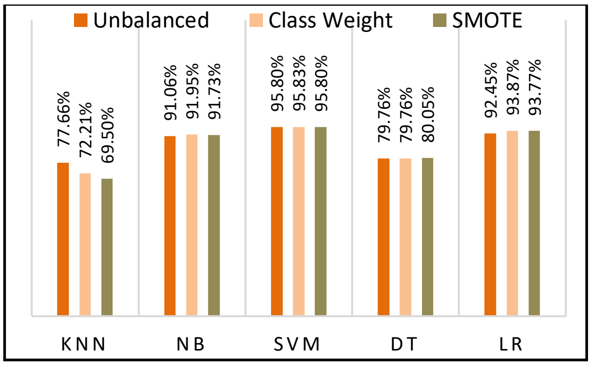

Figure 6. *Recall of the algorithm in comparison graph of 10%:90%*

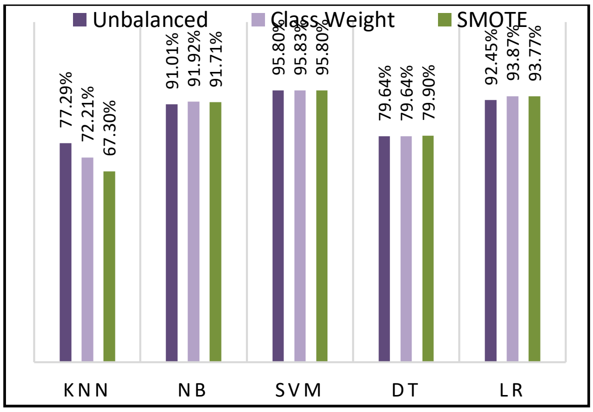

Figure 7. *F1-score of the algorithm in comparison graph of 10%:90%*

Based on images 4-7, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 10%:90% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the Class Weight and SMOTE methods consistently improve accuracy and precision across all algorithms compared to unbalanced data, particularly in the SVM and Logistic Regression algorithms, with the highest values exceeding 90%.

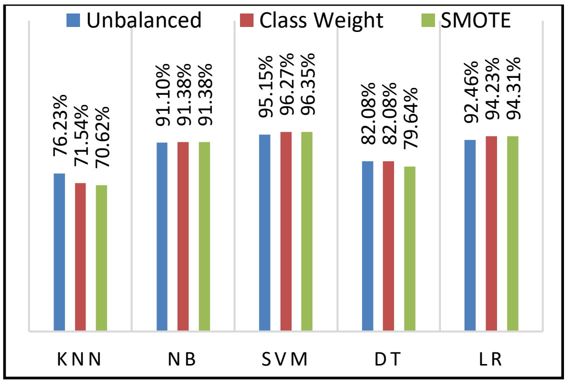

Figure 8. *Accuracy of the algorithm in comparison graph of 20%:80%*

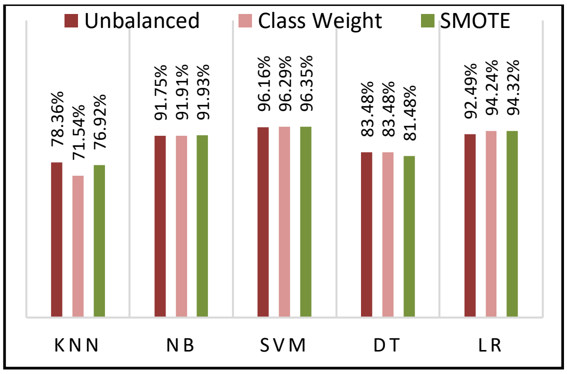

Figure 9. *Precision of the algorithm in comparison graph of 20%:80%*

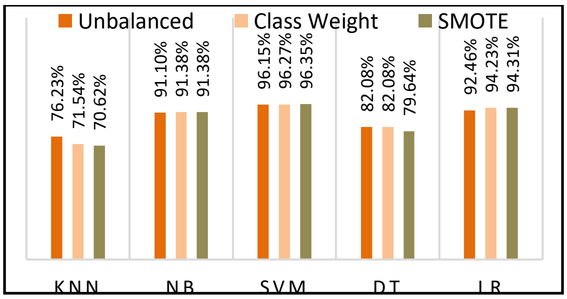

Figure 10. *Recall of the algorithm in comparison graph of 20%:80%*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

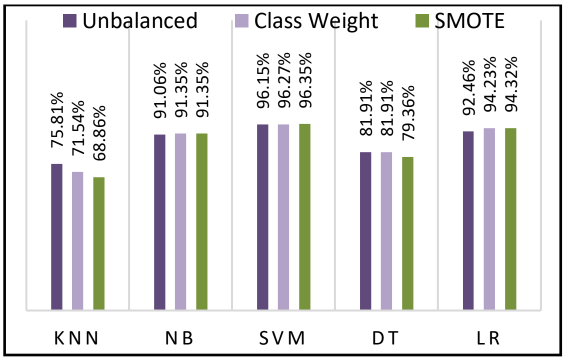

Figure 11. *F1-score of the algorithm in comparison graph of 20%:80%*

Based on images 8-11, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 20%:80% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the SVM and Logistic Regression algorithms achieve the highest values, exceeding 90%.

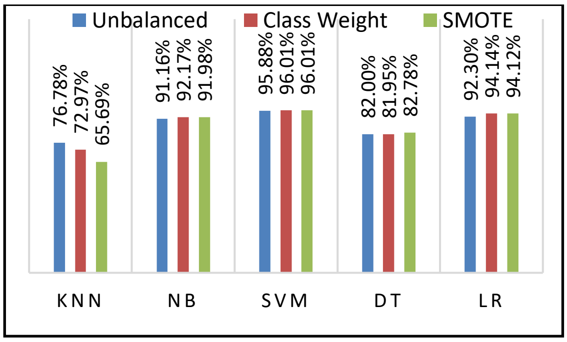

Figure 12. ***Accuracy of the algorithm in comparison graph of 30%:70%***

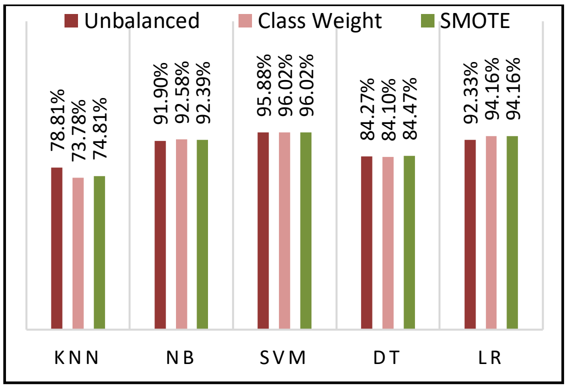

Figure 13. *Precision of the algorithm in comparison graph of 30%:70%*

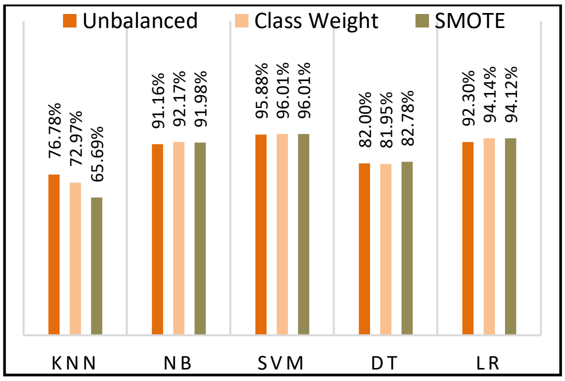

Figure 14. *Recall of the algorithm in comparison graph of 30%:70%*

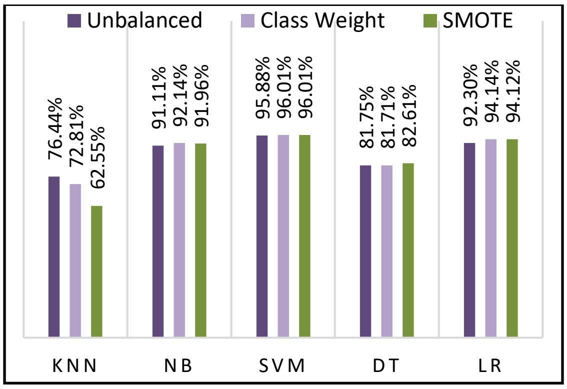

Figure 15. *F1-score of the algorithm in comparison graph of 30%:70%*

Based on images 12-15, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 30%:70% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the SVM and Logistic Regression algorithms achieve the highest values, exceeding 90%.

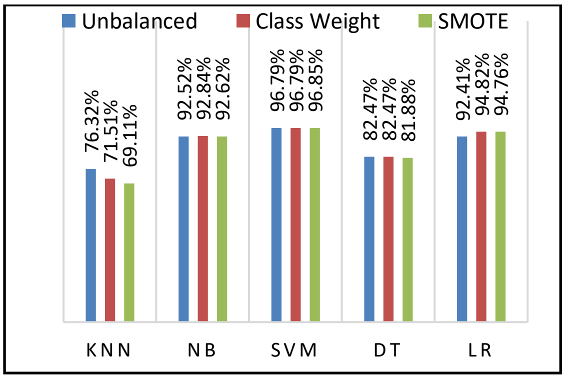

Figure 16. *Accuracy of the algorithm in comparison graph of 40%:60%*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

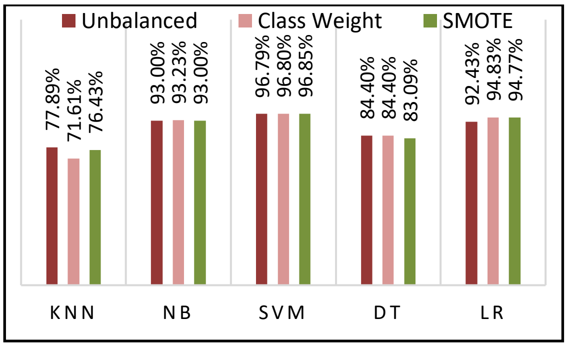

Figure 17. *Precision of the algorithm in comparison graph of 40%:60%*

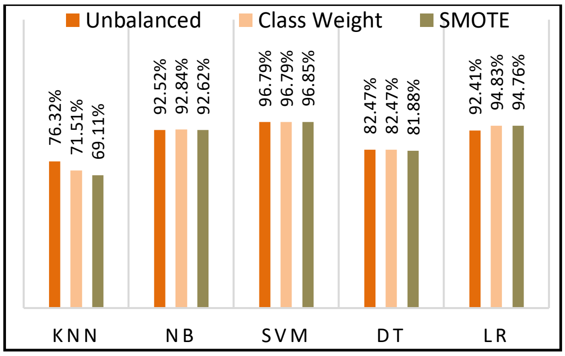

Figure 18. *Recall of the algorithm in comparison graph of 40%:60%*

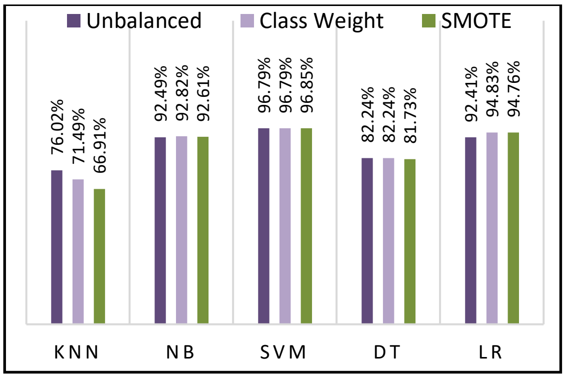

Figure 19. *F1-Score Of The Algorithm In Comparison Graph Of 40%:60%*

Based on images 16-19, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 40%:60% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the SVM and Logistic Regression algorithms achieve the highest values.

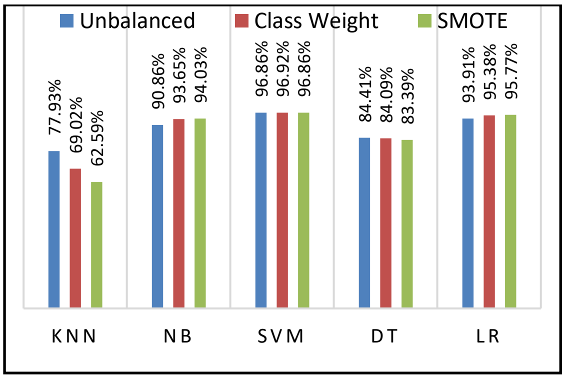

Figure 20. *Accuracy of the algorithm in comparison graph of 50%:50%*

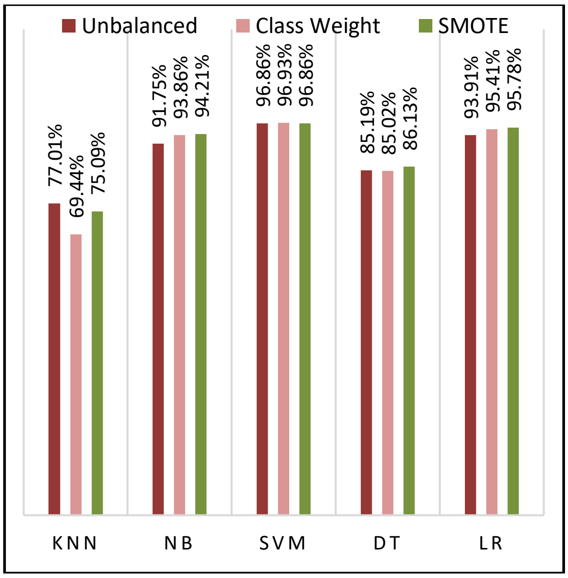

Figure 21. *Precision of the algorithm in comparison graph of 50%:50%*

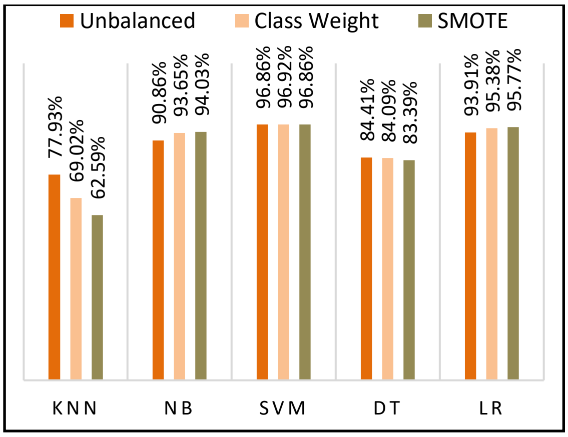

Figure 22. *Recall of the algorithm in comparison graph of 50%:50%*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (339-349)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

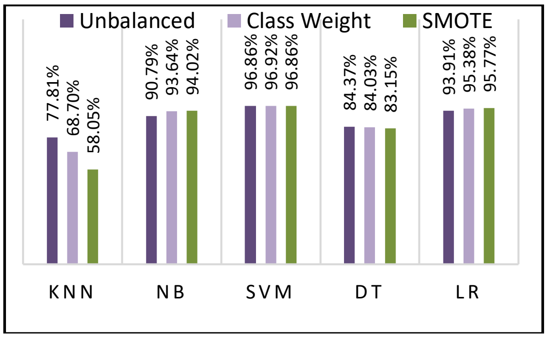

Figure 23. *F1-score of the algorithm in comparison graph of 50%:50%*

Based on images 20-23, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 50%:50% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the SVM and Logistic Regression algorithms achieve the highest values.

3.2 Data Test

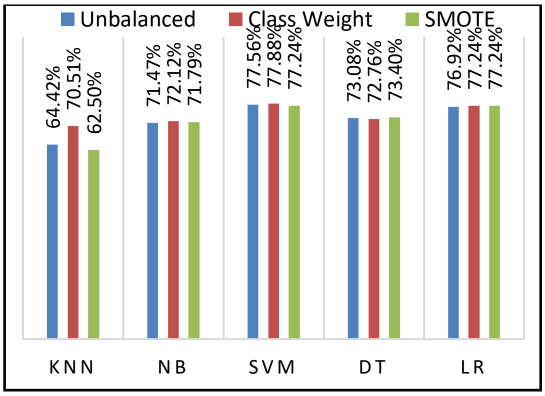

Figure 24. *Accuracy of the algorithm in comparison graph of 10%:90%*

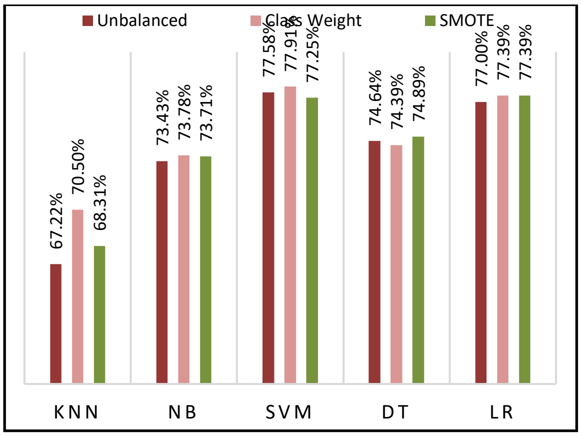

Figure 25. *Precision of the algorithm in comparison graph of 10%:90%*

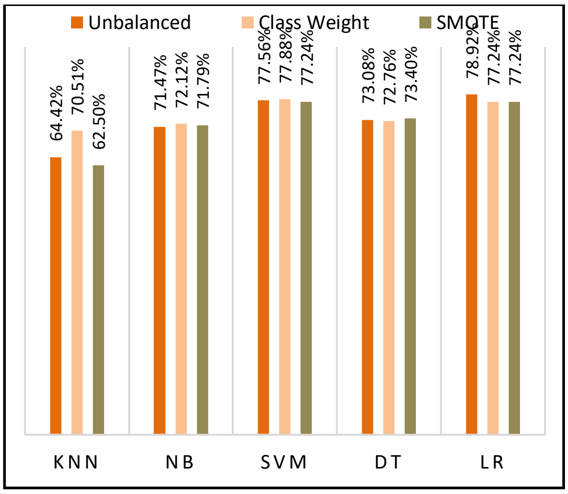

Figure 26. *Recall of the algorithm in comparison graph of 10%:90%*

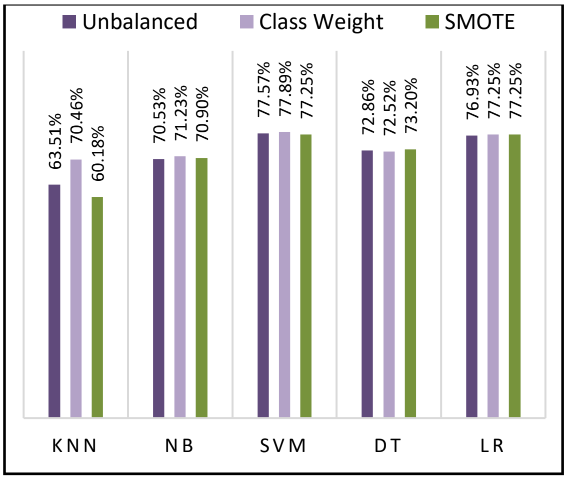

Figure 27. *F1-score of the algorithm in comparison graph of 10%:90%*

Based on images 24-27, the graphs display a comparison of accuracy, precision, recall, and F1-score of five classification algorithms (KNN, Naive Bayes, SVM, Decision Tree, and Logistic Regression) in handling imbalanced data with a 10%:10% ratio. Three methods were applied: without handling (Unbalanced), Class Weight, and SMOTE. The results indicate that the SVM and Logistic Regression algorithms achieve the highest values.

Overall, the SVM algorithms demonstrate the best performance in sentiment analysis of Lampung Robusta coffee on Twitter, making them reliable choices for quick and straightforward results.

CONCLUSION

From this research, it can be concluded that sentiment analysis of Lampung Robusta coffee on Twitter can be effectively conducted using machine learning algorithms. Overall performance based on test set accuracy, with emphasis on model stability and superiority: SVM (Support Vector Machine) consistently provides the highest test accuracy results in almost all split ratios (e.g., 77.24%, 81.09%, 80.56%, 77.96%, and 78.91%). It has high and stable precision, recall, and F1-score values as well. This shows that SVM performs very well in handling oversampled data (SMOTE), possibly because SVM is able to cope well with minority classes. SVM is the best performing algorithm consistently on oversampled data using SMOTE, judging by the combination of accuracy, precision, recall, and F1-score on the test set. Logistic Regression can be used as an alternative if a simpler and easier to explain model is required. It is hoped that future research will consider additional features or the use of deep learning models to further enhance the accuracy of sentiment analysis.

REFERENCES

[1] D. P. Widiyani and J. S. S. Hartono, "Studi Eksplorasi Agroklimat Tanaman Kopi Robusta (Coffea canephora) Kabupaten Tanggamus, Lampung," *J. Agrinika J. Agroteknologi dan Agribisnis*, vol. 5, no. 1, p. 20, 2021, doi: 10.30737/agrinika.v5i1.1523.

[2] F. Fitriani, B. Arifin, and H. Ismono, "Indonesian coffee exports and its relation to global market integration," *J. Socioecon. Dev.*, vol. 4, no. 1, p. 120, 2021, doi: 10.31328/jsed.v4i1.2115.

[3] B. P. Statistik, *Statistik Kopi Indonesia 2020*. 2020.

[4] B. P. Statistik, *Statistik Kopi Indonesia 2021*. 2021.

[5] B. P. Statistik, *Statistik Kopi Indonesia 2022*. 2022.

[6] F. D. la P. Nhan Cach Dang, María N. Moreno-García, "Sentiment Analysis Based on Deep Learning :," *MDPI*, vol. 1, no. 1, pp. 1–29, 2020, doi: 10.3390/electronics9030483.

[7] J. Ilmiah and E. Islam, "Pemanfaatan Media Sosial Dalam Pengaruhnya Terhadap Pembentukan Persepsi dan Reputasi Wisata Halal di Indonesia," vol. 8, no. 03, pp. 3236–3248, 2022.

[8] P. Heydarian, A. Bifet, and S. Corbet, "Understanding market sentiment analysis : A survey," *Wiley*, pp. 1–23, 2024, doi: 10.1111/joes.12645.

[9] G. K. Patra, C. Kuraku, S. Konkimalla, V. Nagesh, B. Manikanth, and S. Kumar, "Journal of Artificial Intelligence & Cloud Computing A Sentiment Analysis of Customer Product Review Based on Machine Learning Techniques in E-Commerce," *J. Artif. Intell. Cloud Comput.*, vol. 2, no. 4, pp. 1–7, 2023, doi: 10.47363/JAICC/2023(2)389.

[10] N. Singh, "Sentiment Analysis Using Machine Learning: A Comparative Study," *Adv. Distrib. Comput. Artif. Intell. J.*, vol. 12, no. 1, 2023, doi: 10.14201/adcaij.26785.

[11] A. Samoggia, "Social media exploration for understanding food product attributes perception : the case of coffee and health with Twitter data," vol. 122, no. 12, pp. 3815–3835, 2020, doi: 10.1108/BFJ-03-2019-0172.

[12] K. Sharma, L. Hota, V. A. Tikkiwal, and A. Kumar, "ScienceDirect Exploring Exploring Twitter Twitter Sentiments Sentiments for for Predicting Predicting Match Match Outcomes Outcomes in in The The Game of Cricket Game of Cricket," *Procedia Comput. Sci.*, vol. 235, pp. 152–162, 2024, doi: 10.1016/j.procs.2024.04.018.

[13] Z. Firmansyah and N. F. Puspitasari, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes," *J. Tek. Inform.*, vol. 14, no. 2, pp. 171–178, 2021, [Online]. Available: https://doi.org/10.15408/jti.v14i2.24024

[14] Z. A. Sudianto Sudianto, Juan Arton Masheli, Nursatio Nugroho, Rafi Wika Ananda Rumpoko, "Jurnal teknik informatika," *J. Tek. Inform.*, vol. 15, no. 2, pp. 110–118, 2022, doi: 10.15408/jti.v15i2.24996.

[15] R. Watrianthos, M. Giatman, W. Simatupang, R. Syafriyeti, and N. K. Daulay, "Analisis Sentimen Pembelajaran Campuran Menggunakan

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.41316

Twitter Data," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 166, 2022, doi: 10.30865/mib.v6i1.3383.

[16] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences (Switzerland)*, vol. 13, no. 7. mdpi.com, Switzerland, pp. 1–21, 2023. doi: 10.3390/app13074550.

[17] R. Patel and K. Passi, "Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning," *Internet of Things*, vol. 1, no. 2, pp. 1–22, 2020, doi: 10.3390/iot1020014.

[18] M. Lamba and M. Madhusudhan, "Sentiment analysis," in *Text Mining for Information Professionals: An Uncharted Territory*, 1st ed., S. N. S. AG, Ed., University of Delhi, India: Springer, 2021, pp. 191–211. doi: 10.1007/978-3-030-85085-2_7.

[19] R. Astri, L. P. Hung, S. B. Sura, A. Kamal, and R. Yuliet, "Sentiment analysis using naive bayes for reviews of visitors to Padang City beach tourism after the COVID-19 pandemic," *E3S Web Conf.*, vol. 464, 2023,doi:10.1051/e3sconf/202346406002.

[20] E. M. Z. Darmawan and A. Fauzan Dianta, "Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM," *Teknol. J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 8–15, 2023.

[21] Y. Mao, Q. Liu, and Y. Zhang, "Journal of King Saud University - Computer and Sentiment analysis methods , applications , and challenges : A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, pp. 1–16, 2024, doi: 10.1016/j.jksuci.2024.102048.

[22] M. Shaden, A. Fadel, S. Achmad, and R. Sutoyo, "ScienceDirect ScienceDirect Sentiment analysis for customer review : Case study of Traveloka Sentiment analysis for customer review : Case study of Traveloka," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.