EDUMATIC Jurnal Pendidikan Informatika

e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 9 No. 1, April, 2025, Hal. 31-40
DOI: 10.29408/edumatic.v9i1.29120

Logistic Regression and Naïve Bayes Comparison in Classifying Emotions on Indonesian X Social Media

Gerald Shabran Rasyad [1] , Warih Maharani [2, *]

[1] Department of Informatics, Telkom University, Indonesia
[2] Data Science, Telkom University, Indonesia
[*] Correspondence: wmaharani@telkomuniversity.ac.id

Received: 30 December 2024 | Revised: 10 January 2025 | Accepted: 16 January 2025 | Published: 9 April 2025

Abstract
Emotions are integral to human interaction and decision-making, often expressed on social media platforms like X, which provides valuable data for sentiment analysis. However, analyzing texts from X poses challenges due to informal language, slang, and unique textual features. This study compares Logistic Regression and Naive Bayes in classifying emotions from Indonesian tweets, addressing gaps in prior research by exploring feature extraction methods, data split ratios, and hyperparameter tuning. Data were collected from 100 Telkom University students, resulting in 8,978 tweets labeled into four emotions: Happy, Sad, Angry, and Fear. After preprocessing, feature extraction methods TF-IDF and Bag of Words (BoW) were applied. Models were trained and tested on 10%, 20%, and 30% data splits, and performance was evaluated using accuracy, precision, recall, and F1-score. Hyperparameter tuning was conducted for Logistic Regression using GridSearch. Results showed Logistic Regression outperformed Naive Bayes, achieving 73.49% accuracy compared to 70.27%, with BoW yielding superior results over TF-IDF. The 20% data split provided the best balance for training and testing. This research demonstrates the effectiveness of Logistic Regression and highlights the importance of tailored feature extraction and parameter optimization for emotion classification in informal text datasets, particularly for Indonesian tweets.

Keywords: bow; emotion classification; logistic regression; naive bayes; tf-idf

INTRODUCTION

Emotion involves an aspect of human communication, interaction, and decision making (Vistorte et al., 2024). This can measure how people react to certain issues. By categorizing this emotion, it can reach a valuable insight for many fields such as customers feedback for commerce, public sentiment towards a topic discussed, and mental health monitoring for doctors and any other at the same field (Arias et al., 2022; Jim et al., 2024; Wankhade et al., 2022). At this age, people always express it on social media, make it easy and fast for the user. X is the most popular social media to do this activity (Irmayani et al., 2021; Zharifa & Ujianto, 2024).

X as a social media platform takes a role to be a resource for studying emotion due to its use informal language, slang, hashtags, and any other unique textual feature aspect that present a challenge in analyzing its text (Haya et al., 2024; Mao et al., 2024; Nip & Berthelier, 2024). The emotions conveyed through social media text can sometimes be unclear, as the emotions written do not always reflect the actual emotional state (Holtgraves, 2022). Additionally, factors such as the use of informal language can complicate the intended meaning of the emotion in the text (Naveen & Trojovský, 2024). Emotion classification on social media is important to understand the true expressions felt by the user when writing the post. (Basha et al., 2021). Therefore, machine learning is important to classify emotions through the text of social media posts (Kabir & Madria, 2021).

A method to classify these emotions are conducted with 2 approach using machine learning models that is Logistic Regression and Naive Bayes. Both methods chosen due to their established efficiency and simplicity in text classification tasks (Pintas et al., 2021). Naïve Bayes commonly known for its simplicity making it easy to implement and efficient in computation (Armansyah & Ramli 2022; Ismail et al., 2020; Hendrawan et al., 2022; Sihombing et al., 2021). However, this method is less effective when dealing with datasets that has too many features, which can cause its accuracy to decrease (Riani et al., 2023). Compared to Naïve Bayes, Logistic Regression can handle complex data and its flexibility to model complex relationship between features and output classes (Balboa et al., 2024). Due to complex handling data, Logistic Regression can get a better accuracy than Naïve Bayes (Wahyuningsih et al., 2024). Although Logistic Regression gained a bigger accuracy. Naive Bayes excels at handling large text classification datasets with independent features, enabling rapid processing efficiently while minimizing the need for extensive resources (Bahtiar et al., 2023).

Research conducted by Ramadhani & Suryono (2024) that compares Logistic Regression and Naïve Bayes, and resulting accuracy Logistic Regression 95% and Naïve Bayes 91% from 5799 data using TF-IDF feature extraction. Their research achieve that Logistic Regression gained a better accuracy than Naïve Bayes. Furthermore, research by Toyibah et al. (2024) also compared Logistic Regression and Naïve Bayes also gained accuracy 89% on Logistic Regression and 85% on Naïve Bayes using 40,000 datasets also using TF-IDF. Their research also achieve that Logistic Regression is better than Naïve Bayes. Both Research used TF-IDF because of its ability to highlight word with relevance in specific document but rarely found across the entire corpus, this also reduce the word that is less informative and provide meaningful representation of the text to enhance method in identifying emotional patterns. However, gaps can be seen by their research. The research by Ramadhani & Suryono (2024) only use TF-IDF as feature extraction making it unclear how the models would perform with other feature extraction. Additionally, Toyibah et al. (2024) used a 20:80 data split, but the rationale behind this choice and its potential impact on model performance was not explicitly discussed. Both of their research did not address the parameter to gain the best performance of the method.

Previous research has shown that Logistic Regression outperforms Naïve Bayes in emotion classification, with TF-IDF as the sole feature extraction technique. However, these studies did not explore alternative feature extraction methods, the impact of different data splits, or the role of hyperparameter optimization. To address these gaps, this research aims to compare Logistic Regression and Naïve Bayes using various feature extraction techniques, such as TF-IDF and Bag of Words, analyze the effects of different data split ratios on model performance, and optimize hyperparameters to achieve the best possible results. By fulfilling these objectives, this study seeks to provide deeper insights into emotion classification, which is crucial for understanding public reactions to issues like the implementation of a 12% VAT, offering valuable guidance for decision-making and policy evaluation.

METHOD

Figure 1 illustrates the research begins with data collection, labelling, preprocessing, splitting, method testing, and evaluation. The data is gathered from Telkom University students who have X social media accounts. The collected data is then processed through data labelling stages, such as assigning emotion labels, followed by preprocessing to convert raw text into a format understandable by the model. The data is then split into training and testing sets and applied to two methods: Logistic Regression and Naïve Bayes.

The data for this research was collected from X social media accounts of Telkom University students. Respondents were gathered between July and September 2024, with 100 students aged 18 to 22 participating. These students agreed to provide their usernames for sampling. The tweet data was collected over one month, from October 1 to October 30, 2024. This dataset forms the basis for the analysis in this research.

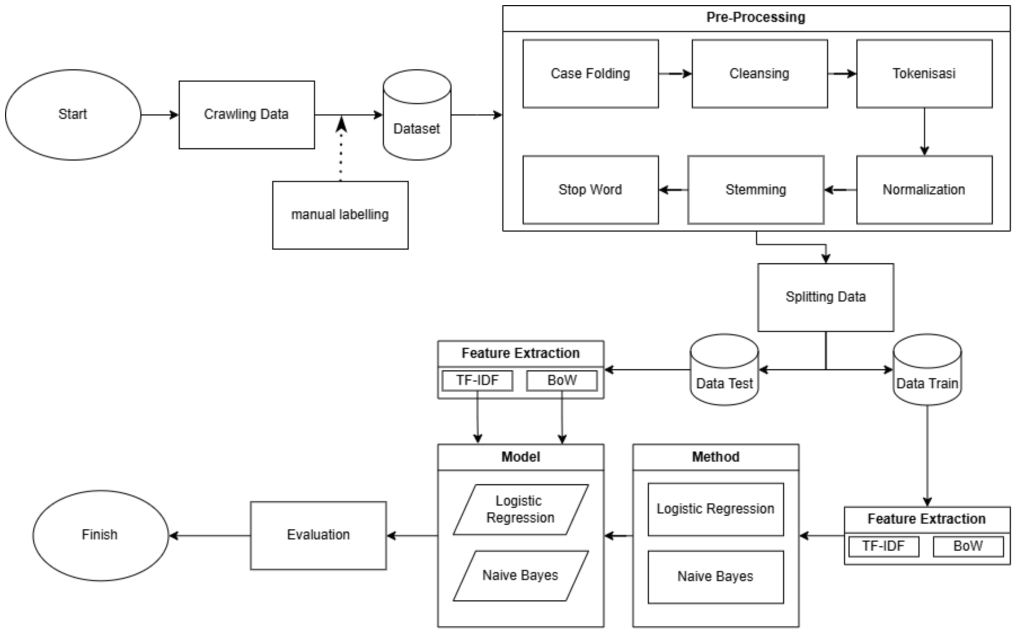

Figure 1. Flowchart research

The labelling process was conducted manually by annotators fluent in Indonesian, aiming to classify the emotions contained in the text into four categories: Happy, Angry, Sad, and Fear. During this process, the annotators not only understood the context of the sentences but also considered punctuation and the implied meaning of the text to ensure more consistent and accurate labelling, resulting in a reliable dataset. For example, the sentence "*aku benar-benar marah karena sikapmu itu*" would be labelled as Angry, while "*saya merasa sedih karena kehilangan itu*" would be labelled as Sad. The final label for each tweet was determined based on the majority vote. If significant disagreements arose, further discussions were conducted until annotators reached an agreement.

After the data was labelled, preprocessing was done to turn the raw data into structured data the machine can understand and to make sure it is consistent across all the data. The steps included case folding to make all the letters lowercase, cleaning to remove irrelevant words such as URLs, symbols, and numbers, tokenization to split sentences into individual words, normalization to change words into their standard forms, stemming to reduce words to their root forms, and stop word removal to remove words that don't have significant meaning, such as "*dan*", "*atau*", "*dari*", "*untuk*", and etc.

The feature extraction methods used in this research are TF-IDF and BoW due to their fast computation ability, simplicity, and effectiveness. These feature extraction techniques can capture patterns in the dataset. Additionally, these methods have provided relevant results for model evaluation. TF-IDF determining how important the word is in a document and BoW calculating the frequency of word occurrences across all processed documents. Both are feature extraction methods as they effectively represent text data.

Multinomial Naïve Bayes chosen because efficient for large text datasets with independent features, making it suitable for emotion classification. Logistic Regression, with its ability to model complex relationships, often achieves higher accuracy on intricate patterns.

The model's performance is assessed using metrics such as accuracy, precision, recall, and F1-score. Evaluating these metrics offers valuable insights into the effectiveness of the model in classifying emotions.

RESULT AND DISCUSSION

Result

The dataset used in this research was collected and comprises a total of 8,978 rows of data was retrieved using the X API through a systematic crawling process. After that the data labelled by annotator that is proficient in Indonesia. The result is categorized into 4 emotion namely Happy, Angry, Sad, and Fear. Figure 2 shows the distribution of tweets across the four emotion labels: Happy, Angry, Sad, and Fear. The data shows that Happy is the most frequently occurring emotion, with 2,723 tweets, accounting for the largest portion of the dataset. This is followed by Angry, with 2,492 tweets, and Sad, with 2,001 tweets. Fear is the least represented emotion, comprising 1,762 tweets. The results of the manual text labelling on each emotion namely Happy, Angry, Sad, and Fear is shown in table 1.

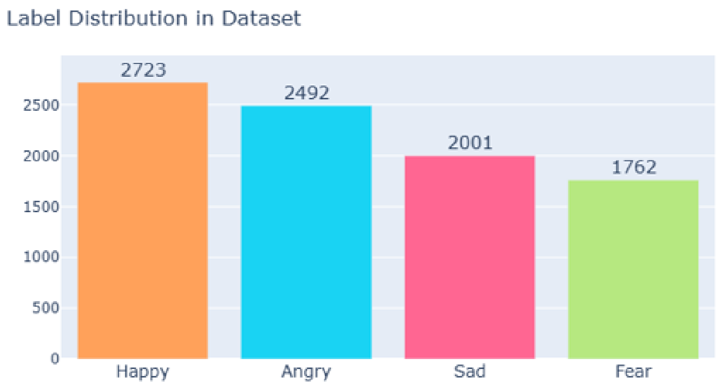

Figure 2. Labels and tweets

Table 1. Labels and tweets

Label	Tweets
Happy	@bashfunction RANDOM BANGET TIBA2 NGUCAPIN wkwk makasi yeaa
Angry	*Gue tinggal makan ada ajaaa masalahnya anjir* https://t.co/ausMfbscZB
Sad	*Ga mood kuliah habisa dibikin sedih webtoon*
Fear	*@kuliahturu akuu juga mau buat yang kek gini tapi Takut salah paham*

To ensure that the data is in a suitable format for emotion classification, preprocessing is necessary. In the preprocessing stage, the text undergoes several steps. First, in the case folding stage, all characters are converted to lowercase. Then, during cleansing, the text is cleaned by removing URLs, mentions, hashtags, and symbols. Tokenization follows, breaking the text into separate tokens. Normalization standardizes slang into formal language. Stemming reduces words to their root form, minimizing variations of the same word. Lastly, stopword removal is performed using the NLTK library and an Indonesian stopwords list, ensuring that only meaningful keywords remain, resulting in a simplified text ready for classification. The process of each step shown in table 2.

Table 2 presents the preprocessing results applied to the example sentence "*Kenapa sih kamu selalu terlambat? Ini benar-benar bikin kesel banget!!!*". The first column lists the preprocessing techniques, while the second column provides the resulting sentence at each step. Starting with Case Folding, the text was converted to lowercase without removing punctuation. In Cleaning, unnecessary characters such as punctuation were removed. Tokenization splits the sentence into individual words or tokens. Normalization standardizes informal words to their formal equivalents (e.g., "*bikin*" to "*membuat*"). Stemming reduces words to their root forms, such as "*terlambat*" to "*lambat*". Finally, Stopword removal eliminates irrelevant words, leaving only key terms like "*kenapa*", "*lambat*", and "*kesal*" to enhance classification accuracy.

Table 2. Preprocessing table result

Preprocessing Techniques	Sentence
Case Folding	kenapa sih kamu selalu terlambat? ini benar-benar bikin kesel banget!!!
Cleaning	kenapa sih kamu selalu terlambat ini benar benar bikin kesel banget
Tokenization	["*kenapa", "sih", "kamu", "selalu", "terlambat", "ini", "benar", "benar", "bikin", "kesel", "banget*"]
Normalization	["*kenapa", "sih", "kamu", "selalu", "terlambat", "ini", "benar", "benar", "membuat", "kesal", "sekali*"]
Stemming	["*kenapa", "sih", "kau", "selalu", "lambat", "ini", "benar", "benar", "buat", "kesal", "sekali*"]
Stopword	["*kenapa", "kau", "lambat", "benar", "buat", "kesal", "sekali*"]

After the data undergoes preprocessing, it will be tested using different splits starting from 10% (898 data test and 8,808 data train), 20% (1,796 data test and 7,182 data train), and 30% (2,694 data test and 6,284 data train), utilizing the scikit-learn library and a fixed random state of 42 to ensure consistency in the data. Additionally, this evaluation will compare TF-IDF and BoW feature extraction methods. The goal of this experiment is to determine which data split and feature extraction method yield the best results. The result of Logistic Regression shown in table 3 and Naive Bayes shown in table 4.

Table 3. Split test result on logistic regression

Data Split	Feature Extraction	Accuracy	Precision	Recall	F1-Score
10%	TF-IDF	71.27%	0.72	0.70	0.71
	BoW	71.63%	0.72	0.70	0.71
20%	TF-IDF	71.59%	0.71	0.71	0.71
	BoW	71.80%	0.73	0.70	0.71
30%	TF-IDF	70.86%	0.71	0.70	0.70
	BoW	71.42%	0.72	0.70	0.71

From the table 3 and 4, at 20% data split ratio, the model receives a sufficient amount of training data, allowing it to learn patterns more effectively while still maintaining enough test data for a representative evaluation. This creates a balanced trade-off between training and testing. At a 10% ratio, performance is lower due to the reduced amount of test data, which limits the model's ability to evaluate its generalization capabilities. Conversely, with a 30% ratio, the model is trained on less data, which limits its ability to capture more complex patterns and nuances in the dataset.

It also shown from table 3 and 4, Bag of Words excels TF-IDF due to its ability to calculate the frequency of word occurrences in a document without considering the importance of the word. In the context of Indonesian tweets for emotion classification, BoW effectively capture the presence of specific words that are strongly associated with emotional expressions. Given that many emotional terms, such as "*senang*", "*marah*", or "*sedih*", frequently appear across multiple tweets, BoW is able to assign high importance to these words, making it well-suited for identifying distinct emotions. Its performance is particularly notable in datasets where the frequency of certain words is a key feature for classification. Unlike TF-IDF, which downplays the importance of frequently occurring words, BoW does not consider the corpus-wide importance of terms, making it more effective when the dataset has clear, frequent words tied to specific emotional states. This straightforward approach results in BoW generating explicit features directly related to the words that appear more often, which aids in better emotional classification performance, especially for common emotions expressed in the dataset.

Table 4. Split test result on naive bayes

Data Split	Feature Extraction	Accuracy	Precision	Recall	F1-Score
10%	TF-IDF	69.49%	0.70	0.68	0.69
	BoW	70.04%	0.71	0.69	0.70
20%	TF-IDF	68.65%	0.69	0.67	0.68
	BoW	70.27%	0.71	0.69	0.70
30%	TF-IDF	68.34%	0.69	0.67	0.68
	BoW	68.67%	0.69	0.68	0.67

Based on the conducted testing on table 3 and 4, the configurations used were 20% data split and Feature Extraction using Bag of Words (BoW). With these two configurations, the next step is to perform Hyperparameter Tuning. It aims to maximize the performance of the model, specifically for the Logistic Regression method, using the parameters listed in Table 5. Regularization values ranged from 0.01 to 10, the lower the value means the model prioritize generalization and can underfit the data. Higher value means prioritize fitting the training data, it can also lead to overfit. This process utilizes GridSearch to optimize the model's accuracy.

Table 5. Parameter tuning to test

Model	Parameter	Values Tested
Logistic Regression	C (Regularization)	[0.01, 0.1, 1,10]
	Solver	[‘newton-cg’,’lbfgs’,’sag’, ‘saga’]

Table 6. Tuned parameter result

Method	Evaluation			
	Accuracy	Precission	Recall	F1-Score
Logistic Regression	73.49%	0.74	0.72	0.73
Naive Bayes	70.27%	0.71	0.69	0.70

Table 6 shows that the parameters improve the accuracy of the Logistic Regression (LR) with the selected parameters solver of lbfgs and regularization set to 1. lbfgs is a suitable parameter for multiclass datasets and is efficient for large data. Additionally, the regularization value of 1 indicates a balance between fitting the training data, preventing both overfitting and underfitting. The accuracy achieved by Logistic Regression is better with a value of 73.49%, compared to Naive Bayes which only achieved 70.27%. Logistic Regression is superior to Naive Bayes in emotion classification because it can model complex relationships between features and the target variable, resulting in better performance, especially in multiclass scenarios. Additionally, LR allows for fine-tuning through regularization, which helps prevent overfitting and improves generalization to unseen data. The ROC and AUC result is shown on figure 3.

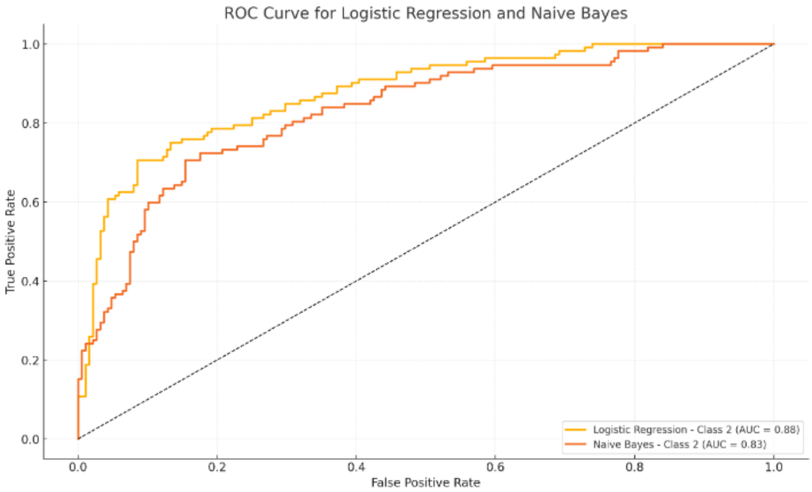

Figure 2. Roc and auc result

The ROC curve compares the performance of Logistic Regression and Naive Bayes for a classification task. The yellow curve, representing Logistic Regression, has a higher area under the curve (AUC = 0.88) compared to the orange curve for Naive Bayes (AUC = 0.83). This indicates that Logistic Regression is better at distinguishing between classes across various decision thresholds. A higher AUC means that Logistic Regression achieves a higher true positive rate (TPR) while maintaining a lower false positive rate (FPR) compared to Naive Bayes. In practical terms, Logistic Regression is more effective at making correct predictions and is more reliable for this dataset.

Discussion

This research provides valuable insights into the comparative performance of Logistic Regression and Naive Bayes for emotion classification on Indonesian tweets. The experimental results reveal that Logistic Regression achieved superior accuracy, with a value of 73.49%, compared to Naive Bayes' 70.27%. These results were obtained using the Bag of Words (BoW) feature extraction method and a 20% train-test split. Logistic Regression also demonstrated higher effectiveness in distinguishing between emotion classes, as evidenced by a higher area under the ROC curve (AUC = 0.88) compared to Naive Bayes (AUC = 0.83).

The research highlights the importance of feature extraction methods tailored to the linguistic properties of Indonesian tweets. The BoW approach proved more effective than TF-IDF, as it emphasizes frequently occurring emotional keywords, such as "*senang*," "*marah*," and "*sedih*," which are crucial for emotion classification. This focus on high-frequency terms allowed BoW to better capture key patterns in the dataset, enhancing classification performance. Additionally, the optimal 20% data split provided a balance between sufficient training data and a representative test set, enabling robust evaluation of the models.

Hyperparameter tuning further optimized Logistic Regression, striking a balance between generalization and performance. By fine-tuning the regularization and solver parameters, the model's predictive accuracy and reliability improved significantly. These findings reinforce the effectiveness of Logistic Regression for handling complex relationships in multiclass classification tasks, particularly for informal text datasets like social media posts.

Logistic Regression consistently outperformed Naive Bayes in this research due to its ability to model complex relationships between features and target variables. Its flexibility and higher AUC values make it more reliable for multiclass classification tasks. In contrast, Naive Bayes' simpler approach is less suited for datasets requiring nuanced interpretations of emotional text. These findings underscore Logistic Regression's broader applicability in handling datasets with informal and variable linguistic structures. The results emphasize its value for tasks such as emotion classification, particularly in contexts involving multilingual and informal text datasets.

The results of this research align with prior research, such as Ramadhani & Suryono (2024), which also reported better performance for Logistic Regression compared to Naive Bayes in sentiment and emotion classification tasks. For instance, achieved 95% accuracy for Logistic Regression and 91% for Naive Bayes using TF-IDF, while Toyibah et al. (2024) reported 89% and 85% accuracy for the respective models on a larger dataset. Both studies demonstrated the superior accuracy of Logistic Regression, but they did not explore alternative feature extraction methods or assess the impact of different data split ratios.

This study addresses these gaps by incorporating both TF-IDF and BoW feature extraction methods, as well as evaluating the effects of 10%, 20%, and 30% data splits on model performance. The findings indicate that BoW outperforms TF-IDF in this context, and a 20% data split provides the best balance for model training and evaluation. Additionally, hyperparameter tuning, which was not discussed in prior studies, significantly improved the performance of Logistic Regression. These methodological advancements enhance the novelty of this research and offer deeper insights into emotion classification on social media datasets.

The practical implications of this research are significant, especially for real-time sentiment monitoring and public opinion analysis. Enhanced accuracy and AUC in classifying emotions provide actionable insights for decision-makers and policymakers. For example, the ability to classify emotions in Indonesian tweets can help understand public reactions to policies or societal events. This capability supports more effective decision-making and policy evaluation. Overall, the findings highlight the potential of tailored machine learning approaches in addressing real-world challenges in emotion classification.

CONCLUSION

This research confirms that Logistic Regression is more effective than Naive Bayes in identifying emotional expressions within Indonesian tweets. The findings highlight the importance of employing appropriate data representation techniques that cater to the linguistic nuances of informal text. This research demonstrates that customizing machine learning workflows such as feature extraction methods, data splitting strategies, and parameter optimization leads to improved performance in multiclass classification tasks. By achieving accurate and reliable emotion detection, this work provides a framework that can be adapted for analyzing other informal text datasets. The results have practical implications, particularly for applications such as public opinion monitoring, decision-making processes, and understanding societal trends through emotion analysis.

REFERENCES

Arias, F., Zambrano Nunez, M., Guerra-Adames, A., Tejedor-Flores, N., & Vargas-Lombardo, M. (2022). Sentiment Analysis of Public Social Media as a Tool for Health-Related Topics. *IEEE Access*, *10*, 74850–74872. https://doi.org/10.1109/ACCESS.2022.3187406

Armansyah, A., & Ramli, R. K. (2022). Model prediksi kelulusan mahasiswa tepat waktu dengan metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 1-10. https://doi.org/10.29408/edumatic.v6i1.4789

Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, *5*(3), 915–927. https://doi.org/10.51519/journalisi.v5i3.539

Balboa, A., Cuesta, A., González-Villa, J., Ortiz, G., & Alvear, D. (2024). Logistic regression vs machine learning to predict evacuation decisions in fire alarm situations. *Safety science*, *174*, 106485. https://doi.org/10.1016/j.ssci.2024.106485

Basha, S. R., Rao, M. S. B., Reddy, P. K. K., & Kumar, G. R. (2021). Emotional Tweets Analysis on Social Media with Short Text Classification Using Various Machine Learning Techniques. *Journal of Computational and Theoretical Nanoscience*, *17*(12), 5477–5482. https://doi.org/10.1166/jctn.2020.9442

Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of naïve bayes algorithm and XGBoost on local product review text classification. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 143-149. https://doi.org/10.29408/edumatic.v6i1.5613

Holtgraves, T. (2022). Implicit communication of emotions via written text messages. *Computers in Human Behavior Reports*, *7*, 1-8. https://doi.org/10.1016/j.chbr.2022.100219

Irmayani, D., Edi, F., Harahap, J. M., Rangkuti, R. K., Ulya, B., & Watrianthos, R. (2021, June). Naives Bayes algorithm for twitter sentiment analysis. In *Journal of Physics: Conference Series* , 1933(1), 1-6. IOP Publishing. https://doi.org/10.1088/1742-6596/1933/1/012019

Ismail, M., Hassan, N., & Bafjaish, S. S. (2020). Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task. *Journal of Soft Computing and Data Mining*, *1*(2), 1–10.

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Kabir, M. Y., & Madria, S. (2021). EMOCOV: Machine learning for emotion detection, analysis and visualization using COVID-19 tweets. *Online Social Networks and Media*, *23*. https://doi.org/10.1016/j.osnem.2021.100135

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, *36*(1), 102048. https://doi.org/10.1016/j.jksuci.2024.102048

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *Iscience*, *27*(10), 110878. https://doi.org/10.1016/j.isci.2024.110878

Nip, J. Y. M., & Berthelier, B. (2024). Social Media Sentiment Analysis. *Encyclopedia*, *4*(4), 1590–1598. https://doi.org/10.3390/encyclopedia4040104

Pintas, J. T., Fernandes, L. A. F., & Garcia, A. C. B. (2021). Feature selection methods for text classification: a systematic literature review. *Artificial Intelligence Review*, *54*(8), 6149–6200. https://doi.org/10.1007/s10462-021-09970-6

Ramadhani, B., & Suryono, R. R. (2024). Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse. *Jurnal Media Informatika Budidarma*, *8*(2), 714-725. https://doi.org/10.30865/mib.v8i2.7458

Riani, A. P., Sulistyowati, N., Ridwan, T., & Voutama, A. (2023). Klasifikasi Emosi Publik Terhadap Larangan Penggunaan Obat Sirup Menggunakan Algoritma Naive Bayes. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, *7*(2), 325-339. https://doi.org/10.46880/jmika.Vol7No2.pp325-339

Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, *5*(2), 233-242. https://doi.org/10.29408/edumatic.v5i2.4089

Toyibah, Z. B., Putri, Y. N., Puandini, P., Widodo, Z. M., & Ni'mah, A. T. (2024). Perbandingan Kinerja Algoritma Multinomial Naïve Bayes dan Logistic Regression pada Analisis Sentimen Movie Ratings IMDB. *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, *10*(2), 181–189. https://doi.org/10.21107/edutic.v10i2.28150

Vistorte, A. O. R., Deroncele-Acosta, A., Ayala, J. L. M., Barrasa, A., López-Granero, C., & Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. *Frontiers in psychology*, *15*, 1-13. https://doi.org/10.3389/fpsyg.2024.1387089

Wahyuningsih, T., Manongga, D., Sembiring, I., & Wijono, S. (2024). Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments. *Procedia Computer Science*, *234*, 349–356. https://doi.org/10.1016/j.procs.2024.03.014

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Zharifa, A. H. A., & Ujianto, E. I. H. (2024). Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 754-763. https://doi.org/10.29408/edumatic.v8i2.28048