SENTIMENT ANALYSIS ON ARTIFICIAL INTELLIGENCE TECHNOLOGY USING NAIVE BAYES CLASSIFIER

Lintang Dwi Cahya[1], Adityo Permana Wibowo[2]
[1]Informatics, University of Technology Yogyakarta, Yogyakarta, Indonesia
[2]Information System, University of Technology Yogyakarta, Yogyakarta, Indonesia
Email: lintangcahya.lc@gmail.com, adityopw@uty.ac.id

Abstract: *The advancement of Artificial Intelligence (AI) technology has significantly impacted various aspects of life. However, these technological developments often elicit diverse responses from the public, including enthusiasm and concern. This study aims to analyze public sentiment toward AI developments using the Naive Bayes Classifier algorithm. The research collected AI-related tweets through data crawling, preprocessing, sentiment labeling, and feature extraction using TF-IDF. To address the class imbalance in the data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The Naive Bayes Classifier model was then used to classify public sentiment into positive, negative, and neutral categories. Evaluation results indicate that the Naive Bayes Classifier achieved an overall accuracy of 70%, with a precision of 72%, recall of 70%, and F1-score of 71%. Although the model is relatively effective in identifying positive sentiments, challenges remain in distinguishing negative and neutral sentiments accurately. Factors affecting model performance, such as data preprocessing quality and limited dataset diversity, are discussed as areas for improvement in future research.*

Keywords: *Artificial Intelligence, Naïve Bayes Classifier, Public Opinion, Sentiment Analysis, Text Mining*

INTRODUCTION

In the era of rapidly advancing technology, Artificial Intelligence (AI) has brought significant changes to various aspects of life. It enables the automation of complex processes and creates opportunities in data analysis. One important application of data analysis is sentiment analysis, which examines public opinion and uncovers insights into societal views and feelings on specific topics or issues. (Liu, 2020). Through sentiment analysis, organizations, companies, and policymakers can gain valuable insights into public perception, particularly regarding technological advancements like AI, which often generate diverse responses ranging from enthusiasm to concern.

Sentiment analysis, or opinion mining, is a technique for extracting, interpreting, and measuring opinions and emotions from human-generated text. This method enables the identification of attitudes, evaluations, emotions, and opinions toward specific entities within the text, playing a crucial role in understanding societal mindsets. (Medhat et al., 2014). In the context of AI advancements, sentiment analysis can provide valuable insights into how society perceives this technology, ultimately aiding policymakers and researchers in addressing ethical, social, and economic challenges associated with its application. (Ullah et al., 2021).

The Naïve Bayes Classifier method has proven effective in text classification tasks, including sentiment analysis, due to its simplicity, efficiency, and strong accuracy, particularly when working with large datasets. (Sailunaz & Alhajj, 2019). Naïve Bayes has been widely used in research focused on text and sentiment classification and can deliver robust results on large datasets because of its efficient probabilistic approach. (Joshi et al., 2010). This method is highly suitable for identifying and classifying positive, negative, and neutral sentiments regarding AI, given the need to understand sentiment context more deeply.

By analyzing public opinion on AI, this study aims to explore positive, negative, and neutral societal attitudes toward technology. The findings of this research are expected to provide a solid foundation for the development of policies and AI implementation strategies that are more responsive to public needs.

METHOD

The stages undertaken in this research include data collection, data cleaning, data preprocessing, feature extraction, modeling, and evaluation, as illustrated in Figure 1.

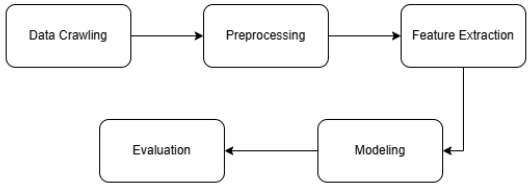

Figure 1. Research Flow

A. Data Crawling

Data was collected from the social media platform X using the tweet-harvest tool, which extracts tweets based on keywords related to advancements in artificial intelligence technology. In this process, multiple keywords were used, such as "Artificial Intelligence," "AI technology advancements," "AI in everyday life," and other related terms. As a result, a total of 2,514 tweets were successfully gathered for further analysis.

B. Preprocessing

Preprocessing was conducted to transform raw data into clean data, where text data was standardized in form and format to prepare it for further processing stages. This process includes cleaning, case folding, tokenization, stemming, and filtering. (Darwis et al., 2021). The cleaning step removes irrelevant characters or symbols; case folding standardizes text case; tokenization breaks down text into words; stemming reduces words to their root form; and filtering removes meaningless words, producing consistent text data that is ready for the next steps. (Aggarwal, 2015).

After preprocessing, sentiment labeling was conducted using TextBlob, a Python library for natural language processing capable of classifying text into positive, neutral, or negative sentiment. TextBlob has proven effective for automatic sentiment labeling in natural language data, providing reasonably accurate results in initial sentiment classification. (Bird et al., 2009).

C. Feature Extraction

Following preprocessing, feature extraction was conducted using Term Frequency-Inverse Document Frequency (TF-IDF). In this process, words frequently occurring in the document are weighted. (Hasibuan & Serdano, 2022). Thus, each tweet is transformed into a feature vector representing the weight of each word in the text, facilitating sentiment classification. (Robertson, 2004).

To address the class imbalance in the dataset before classification, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. This method generates synthetic samples of the minority class by interpolating between existing samples, resulting in a more balanced class distribution. (Chawla et al., 2002). SMOTE has been shown to enhance model sensitivity to the minority class, reduce bias toward the majority class, and help prevent overfitting (Han et al., 2005).

D. Naive Bayes Classifier

In the classification phase, the Naïve Bayes Classifier algorithm was used to predict public sentiment regarding AI advancements. Naïve Bayes Classifier, a classification algorithm based on Bayes' probability theory, is known for its simplicity and effectiveness, especially in text classification tasks like sentiment analysis. (Aydogan & Akcayol, 2016). This algorithm assumes that each feature in the dataset is independent of others (naïve assumption), simplifying the calculation of class probabilities. (Jordan & Mitchell, 2015).

This model is highly suitable for classifying text data involving numerous words or terms, particularly with large-scale datasets, as it can produce accurate classification results with efficient processing time (Yadav & Vishwakarma,

2020). Additionally, to enhance model performance, hyperparameter tuning was performed using grid search to determine optimal parameters, such as smoothing values for the word probability distribution, especially for words with low occurrence in the documents (Rennie et al., n.d.).

E. Evaluation

Model evaluation was performed to assess the performance of the Naïve Bayes Classifier in classifying public sentiment. The evaluation metrics used include accuracy, precision, recall, and F1-score, each providing a different perspective on model prediction accuracy. (Powers & Ailab, n.d.). Accuracy measures the proportion of correct predictions, precision calculates accuracy for correctly predicted positive cases, recall assesses correct detection of all true positive cases, and F1-score balances precision and recall.

Thus, employing various evaluation metrics provides a comprehensive view of the model's ability to classify sentiment accurately and without bias toward any class. This evaluation aims to provide valid and reliable results as a basis for public sentiment analysis on AI technology advancements.

RESULTS AND DISCUSSION

This research aims to develop a sentiment analysis program capable of identifying public perspectives on Artificial Intelligence (AI) technology. Using a dataset of 2,244 records obtained through data crawling from Twitter, this analysis includes sentiment categories of positive, negative, and neutral. The findings of this research are expected to provide insights for researchers and developers in determining the direction of AI technology development.

The methodology applied in this research consists of several stages, including data cleaning and preprocessing, sentiment labeling using TextBlob, and feature extraction with Term Frequency-Inverse Document Frequency (TF-IDF). Additionally, data balancing is performed through the Synthetic Minority Over-sampling Technique (SMOTE), and the classification model used is Multinomial Naive Bayes with hyperparameter tuning via GridSearch. This approach is expected to produce results that more accurately represent public views on AI technology.

A. Data Crawling

The crawling process was conducted to collect a dataset containing public opinions on technological advancements, specifically in the field of Artificial Intelligence (AI). The primary goal of this data collection is to perform sentiment analysis on public views regarding the development and impact of AI technology.

Data was collected using a tool named tweet-harvest, allowing researchers to access and download tweets from the social media platform Twitter. Various keywords were used in the search process, including: "Artificial Intelligence," "AI Technology," "AI Development," "Machine Learning," "AI Impact," and "Latest AI Innovations." The use of diverse keywords aims to capture a wide spectrum of public opinion on AI and related technologies.

The data collection process was time-consuming due to Twitter API limitations, which restrict the amount of data retrievable within a specific time frame. Examples of collected tweets are shown in Table 1.

Table 1. Tweet Data

NO	Tweet
1	Maybe read this ss again. Penggunaan AI itu harmful banget karena nantinya bisa digunakan buat hal berbahaya termasuk deepfake (mengubah foto atau video muka/badan kalian sebagai orang lain untuk bikin false information) propaganda dan bahkan revenge porn. Kalian ga takut kah?
2	Kata Dharma AI itu artificial intelligence. Dari kata intelijen alias mata-mata. Jadi AI itu alat mata-mata buatan untuk mengawasi kita. Problem

	Jakarta yang kompleks dihadapi dengan teori konspirasi
3	Yg jadi problem dari AI adalah kecurangan dan pelanggaran etika yg menyepelekan nilai orisinalitas. Mangkanya senimanlah yg pertama kali paling keras protes. Itu krn seniman sangat menghargai proses dan hasil itu ujung2nya cuma target yg dilupakan stlh dicapai.
	……..
2243	Artificial Intelligence (AI) adalah teknologi yang memungkinkan mesin dan sistem untuk meniru kecerdasan manusia. Dengan kemampuan belajar menganalisis data dan membuat keputusan AI telah menjadi revolusi dalam berbagai industri. https://t.co/bGRjkGGpWq
2244	@mithamiwuwu @0xhujan Nah ini masalahnya. Regulasi AI sepertinya masih panjang tapi dampaknya udah besar. Ya bener akhirnya kita yg harus pinter pake etika buat gunain AI ini.

B. Preprocessing

The preprocessing stage is conducted to clean raw text data, ensuring it is ready for sentiment analysis. This step is crucial for enhancing data quality and improving model accuracy in detecting sentiment patterns within the text. The preprocessing steps applied include:

1. *Handling Missing Values*: The first step involves examining the dataset for any missing values in each attribute. If missing data is found, it is either removed or filled with appropriate values to avoid impacting the analysis results.
2. *Cleaning*: This step involves cleaning the data by removing irrelevant characters or symbols such as punctuation marks, emojis, links, and other special characters using RegEx. This ensures that the data is free from elements that hold no meaning in sentiment analysis.
3. *Case Folding*: All text is converted to lowercase to standardize word format, so "AI" and "ai" are treated the same by the model. Case folding reduces word variations caused by differences in capitalization.
4. *Tokenization*: Text is broken down into individual words (tokens), allowing the model to analyze each word separately. Tokenization is essential for isolating words to facilitate more accurate sentiment classification.
5. *Stemming*: Using the Sastrawi stemmer, words are reduced to their root form. For instance, “mengembangkan” becomes “kembang.” Sastrawi stemmer, specifically designed for the Indonesian language, helps maintain word consistency.
6. *Filtering*: Common words such as "dan," "yang," and "atau" are removed to focus on more relevant words, enhancing the analysis by reducing noise.
7. *Translating and Labeling*: After text processing, the data is translated into English and labeled using TextBlob. TextBlob classifies the text as positive, neutral, or negative sentiment based on the sentence context.

The results of the preprocessing steps can be seen in Table 2.

Table 2. Preprocessing Result

Clean	English	Label
maybe read this ss again guna ai harmful banged bahaya deepfake ubah foto video mukabadan orang bikin false information propaganda revenge porn ga takut kah	maybe read this ss again using ai is very harmful, the dangers of deepfake are changing photos, videos, people's faces, making false information, propaganda, revenge porn, aren't you afraid?	negative
dharma ai artificial intelligence intelijen alias matamata ai alat matamata buat awas problem jakarta	dharma ai artificial intelligence intelligence alias spy ai spy tool to watch out for complex Jakarta problems facing	negative

kompleks hadap teori konspirasi	conspiracy theories	
yg problem ai curang langgar etika yg sepele nilai orisinalitas mangkanya seniman yg kali keras protes krn seniman harga proses hasil ujung target yg lupa stlh capai	The problem is that AI cheats and violates trivial ethics, the value of originality, that's why artists often protest loudly because artists pay for the process and the end result, the target is forgotten after it is achieved.	positive
……..		
regulasi ai dampak udah ya bener yg pinter pake etika gunain ai	The impact of AI regulations is correct for those who are smart in using ethics when using AI	positive

C. Feature Extraction

In this study, feature extraction was performed using the Term Frequency-Inverse Document Frequency (TF-IDF) method to determine the weight of each word in the document. TF-IDF highlights significant words, giving higher weights to rare yet contextually relevant words in sentiment analysis by transforming them into feature vectors. TF-IDF consists of two main components: Term Frequency (TF) and Inverse Document Frequency (IDF). The formula for TF can be seen in Equation 1:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

TF measures how often a word appears in a particular document, calculated by dividing the occurrence of word t in document d by the total number of words in d. The IDF formula is shown in Equation 2:

$$IDF(t) = \log\left(\frac{N}{n_t}\right) \quad (2)$$

IDF measures the rarity of a word across the entire document set. For a word t that appears in N documents, it is calculated by dividing the total number of documents by the number of documents containing t. The TF and IDF components are then combined to obtain the TF-IDF weight of word t in document d, as shown in Equation 3:

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \quad (3)$$

TF-IDF is obtained by multiplying TF and IDF, thereby giving higher weight to words that appear frequently in a single document but rarely in others, marking them as unique and informative.

In this study, TF-IDF parameters were configured using the TfidfVectorizer to optimize feature extraction for sentiment analysis. The max_features parameter was set to 2000, limiting the feature set to the top 2000 words with the highest TF-IDF scores. This approach ensures that only the most influential words are considered in the analysis, reducing noise and computational complexity. The ngram_range was defined as (1,2), enabling the inclusion of both unigrams and bigrams. This configuration allows the model to capture not only single-word features but also contextual relationships within word pairs, which may provide deeper insights into sentiment expressions.

To further refine the feature selection, max_df was set to 0.85, excluding terms that appear in more than 85% of documents, as such terms are likely non-discriminative. Similarly, min_df was set to 3, ensuring that only terms appearing in at least three documents are included. These thresholds help to focus the analysis on words that are both relevant and representative for sentiment classification.

An imbalance issue in data distribution is shown in Figure 2. This distribution indicates an imbalance in the number of instances across sentiment categories, where the "positive" class has the most data, followed by the "neutral" class, with the "negative" class having the least. Such imbalance can affect the model's predictive performance, especially in predicting

minority classes, as the model may be biased towards the majority class.

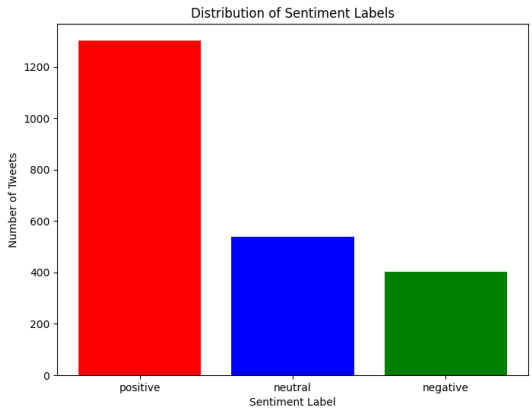

Figure 2. Imbalance Data Before SMOTE

To address this issue, a specialized technique like SMOTE was used. The outcome of SMOTE can be seen in Figure 3:

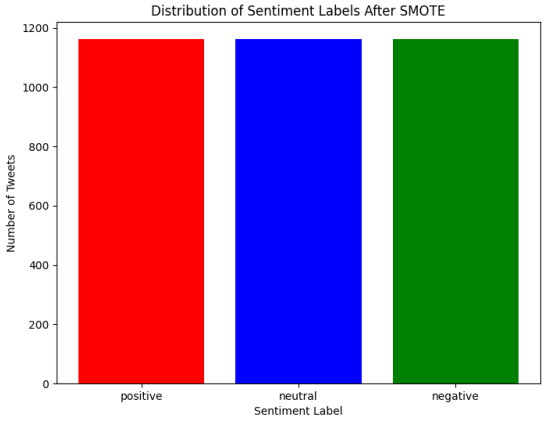

Figure 3. Label Distribution After SMOTE

After applying SMOTE, as shown in Figure 3, the sentiment labels became more balanced across all three classes. This adjustment aids the model in handling data imbalance, improving sensitivity to minority classes, and reducing the likelihood of bias towards the majority class.

D. Naïve Bayes Classifier

The Multinomial Naive Bayes (MNB) algorithm is utilized as the primary classification method in sentiment analysis due to its suitability for this type of task. MNB is a variant of the Naive Bayes algorithm specifically designed to handle categorical data, particularly text, where word frequency within a document serves as a primary feature. This algorithm operates based on the assumption of feature independence, where the probability of a given class, given certain features, is calculated using conditional probability.

One of MNB's key advantages lies in its ability to handle data with a multinomial distribution, making it well-suited for text classification problems. The model training process involves estimating the prior probability for each class as well as the conditional probabilities for each word within the document. These probabilities are then used to predict the class of new data. As a result, MNB is not only efficient in terms of processing time but also effective in producing accurate classifications on preprocessed datasets. The formula for MNB can be seen in Equation 4:

$$P(C|X) = \frac{P(C)\cdot P(X|C)}{P(X)} = \frac{P(C)\cdot \prod_{i=1}^{n} P(x_i|C)}{P(X)} \quad (4)$$

The performance of the Multinomial Naive Bayes (MNB) model was evaluated using a 90:10 train-test data split, where only 10% of the overall data was used as test data.

E. Evaluation

To evaluate the model's performance in sentiment classification, a confusion matrix will first be employed. The confusion matrix is an effective tool for evaluating model predictions by displaying the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Through this visualization, we can identify the model’s strengths and weaknesses in categorizing each class, as well as gain deeper insights into areas requiring improvement. This analysis is critical in determining subsequent steps to enhance the model’s performance in sentiment analysis. The confusion matrix visualization is presented in Figure 4.

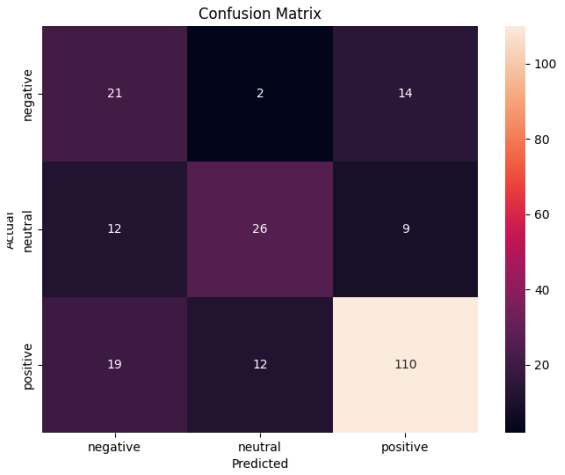

Figure 4. Confusion Matrix

Based on the figure, the model performs best in predicting the positive class, with 110 correct predictions. However, some misclassifications occur: 12 positive instances are classified as neutral and 19 as negative. For the neutral class, 26 instances are correctly classified, but 12 are misclassified as negative and 9 as positive. For the negative class, 21 instances are accurately classified, while 14 are incorrectly labeled as positive and 2 as neutral.

These results indicate that the model is more accurate in predicting the positive class compared to neutral and negative classes, suggesting that it faces challenges in distinguishing ambiguous sentiments or instances with similar characteristics across classes. Consequently, this evaluation highlights opportunities to further improve the model, particularly in increasing accuracy for neutral and negative classes.

In addition to the confusion matrix, the model's performance in sentiment classification can also be assessed using metrics such as Accuracy, Precision, Recall, and F1-score for each class. These performance metrics are summarized in a classification report, as shown in Figure 5.

Classification Report:

	precision	recall	f1-score	support
negative	0.40	0.57	0.47	37
neutral	0.65	0.55	0.60	47
positive	0.83	0.78	0.80	141
accuracy			0.70	225
macro avg	0.63	0.63	0.62	225
weighted avg	0.72	0.70	0.71	225

Figure 5. Classification Report

Precision, recall, and F1-score are essential metrics for evaluating the performance of classification models. Precision measures the proportion of correctly predicted positive instances, indicating the model's reliability in avoiding false positives. Recall, on the other hand, calculates the proportion of actual positive instances correctly identified, reflecting the model's ability to capture all relevant instances. The F1-score, as the harmonic mean of precision and recall, provides a balanced measure, particularly useful when dealing with imbalanced class distributions.

The model was tested using 10% of the dataset, equivalent to 225 samples out of the total data. The evaluation results indicate an overall accuracy of 70%. For the negative class, the model achieved a precision of 40% and a recall of 57%, highlighting challenges in detecting and correctly classifying negative sentiments. The neutral class showed moderate performance, with a precision of 65% and a recall of 55%, suggesting some inconsistencies in identifying neutral sentiments. Conversely, the positive class demonstrated strong performance, achieving a precision of 83% and a recall of 78%, indicating the model's effectiveness in identifying positive sentiment.

While the model's overall accuracy of 70% is encouraging, the variability in performance across sentiment classes underscores areas for improvement. The low precision and recall for the negative class suggest that the model struggles to generalize negative sentiment patterns effectively, possibly due to an imbalance in the dataset or insufficient

representativeness of the training data. Similarly, the neutral class results indicate potential ambiguities in distinguishing neutral expressions from positive or negative sentiments.

To address these challenges, future efforts should focus on enhancing the quality and diversity of the dataset, particularly for underrepresented sentiment classes. To improve the model's ability to capture contextual nuances and achieve more balanced performance across sentiment classes.

CONCLUSION

Based on the results and discussion, the implementation of the Naive Bayes Classifier algorithm for sentiment analysis has been carried out effectively, achieving an accuracy rate of 69.7%, rounded to 70%. The sentiment analysis results on advancements in artificial intelligence technology demonstrate an average precision of 72%, recall of 70%, and F1-score of 71%. While these results indicate the model’s capability in identifying sentiments, its performance is still considered suboptimal, particularly in identifying negative and neutral sentiments.

Several factors may influence these outcomes, including the potential suboptimality of the data preprocessing stage, where improvements to techniques such as cleaning, stemming, and stopword removal could enhance the quality of data used. Additionally, a lack of diversity in the dataset may limit the model’s ability to understand the broader context of expressed sentiments. Much of the data used may not encompass diverse perspectives or expressions related to advancements in artificial intelligence, potentially affecting class representation within the model.

Of data used. Additionally, a lack of diversity in the dataset may limit the model’s ability to understand the broader context of expressed sentiments. Much of the data used may not encompass diverse perspectives or expressions related to advancements in artificial intelligence, potentially affecting class representation within the model.

To improve the model's performance in the future, it is recommended to gather a more varied and representative dataset, explore advanced preprocessing techniques, and consider alternative algorithms such as Transformer-based models like BERT. These methods could provide better contextual understanding and more accurate sentiment classification.

This research highlights the potential of machine learning techniques in analyzing public sentiment, particularly in artificial intelligence technology. The insights gained from this study can serve as a foundation for further development and optimization, ultimately aiding researchers and developers in aligning technological advancements with societal expectations and needs.

REFERENCES

Aggarwal, C. C. (2015). *Data Mining*. Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8

Aydogan, E., & Akcayol, M. A. (2016). *A Comprehensive Survey for Sentiment Analysis Tasks Using Machine Learning Techniques*. IEEE.

Bird, S., Klein, E., & Edward Loper. (2009). *Natural Language Processing with Python*.

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. In *Journal of Artificial Intelligence Research* (Vol. 16).

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, *15*(1), 131–145.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). LNCS 3644 - Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In *LNCS* (Vol. 3644).

Hasibuan, M. S., & Serdano, A. (2022). Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naive Bayes. *JRST (Jurnal Riset Sains Dan Teknologi)*, *6*(2), 199. https://doi.org/10.30595/jrst.v6i2.15145

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning:Trends, perspectives, and prospects. *Science*, *349*(6245), 253–255. https://doi.org/10.1126/science.aac4520

Joshi, M., Das, D., Gimpel, K., & A. Smith, N. (2010). N10-1038. *Human Language Technologies*, 293–296.

Liu, B. (2020). *Sentiment Analysis Essentials Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*.

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

Powers, D. M. W., & Ailab. (n.d.). *EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION*.

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (n.d.). *Tackling the Poor Assumptions of Naive Bayes Text Classifiers*.

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, *60*(5), 503–520. https://doi.org/10.1108/00220410410560582

Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, *36*. https://doi.org/10.1016/j.jocs.2019.05.009

Ullah, A., Zhang, Q., & Ahmed, M. (2021). The influence of intellectual property rights protection on contribution efforts of participants in online crowdsourcing contests. *Computers in Human Behavior*, *123*. https://doi.org/10.1016/j.chb.2021.106869

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5