Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Analysis of SVM and BERT for Sentiment and Sarcasm Detection in the Boycott of Israeli Products on Platform X

Siti Sarah Sabrina[1)*], Diqy Fakhrun Shiddieq[2)], Fikri Fahru Roji[3)]
[1,2,3)] Digital Business Departement, Garut University, Garut, Indonesia
[1)]24025121063@fekon.uniga.ac.id, [2)]diqy@uniga.ac.id, [3)]fikri@uniga.ac.id

Submitted :April 24, 2025 | **Accepted** : May 8, 2025 | **Published** : May 13, 2025

Abstract: The Israel-Palestine conflict has triggered a global consumer movement, including a widespread boycott of Israeli-affiliated products in Indonesia. As this campaign gains momentum on digital platforms like X (formerly Twitter), understanding public sentiment becomes crucial—not only for gauging public opinion but also for anticipating potential socio-economic impacts. This study evaluates the effectiveness of two sentiment analysis models—Support Vector Machine (SVM) and Bidirectional Encoder Representations from Transformers (BERT)—in classifying sentiment and detecting sarcasm related to the boycott campaign. A total of 5,637 Indonesian-language tweets were manually labeled into positive, neutral, and negative categories, with sarcasm detection performed using a fine-tuned IndoBERT, model which classified tweets into two categories: sarcastic and non-sarcastic. The models were assessed using accuracy, precision, recall, F1-score, and computational efficiency. Results show that BERT outperforms SVM in both sentiment classification (accuracy: 69.26% vs. 64.58) and sarcasm detection (accuracy: 92.20% vs. 86.15%). However, the improved performance of BERT comes at the cost of considerably higher computational time than SVM in both sentiment and sarcasm tasks. These findings highlight a trade-off between contextual comprehension and real-time efficiency. Future research may explore ensemble methods or threshold-tuning to optimize this balance. The practical implications of this research lie in its application for real-time public discourse monitoring and data-driven policy development. By improving the detection of nuanced expressions such as sarcasm, this study contributes to more accurate sentiment interpretation in polarized digital environments.

Keywords: BERT, Classification, Sarcasm detection, Sentiment analysis, SVM.

INTRODUCTION

The Israel-Palestine conflict significantly affects global political, economic, and social dynamics. In Indonesia, the conflict has triggered a solidarity movement, notably through boycotts of products associated with Israel. This movement has gained substantial traction on platform X, where Indonesians actively express dissent toward Israeli policies (Azra et al., 2023; Wulandari & Saefudin, 2024).

The boycott has also had a considerable economic impact, boosting support for local products and expanding the market for domestic businesses, while also influencing social stability (Wahdiyana et al., 2024). In contrast, multinational companies with ties to Israel have experienced declining sales and significant reputational damage, posing a threat to their long-term sustainability (Witro, 2024). These developments highlight the powerful role of consumer movements in reshaping domestic market dynamics. They also underscore the importance of sentiment analysis in measuring public opinion, understanding polarized views, and assessing the socio-economic effects of the campaign (Aniket et al., 2024; Novianti et al., 2024; Patsiaouras, 2022). For policymakers and businesses, sentiment analysis offers valuable insights, enabling more effective and timely responses to challenges that may disrupt social stability (Aniket et al., 2024; Tan et al., 2023).

Sentiment analysis, a subfield of text mining, focuses on the systematic study of public emotions, opinions, and attitudes expressed in textual form (Ma'aly et al., 2024). This study utilizes platform X to analyze sentiments associated with the boycott of products linked to Israel, as X is a widely-used text-based social media platform in Indonesia, boasting 24.85 million users (Paath et al., 2024). Given its popularity, X provides a relevant medium

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

for capturing public opinion regarding the boycott (Seki, 2024). Tweets and other textual data extracted from platform X are subsequently analyzed using sentiment analysis techniques (Ma'aly et al., 2024). This study aims to classify sentiment into three distinct types—positive, negative, and neutral—representing support, disapproval, or neutrality concerning the boycott campaign (J, 2024).

Despite its potential, sentiment analysis faces significant challenges, particularly when dealing with linguistic complexities, ambiguity, sarcasm, and slang (Thomas et al., 2021). Sarcasm is a linguistic expression where the intended sentiment contrasts with the literal meaning, often used to convey ridicule or criticism implicitly (Sharma et al., 2022). The detection of sarcasm is particularly difficult because it often masks true sentiment and requires models to understand nuanced context beyond surface-level words. Support Vector Machine (SVM), As a classical machine learning technique, SVM has shown strong performance in sentiment analysis, largely attributed to its capability to manage both linear and nonlinear patterns within data (Rewatkar et al., 2024). Numerous studies have demonstrated that SVM achieves high accuracy and precision in sentiment classification across diverse domains (Fadlil et al., 2024; Khan et al., 2024; Mantik & Singgalen, 2024). However, while SVM is effective, it struggles to capture the contextual nuances present in complex sentence structures, which are essential in sentiment analysis. This is where newer models like Bidirectional Encoder Representations from Transformers (BERT) come in. (González-Carvajal & Garrido-Merchán, 2020). In numerous NLP applications, BERT has demonstrated outstanding performance, particularly in contextual comprehension, outperforming conventional methods such as SVM and LSTM in sentiment classification (Sanwal & Mazhar, 2023). Furthermore, recent studies show that fine-tuned BERT models can accurately detect sarcasm in Indonesian text, highlighting its potential in understanding sentiment beyond simple classifications (Fitrianto & Editya, 2024; Imawan et al., 2025).

This study seeks to answer which model, SVM or BERT, is more effective in capturing sentiment and sarcastic expressions related to the boycott campaign on platform X. While prior studies have applied classical machine learning or deep learning models for sentiment analysis, few have directly compared these approaches within the specific socio-political context of the Israeli product boycott in Indonesia. Existing research often overlooks the challenge of detecting sarcasm—a linguistic feature prevalent in social media discourse. This study addresses this gap by evaluating the performance of SVM and BERT in sentiment classification and incorporating sarcasm detection using BERT to assess each model's contextual understanding. This dual focus on performance comparison and sarcasm detection in a real-world campaign context sets this research apart. The study assesses both algorithms using evaluation indicators—accuracy, precision, recall, computational efficiency, and contextual understanding—applied to sentiment analysis of boycott-related content on platform X (Zhang, 2024).

LITERATURE REVIEW

Analyzing sentiment is now a crucial method for capturing public perspectives across social media environments.. It has been widely applied across domains such as finance, healthcare, and political discourse. One of the main challenges lies in the informal language and slang commonly found on social media, which can complicate the process of sentiment classification and affect model performance. Support Vector Machine (SVM) continues to be a widely used method for sentiment classification, praised for its robustness with high-dimensional data and efficiency on smaller datasets. However, results vary depending on the context of the data. For instance, SVM achieved 74.44% accuracy in a study on Palestine-Israel conflict sentiment (Andriawan & Ernawati, 2024), yet outperformed Naïve Bayes with 86.14% accuracy in product review classification on Shopee (Jannah & Kusnawi, 2024). Similarly, SVM reached 85% testing accuracy in analyzing YouTube comments, surpassing Random Forest (Syafia et al., 2023).

The emergence of transformer architectures like BERT has significantly advanced sentiment analysis, particularly in interpreting intricate language patterns, including sarcasm and subtle expressions. Studies have demonstrated BERT's superiority in sarcasm detection. For example, (Pandey & Singh, 2023) showed the effectiveness of a BERT-LSTM model for sarcasm detection in code-mixed texts, while (Wang, 2024) confirmed BERT's contextual strength through its integration with fuzzy logic.

To evaluate sentiment classification models effectively, it is important to use multiple evaluation metrics. Although accuracy offers a broad picture of model effectiveness, it can be unreliable when dealing with imbalanced datasets. Therefore, metrics like precision, recall, and F1-score are used to gain deeper insight into how well the model distinguishes between sentiment categories.. These metrics have been consistently used in recent studies. For instance, (Prasetyo et al., 2024) applied them to compare MultinomialNB, SVM, and BERT for analyzing sentiment on Garuda Indonesia's Twitter account, where BERT outperformed the others. Similarly (ElMassry et al., 2024) used the same metrics to evaluate SVM, LSTM, and BERT across both balanced and imbalanced datasets, with fine-tuned BERT showing the best overall performance.

Beyond conventional metrics, this study also utilizes the Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves to enhance evaluation depth, particularly in scenarios involving class imbalance. The PR curve is effective in measuring model performance on underrepresented classes, whereas the ROC curve helps

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

analyze the balance between correctly identified positives and incorrectly flagged negatives.The importance of these curves is supported by (Richardson et al., 2024), who demonstrated their effectiveness through simulations and real-world datasets, affirming their value in imbalanced sentiment classification tasks.

METHOD

The methodology employed in this study encompasses several stages designed to ensure accurate and reliable sentiment and sarcasm classification. Each step—ranging from data collection to evaluation—is carefully structured to address the research objectives and challenges discussed earlier. The overall research workflow is summarized in the following diagram:

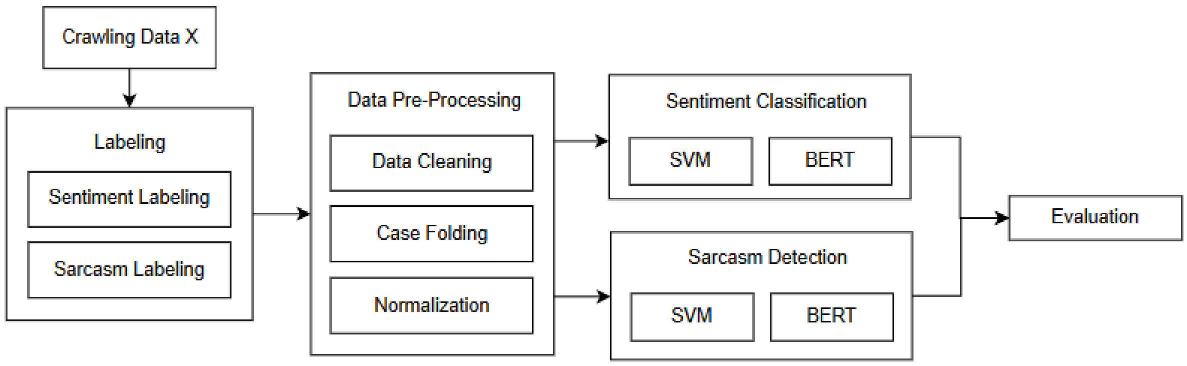


Fig. 1 Diagram Method

Data Crawling

Data was collected from Twitter search results based on specific keywords and date ranges, focusing on tweets containing "Boikot produk Israel, "Boikot MCD," or "Boikot Starbucks" from March 1, 2023, to September 15, 2024. This process resulted in 5,637 tweets prior to pre-processing. The majority of the dataset was in Indonesian, reflecting the prevalent use of the language in the collected tweets.

Data Labeling

The manual labeling process was crucial in this study due to the contextual nuances within the boycott campaign against Israeli products. Many tweets explicitly support the movement while expressing negative sentiments toward Israel or its affiliates, which can mislead automated sentiment analysis. To ensure contextual accuracy, manual labeling was conducted, resulting in 908 negative, 1,453 neutral, and 3,087 positive tweets. In this context, sentiment labeling was carried out based on the pragmatic meaning of each tweet rather than its literal wording. Tweets expressing sharp criticism of Israeli-affiliated products were labeled as positive, as they reflected support for the boycott movement. Rather than focusing on tone, sentiment was defined by the intended stance of the message. To accommodate this, sentiment classification was applied at the sentence level, treating each tweet as a standalone textual unit. This approach is well-suited to the concise nature of tweets and allows for the nuanced interpretation of individual expressions, as emotional and confrontational language is often used to signal support in social media discourse (Azzahro, 2024).

Table 1. Sample Tweets and Their Sentiment Labels

Tweet	Sentiment Label
Ayo boikot produk Israel dr negara ini krn Israel adlh negara penajajah Palestina	2 (Positive)
Tanggapan dia soal gerakan boikot produk pro Israel spt apa?	1 (Neutral)
Paling TOLOL kumpulan boikot ini... Gak ada gunanya kalian boikot produk Israel	0 (Negative)

To further capture indirect expressions, sarcasm detection was applied using a pre-trained BERT-based model. Upon processing, the model generates raw scores (logits) for each class, which are then transformed into probability values through the softmax activation function. Based on these probabilities, a threshold of 0.5 is applied to determine the classification outcome. If the predicted probability for the sarcastic class is equal to or greater than 0.5, the tweet is classified as sarcastic. Conversely, if the probability is below 0.5, the tweet is labeled

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

as non-sarcastic. The choice of a 0.5 threshold is based on standard practice in binary classification, as it provides a balanced trade-off between sensitivity and specificity.

Table 2. Sample Tweets with Their Sarcasm and Sentiment Label

Tweet	Sarcasm Label	Sentiment Label
Apalah daya saya yang cuma remahan rengginang ini yang cuma bisa boikot starbucks sementara bohir seneng produk israel	Sarcasm	Positive
Protes penjajahan israel sah-sah saja tapi boikot starbucks? Brilliant banget logikanya	Sarcasm	Negative
Bagus sih biokot-boikot gitu tapi apakah pendanaan israel cuman dari starbuck dan produk famp	Sarcasm	Neutral
Boikot starbucks dan mcdoald's dan juga amway karena mereka dukung israel.	Non-Sarcasm	Positive
Aku benci israel tapi kalau sampai boikot produk tidak bisa kayaknya	Non-Sarcasm	Negative
Silakan saja kalau kamu memilih bersikap diam soal isu palestina, itu hak kamu juga sebagai manusia	Non-Sarcasm	Neutral

This process yielded 1,246 sarcastic and 4,202 non-sarcastic tweets. Incorporating sarcasm detection enhances sentiment interpretation by distinguishing between literal and ironic expressions, contributing to more reliable analysis outcomes.

Pre-processing Data

Data preparation is a crucial step to ensure the quality and reliability of the analysis by removing noise and inconsistencies. This enhances the validity of the results and strengthens the research's understanding, where raw data is cleaned and processed into a usable format for the models (Singgalen, 2024). The pre-processing stages include data cleaning, which involves removing punctuation, unnecessary characters, and correcting spelling errors and duplicates; case folding to convert all text to lowercase; normalization to change words to their standard forms; tokenization to split text into individual words (Selian et al., 2024) (Dwiki et al., 2021). In this study, a total of 5,637 tweets related to the boycott of Israeli products were initially collected from the X platform. After undergoing pre-processing to remove irrelevant or noisy content, the final dataset consisted of 5,447 usable tweets.

In this study, the collected data was processed to improve quality and ensure consistency in the analysis. The pre-processing steps included data cleaning by removing URLs, hashtags, mentions, numbers, and non-alphabetic characters. Case folding was then applied to convert all text to lowercase. For the SVM model, normalization was implemented to convert non-standard words into their standard forms using a predefined dictionary, such as changing 'abis' to 'habis,' 'ad' to 'ada,' and 'adlh' to 'adalah'.

Modeling

The method selection was based on previous research highlighting the strengths of both classical and transformer-based models in sentiment analysis tasks (Fitrianto & Editya, 2024; González-Carvajal & Garrido-Merchán, 2020; Rewatkar et al., 2024). This study implements two distinct modeling processes on the same dataset, each addressing a different classification objective. The first modeling task focuses on sentiment classification, where tweets are labeled as positive, neutral, or negative based on the expressed attitude toward the boycott campaign. The second modeling task is dedicated to sarcasm detection, which classifies tweets as either sarcastic or non-sarcastic, aiming to capture implicit and ironic expressions that may interfere with sentiment interpretation.

During the modeling process, both classical machine learning and deep learning approaches are utilized. Support Vector Machine (SVM), recognized for its capability to model linear and non-linear patterns, is used to classify sentiments and sarcasm detection in tweets related to the boycott of Israeli products (Rewatkar et al., 2024). This algorithm is particularly well-suited for high-dimensional text classification tasks. In this study, textual

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

input is converted into numerical form using Term Frequency-Inverse Document Frequency (TF-IDF), limited to 5,000 features. The SVM classifier is trained with a linear kernel, a fixed random state to maintain reproducibility, and probability estimation enabled.

In contrast, the Bidirectional Encoder Representations from Transformers (BERT) model is utilized to leverage its deep, bidirectional, and contextual understanding, which is highly effective for natural language processing tasks (González-Carvajal & Garrido-Merchán, 2020). This study adopts the IndoBERT-base model as the pre-trained transformer, fine-tuned for sentiment classification across three labels. For BERT, tokenization is performed using the official BERT Tokenizer, with padding and truncation applied to maintain uniform sequence lengths. The model is fine-tuned using the AdamW optimizer, set at a learning rate of 2e-5, with a batch size of 16, and trained for three epochs.

Evaluation

After the models were developed, a thorough evaluation was conducted using a test dataset to measure their performance. To evaluate the sentiment classification models, six key metrics are used: Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curve, and Precision-Recall (PR) curve. These performance indicators provide a holistic view of how well each model performs in terms of classification accuracy and reliability.

Accuracy

Accuracy refers to the ratio of correctly predicted instances to the total number of observations in the dataset.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

TP (True Positive) = Cases where the model correctly identifies a positive instance

TN (True Negative) = Instances accurately predicted as negative

FP (False Positive) = Negative instances incorrectly classified as positive

FN (False Negative) = Positive instances that the model fails to identify correctly

Precision

Precision quantifies the proportion of correctly predicted positive observations to the total predicted positives

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

High precision indicates a lower false positive rate, making it a crucial metric for cases where false positives must be minimized (Couvy-Duchesne et al., 2020).

Recall (Sensitivity)

Recall refers to the model's effectiveness in retrieving all actual positive samples within the test data.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

A high recall value indicates that the model captures most of the positive instances, reducing false negatives (Bold et al., 2022).

F1-Score

F1-score captures the balance between precision and recall, making it a comprehensive indicator of classification performance

$$F1\ Score = 2x\ \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

This metric is particularly useful when the dataset is imbalanced (Riyanto et al., 2023).

ROC Curve

By plotting the True Positive Rate against the False Positive Rate across different thresholds, the ROC curve offers insight into a model's discriminative performance.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i2.14723 p-ISSN : 2541-044X

$$TPR = \frac{TP}{TP+FN}\ ,\ FPR = \frac{FP}{FP+TN} \qquad (5)$$

A higher ROC curve indicates better model performance in distinguishing between classes (Richardson et al., 2024).

Precision-Recall (PR) Curve

The PR curve visualizes how precision and recall interact across different classification thresholds, proving especially beneficial when dealing with imbalanced data.

$$PRCurve = (Recall, Precision) \qquad (6)$$

A model with a higher PR curve performs better in maintaining the trade-off between precision and recall is a key factor in the effectiveness of sentiment classification models (Riyanto et al., 2023).

Beyond these conventional metrics, the evaluation also examined processing speed by analyzing the time required for each model to classify tweets, offering insights into computational efficiency, particularly for real-time applications. Additionally, the assessment considered the models' ability to interpret complex sentiment expressions, especially sarcasm, which often carries implicit or nuanced meanings. This aspect was evaluated qualitatively by reviewing the classification results on tweets containing ambiguous language. Sarcasm detection was performed using a BERT-based model, where tweets were labeled as either non-sarcastic or sarcastic based on contextual understanding. The results of these evaluations were then analyzed to compare the performance of classical machine learning algorithms like SVM with the BERT model. This comparison aimed to highlight the advantages and limitations of each approach, particularly in their ability to balance classification accuracy, computational efficiency, and contextual comprehension in sentiment analysis related to the Israeli product boycott campaign on platform X.

RESULT

Performance Metrics

The evaluation metrics of SVM and BERT in sentiment classification are shown in Table 3.

Table 3. Performance Comparison of SVM and BERT in Sentiment Classification

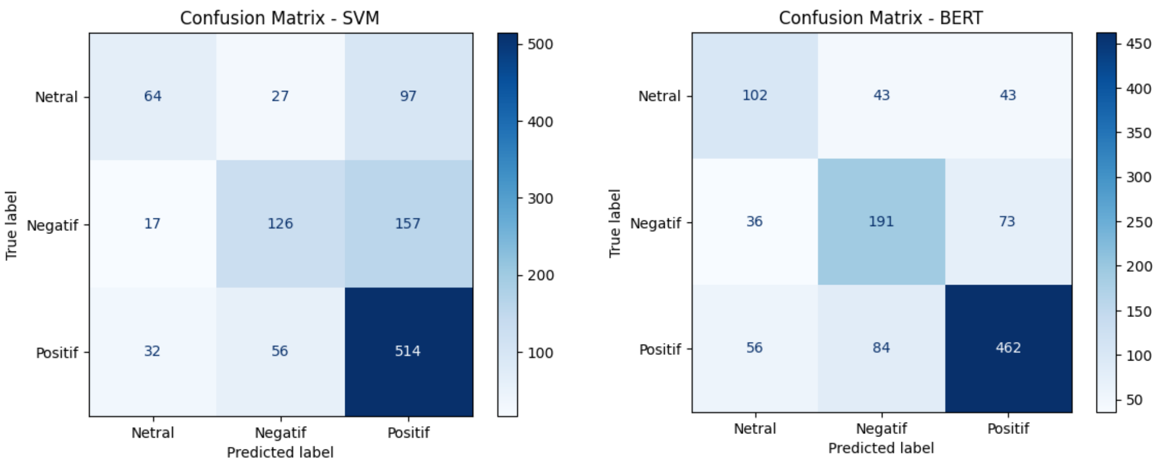
Model	Accuracy	Precision	Recall	F1-Score	Computational Time (s)
SVM	64.58%	63,32%	64,58%	62,40%	18.81
BERT	69.26%	69,74%	69,26%	69,47%	194.76

Table 3 shows the performance results of the SVM and BERT models in the sentiment classification task. The table includes values for accuracy, precision, recall, F1-score, and computational time for each model. BERT achieved higher values across all evaluation metrics compared to SVM. In terms of computational time, BERT required more processing time than SVM.

In this study, accuracy and F1-score are selected as the primary evaluation metrics for sentiment classification. Accuracy provides a general overview of the model's overall correctness, while F1-score offers a balanced measure that takes into account both precision and recall. These two metrics together give a comprehensive view of model performance, particularly in capturing the nuances of sentiment across multiple classes.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X



(a) (b)

Fig 2. Confusion Metrics: (a) SVM), (b) BERT

The figure above presents the confusion matrices of two compared models, Support Vector Machine (SVM) and BERT, in the sentiment analysis task related to the boycott campaign against Israeli products on Twitter. From the confusion matrices, it is evident that both models performed best in classifying positive sentiments. The SVM model correctly classified 514 out of 602 positive tweets (85.4%), while BERT correctly classified 462 tweets in the same class (76.7%). In the neutral class, SVM correctly identified 64 out of 188 neutral tweets (34.0%), while BERT achieved higher accuracy with 102 correct predictions (54.3%). Furthermore, SVM misclassified 97 neutral tweets as positive, compared to 43 such misclassifications by BERT. For negative sentiment classification, BERT outperformed SVM by correctly identifying 191 out of 300 negative tweets (63.7%), whereas SVM only classified 126 correctly (42.0%). SVM also misclassified 157 negative tweets as positive, while BERT reduced this number to 73.

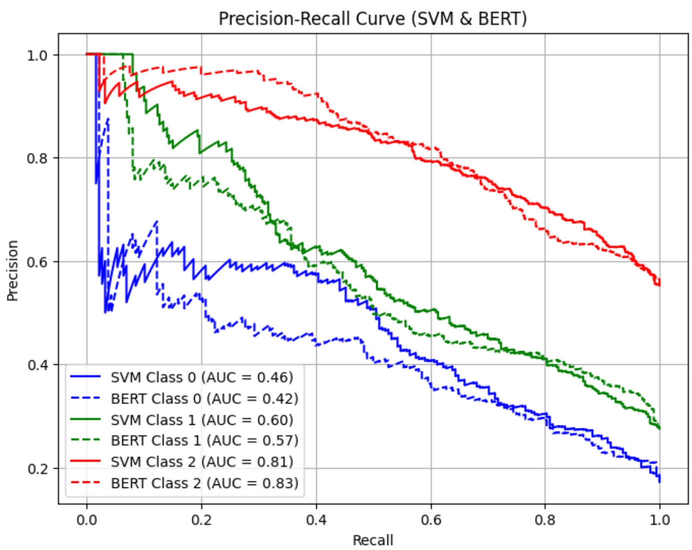


Fig 3. Precision - Recall Curve

Figure 3 presents the Precision-Recall (PR) curve for the SVM and BERT models in the sentiment analysis task regarding the boycott campaign of Israeli-affiliated products on Twitter. The PR curve evaluates the trade-off between precision and recall, which is crucial for handling imbalanced datasets. The primary metric for assessing classification performance is the Area Under the Curve (AUC-PR).

For Class 0 (Negative), the SVM model achieved an AUC of 0.46, while the BERT model had a slightly lower AUC of 0.42. This indicates that both models struggle to maintain high precision as recall increases. The fluctuating curve at the beginning suggests a level of uncertainty in predictions for this class.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14723
e-ISSN : 2541-2019
p-ISSN : 2541-044X

For Class 1 (Neutral), the SVM model attained an AUC of 0.60, outperforming BERT, which obtained an AUC of 0.57. This suggests that SVM is slightly better at handling neutral-class predictions, although the overall performance for this class remains moderate.

For Class 2 (Positive), both models demonstrated strong performance, with SVM achieving an AUC of 0.81 and BERT slightly outperforming with an AUC of 0.83. The relatively stable PR curve and high AUC values indicate that both models effectively classify tweets supporting the boycott campaign.

Overall, SVM shows a slight advantage in classifying Class 1 (Neutral), while BERT exhibits comparable performance in Class 2 (Positive). However, for Class 0 (Negative), both models struggle to achieve an optimal precision-recall balance. These findings highlight the need for further improvements, particularly in distinguishing negative and neutral sentiments.

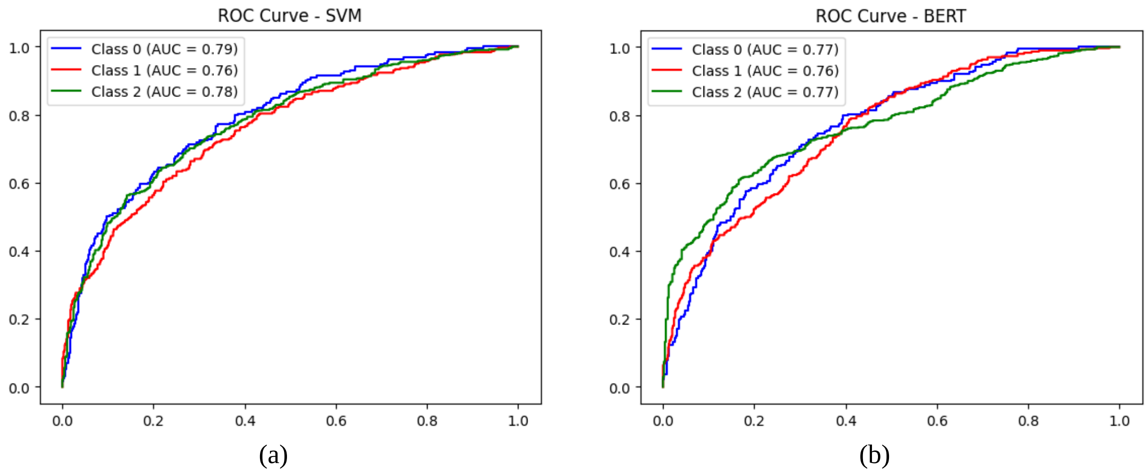


Fig. 4 ROC Curve: (a) SVM, (b) BERT

Figure 4 presents the Receiver Operating Characteristic (ROC) curves for the SVM and BERT models in the sentiment analysis task regarding the boycott campaign of Israeli-affiliated products on TwitterThe ROC curve assesses how well a model can separate different sentiment categories, with the AUC (Area Under the Curve) being a crucial indicator of its classification effectiveness.

For the SVM model (left figure), the AUC values are 0.79 for Class 0 (Negative), 0.76 for Class 1 (Neutral), and 0.78 for Class 2 (Positive). The ROC curves for all three classes show relatively stable patterns, indicating that SVM maintains a consistent ability to differentiate between sentiment categories.

In contrast, for the BERT model (right figure), the AUC values are 0.77 for Class 0 (Negative), 0.76 for Class 1 (Neutral), and 0.77 for Class 2 (Positive). The performance of BERT appears comparable to SVM, as the AUC values for all classes are within a close range. However, BERT's ROC curves exhibit more fluctuations in the early stages, particularly for Class 2, which may indicate some instability in classification confidence at lower thresholds.

Overall, both models demonstrate similar classification capabilities in distinguishing sentiment classes. While SVM slightly outperforms BERT in Class 0 (Negative) and Class 2 (Positive), the performance gap is minimal, suggesting that BERT remains a competitive alternative for sentiment classification in this context.

Performance Analysis in Sarcasm Detection

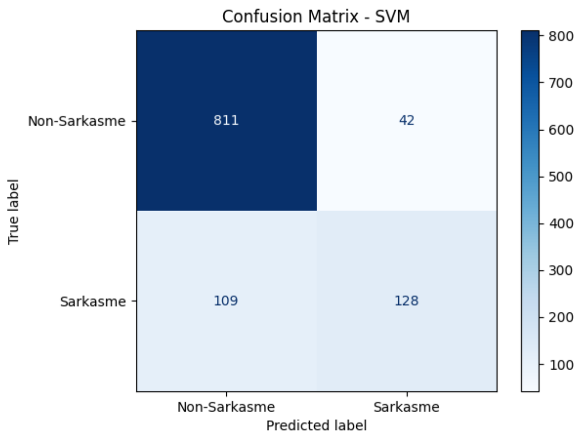
In this study, we compare the performance of Support Vector Machine (SVM) and BERT in classifying sarcastic sentiment. The evaluation is conducted using Accuracy, Precision, Recall, and F1-score metrics, as well as the training time for each model. The evaluation results are presented in Table 4, while the performance of each model in distinguishing sarcasm and non-sarcasm is visualized through the confusion matrices in Figures 5.

Table 4. Performance Comparison of SVM and BERT in Sarcasm Classification

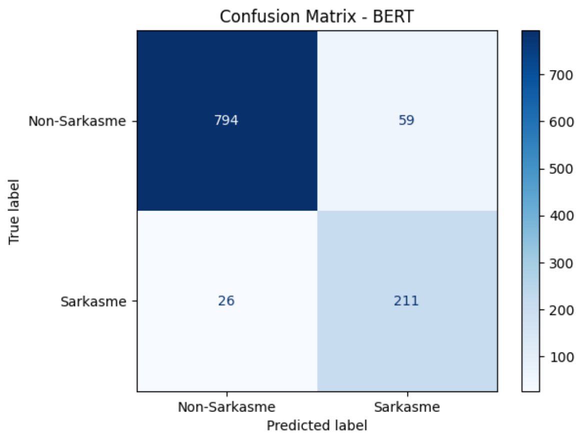
Model	Accuracy	Precision	Recall	F1-Score	Computational Time (s)
SVM	86,15%	85,36%	86,15%	85,27%	10.99
BERT	92,20%	92,77%	92,20%	92,38%	191.92

*name of corresponding author

The evaluation results are presented in Table 4, while the performance of each model in distinguishing sarcasm and non-sarcasm is visualized through the confusion matrices in Figures 5.



(a)



(b)

Fig 5. Confusion Matrix for Sarcasm Classification: (a) SVM), (b) BERT

Based on Figure 5 the SVM model correctly classified 811 out of 853 non-sarcastic tweets (95.1%), while BERT correctly classified slightly fewer non-sarcastic tweets with 794 correct predictions (93.1%). In contrast, BERT performed significantly better in classifying sarcastic tweets, with 211 correct predictions out of 237 (89.0%), compared to SVM, which only classified 128 sarcastic tweets correctly (54.0%). Additionally, SVM misclassified 109 sarcastic tweets as non-sarcastic, while BERT made only 26 such misclassifications.

DISCUSSIONS

This study explores and compares the performance of Support Vector Machine (SVM) and BERT models in sentiment classification and sarcasm detection of tweets related to the boycott of products affiliated with Israel on Platform X.

One of the key insights emerging from this study is the role of pragmatic interpretation in sentiment labeling. Tweets were labeled not solely based on their literal wording, but on the intended stance they conveyed—particularly in the context of politically charged discourse such as the boycott of Israeli-affiliated products. For instance, tweets expressing strong criticism toward these products were classified as positive, as they implicitly supported the boycott movement. This approach aligns with the findings of (Azzahro, 2024), who observed that support for the BDS (Boycott, Divestment, and Sanctions) movement is frequently communicated through emotionally charged, confrontational language on social media platforms. However, this method also introduces new challenges, especially when intersecting with sarcasm detection. Sarcastic expressions often subvert literal meanings, making it difficult to distinguish between sincere praise and ironic critique. For example, a tweet that appears to praise an Israeli product may in fact be a sarcastic condemnation. This underscores the importance of distinguishing between sentiment classification—which focuses on the speaker's attitude—and sarcasm detection—which deals with linguistic inversion. As a result, models tasked with analyzing sentiment in socially or politically sensitive contexts must incorporate deeper contextual awareness to avoid misclassification, particularly in the presence of irony or implicit expressions.

The findings of this study are consistent with previous research by (Bhargava et al., 2025), which found that the sarcasm in social media posts tends to reduce the accuracy of conventional sentiment analysis models. Similarly, (Slim et al., 2024) emphasized the necessity of incorporating sarcasm detection to enhance the performance of sentiment classification, particularly on platforms like. However, unlike many previous studies that rely solely on lexical features, this study adopts a contextual and pragmatic labeling approach, which prioritizes communicative intent over literal linguistic expression.

This conceptual shift highlights the importance of developing adaptive sentiment labeling methods that are better aligned with the dynamic and often ambiguous nature of social media language, especially within controversial or emotionally charged topics. Furthermore, while sarcasm has often been treated as "noise" to be filtered out in traditional models, this study treats sarcasm as a critical signal that reveals deeper, often hidden layers of sentiment.

Experimental results indicate that BERT generally outperforms SVM, particularly in terms of accuracy, precision, recall, and F1-score. These results reinforce the findings of (M. Sharma et al., 2024) and (Prasetyo et

*name of corresponding author

al., 2024), who demonstrated that BERT's ability to capture semantic relationships and contextual nuance makes it especially effective for analyzing complex and emotionally expressive language on social media.

Nevertheless, in the analysis of the Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves, SVM shows superior performance. These metrics are particularly useful in imbalanced data scenarios, as was the case in this study, where negatively labeled tweets were significantly outnumbered by positive ones. SVM's stability in such conditions suggests that it is better equipped to handle class imbalance and maintain discrimination power for minority classes. Conversely, despite its overall strength, BERT tends to make more conservative predictions in imbalanced settings—likely due to its sensitivity to the distribution of training data.

The class imbalance within the dataset is a key factor contributing to the performance gap between the two models. Moreover, the linguistic complexity of the tweets—including slang, hyperbole, and sarcasm—presents unique challenges, particularly for models like SVM that rely heavily on surface-level features. Applying data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique), as suggested by (Singgalen, 2024), may enhance BERT's performance in future research involving similarly skewed datasets.

Another critical distinction lies in computational efficiency. BERT requires significantly more processing time and computational resources due to its complex architecture and high parameter count. In contrast, SVM is lightweight and fast, making it a practical choice for low-resource environments or real-time processing needs, although it is less effective than BERT in capturing deep contextual meaning.

CONCLUSION

This study's outcomes reveal that BERT generally outperforms SVM in sentiment analysis of the boycott campaign against Israeli products on the X platform. BERT achieves higher accuracy (69.26% vs. 64.58%), precision (69.74% vs. 63.32%), recall (69.26% vs. 64.58%), and F1-score (69.47% vs. 62.40%), demonstrating its superior ability to understand complex linguistic contexts. However, based on the Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves, SVM performs better in handling class imbalance and maintaining the trade-off between precision and recall, particularly in distinguishing negative and neutral sentiments.

In sarcasm classification, BERT also outperforms SVM, achieving an accuracy of 92.20% compared to SVM's 86.15%. Additionally, BERT's F1-score is higher at 92.38%, whereas SVM achieves 85.27%, further reinforcing BERT's strength in detecting implicit linguistic expressions.

Despite its superior performance, BERT requires significantly longer computational time (194.76 seconds for sentiment classification and 191.92 seconds for sarcasm detection), making it less efficient for real-time applications compared to SVM, which only takes 18.81 seconds and 10.99 seconds, respectively. Therefore, model selection should be aligned with the objectives of sentiment analysis. If the primary goal is overall accuracy and better contextual understanding, BERT is the preferred choice. However, if computational efficiency and stability in imbalanced data conditions are prioritized, SVM can be a more practical alternative.

To harness the strengths of both approaches, future research may consider improving BERT's effectiveness through threshold calibration or the use of ensemble learning techniques. This could further improve sentiment classification accuracy, particularly in distinguishing sarcasm and negative sentiments more effectively.

REFERENCES

Andriawan, M. G., & Ernawati, T. (2024). Penggunaan algoritma Naïve Bayes dan Support Vector Machine untuk analisis sentimen konflik Palestina dan Israel pada platform X. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3). https://doi.org/10.23960/jitet.v12i3.4943

Aniket, G., Pratik, M. M., Lukesh, N. O., & Atharva, R. Y. (2024). Sentiments of public opinion. *International Research Journal of Modernization in Engineering Technology and Science*.

Azra, F. A. Z., Yudhianti, O., & Chrisworo, O. P. (2023). Perbandingan aksi, reaksi, dan hubungan internasional berbagai negara terhadap konflik Muslim Rohingya–Myanmar dan Muslim Palestina–Israel. *Triwikrama: Jurnal Multidisiplin Ilmu Sosial*.

Azzahro, S. F. (2024). Sentiment analysis on boycott movement: NVivo approach. *Jurnal Ilmiah Ekonomi Kita*, *13*(2), 393–414. https://doi.org/https://doi.org/10.46367/iqtishaduna.v13i2.2185

Bhargava, N., Radaideh, M. I., Kwon, O. H., Verma, A., & Radaideh, M. I. (2025). On the Impact of Language Nuances on Sentiment Analysis with Large Language Models: Paraphrasing, Sarcasm, and Emojis. http://arxiv.org/abs/2504.05603

Bold, R., Al-Khateeb, H., & Ersotelos, N. (2022). Reducing False Negatives in Ransomware Detection: A Critical Evaluation of Machine Learning Algorithms. *Applied Sciences*, *12*(24), 12941. https://doi.org/10.3390/app122412941

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Couvy-Duchesne, B., Zhang, F., Kemper, K. E., Sidorenko, J., Wray, N. R., Visscher, P. M., Colliot, O., & Yang, J. (2020). Linear Mixed Models Minimise False Positive Rate and Enhance Precision of Mass Univariate Vertex-Wise Analyses of Grey-Matter. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 404–407. https://doi.org/10.1109/ISBI45749.2020.9098719

Dwiki, A., Putra, A., & Juanita, S. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(2), 636–646. http://jurnal.mdp.ac.id

ElMassry, A. M., Alshamsi, A., Abdulhameed, A. F., Zaki, N., & Belkacem, A. N. (2024). Machine Learning Approaches for Sentiment Analysis on Balanced and Unbalanced Datasets. *2024 IEEE 14th International Conference on Control System, Computing and Engineering (ICCSCE)*, 18–23. https://doi.org/10.1109/ICCSCE61582.2024.10695972

Fadlil, A., Riadi, I., & Andrianto, F. (2024). Improving Sentiment Analysis in Digital Marketplaces through SVM Kernel Fine-Tuning. *International Journal of Computing and Digital Systems*, *16*(1), 159–171. https://doi.org/10.12785/ijcds/160113

Fitrianto, R. A., & Editya, A. S. (2024). Klasifikasi Tweet Sarkasme Pada Platform X Menggunakan Bidirectional Encoder Representations from Transformers. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(3), 366–371. https://doi.org/10.47233/jteksis.v6i3.1344

González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. *Journal of Computational and Cognitive Engineering*, *2*(4), 352–356. https://doi.org/10.47852/bonviewJCCE3202838

Imawan, F., Shiddieq, D. F., & Roji, F. F. (2025). Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert. *CESS (Journal of Computer Engineering, System and Science)*, *10*(1), 136. https://doi.org/10.24114/cess.v10i1.66455

J, B. (2024). A study and development of application on sentiment analysis. *International Scientific Journal of Engineering and Management*, *03*(03), 1–7. https://doi.org/10.55041/ISJEM01354

Jannah, N. Z. B., & Kusnawi, K. (2024). Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces. *Sinkron*, *8*(2), 727–733. https://doi.org/10.33395/sinkron.v8i2.13559

Khan, T. A., Sadiq, R., Shahid, Z., Alam, M. M., & Mohd Su'ud, M. (2024). Sentiment Analysis using Support Vector Machine and Random Forest. *Journal of Informatics and Web Engineering*, *3*(1), 67–75. https://doi.org/10.33093/jiwe.2024.3.1.5

Ma'aly, A. N., Pramesti, D., Fathurahman, A. D., & Fakhrurroja, H. (2024). Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-Label Classification with a Deep Learning Algorithm. *Information*, *15*(11), 705. https://doi.org/10.3390/info15110705

Mantik, J., & Singgalen, Y. A. (2024). Performance evaluation of SVM with synthetic minority over-sampling technique in sentiment classification. In *Mantik Journal* (Vol. 8, Issue 1). Online.

Novianti, D. N., Shiddieq, D. F., Roji, F. F., & Susilawati, W. (2024). Komparasi Algoritma Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Metaverse. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 231–239. https://doi.org/10.57152/malcom.v4i1.1061

Paath, R. N. P., Suprapto, A. Q., Hayunaji, H., & Irfandianto, M. E. (2024). Political Labeling on Social Media: a Netnographic Study of "SDM Rendah" in Indonesia's 2024 Presidential Election on X (Twitter). *Eduvest - Journal of Universal Studies*, *4*(10), 8798–8811. https://doi.org/10.59188/eduvest.v4i10.38818

Pandey, R., & Singh, J. P. (2023). BERT-LSTM model for sarcasm detection in code-mixed social media post. *Journal of Intelligent Information Systems*, *60*(1), 235–254. https://doi.org/10.1007/s10844-022-00755-z

Patsiaouras, G. (2022). Marketplace cultures for social change? New social movements and consumer culture theory. *Journal of Marketing Management*, *38*(1–2), 17–47. https://doi.org/10.1080/0267257X.2021.1996443

Prasetyo, B., Al-Majid, A. Y., & Suharjito. (2024). A Comparative Analysis of MultinomialNB, SVM, and BERT on Garuda Indonesia Twitter Sentiment. *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, *12*(2), 445–454. https://doi.org/10.33558/piksel.v12i2.9966

Rewatkar, B., Barhate, A., & Verma, P. (2024). Application of Machine Learning Algorithms for Analyzing Sentiments Using Twitter Dataset. *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 1392–1397. https://doi.org/10.1109/ICSCSS60660.2024.10624990

Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M., & Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, *5*(6), 100994. https://doi.org/10.1016/j.patter.2024.100994

Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications*, *14*(6). https://doi.org/10.14569/IJACSA.2023.01406116

*name of corresponding author

Sanwal, M., & Mazhar, M. M. (2023). Performance Comparison of Machine Learning and Deep Learning Models for Sentiment Analysis of Hotel Reviews. *International Journal of Information Technology and Applied Sciences*, *5*(1), 2709–2208. https://doi.org/10.5281/zenodo.8225185

Seki, Y. (2024). *Citizen Sentiment Analysis*. https://doi.org/10.5772/intechopen.113030

Selian, R. I. P., Vitianingsih, A. V., Kacung, S., Maukar, A. L., & Rusdi, J. F. (2024). Sentiment Analysis of Public Responses on Social Media to Satire Joke Using Naive Bayes and KNN. *Sinkron*, *8*(3), 1443–1451. https://doi.org/10.33395/sinkron.v8i3.13721

Sharma, D. K., Singh, B., Agarwal, S., Kim, H., & Sharma, R. (2022). Sarcasm Detection over Social Media Platforms Using Hybrid Auto-Encoder-Based Model. *Electronics*, *11*(18), 2844. https://doi.org/10.3390/electronics11182844

Sharma, M., Kushwaha, R., & Srivastava, Mr. A. (2024). Sarcasm Detection on Social Media: Addressing the Issues of Short and Long Texts Using Machine Learning and Deep Learning Approaches. *International Journal of Scientific Research in Engineering and Management*, *08*(008), 1–8. https://doi.org/10.55041/IJSREM37364

Singgalen, Y. A. (2024a). Sentiment Classification of Food Influencer Content Reviews using Support Vector Machine Model through CRISP-DM Framework. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *5*(3), 517. https://doi.org/10.30865/json.v5i3.7509

Singgalen, Y. A. (2024b). Sentiment Classification Using BERT-CNN and SMOTE: A Case Study on Hotel Reviews Dataset. *Technology and Science (BITS)*, *6*(3). https://doi.org/10.47065/bits.v6i3.6309

Slim, S. O., Aboutabl, A. E., & Yacoub, A. D. (2024). A Survey of Sentiment Analysis and Sarcasm Detection. *International Journal of Electrical and Computer Engineering Systems*, *15*(1), 69–78. https://doi.org/10.32985/ijeces.15.1.7

Syafia, A. N., Hidayattullah, M. F., & Suteddy, W. (2023). Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, *8*(3).

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. In *Applied Sciences (Switzerland)* (Vol. 13, Issue 7). MDPI. https://doi.org/10.3390/app13074550

Thomas, S., Informasi, T., & Shanti Bhuana, I. (2021). Studi Analisis Metode Analisis Sentimen pada YouTube. *JIFOTECH (Journal of Information Technology)*, *1*(1).

Wahdiyana, F., Sari, S. P., Muslikhah, F. P., Ernawati, E., & Nirwana, T. P. (2024). Strategi Pencegahan Penggunaan Produk Israel dalam Gerakan Solidaritas terhadap Palestina di Kp. Tegal Waru Kecamatan Ciampea. *Jurnal Pengabdian Kepada Masyarakat Program Studi Komunikasi Penyiaran Islam*. http://jurnal-inais.id/index.php/SDDJ

Wang, T. (2024). A BERT-based with fuzzy logic sentimental classifier for sarcasm detection. *Applied and Computational Engineering*, *88*(1), 201–206. https://doi.org/10.54254/2755-2721/88/20241727

Witro, D. (2024). State Islamic University Students' Perceptions of IsraelAffiliated Products: A Study After the Fatwa of Indonesian Ulema Council No. 83 of 2023 Concerning the Law on Support for the Palestinian Struggle. *Al-Manahaj: Jurnal Kajian Hukum Islam*, 145–160.

Wulandari, N., & Saefudin. (2024). Kampanye Gerakan Boikot Produk Pendukung Israel terhadap McDonald's Indonesia pada Akun TikTok @calon_surga8. *Argopuro: Jurnal Multidisiplin Ilmu Bahasa*. https://doi.org/10.8734/argopuro.v1i2.365

Zhang, Xiaoxia. , X. Qi. , T. Z. (2024). *Performance evaluation of Reddit Comments using Machine Learning and Natural Language Processing methods in Sentiment Analysis*. https://doi.org/doi: 10.48550/arxiv.2405.16810

*name of corresponding author