SAINTIFIK@: JURNAL PENDIDIKAN MIPA

E-ISSN: 2598-3822 | P-ISSN: 2087-3816

Available online at https://ejournal.unkhair.ac.id/Saintifik

Deep Learning sebagai Implementasi Teori Matematika untuk Model Dinamika Sentimen: Studi Kasus Isu 'BBM Oplosan' Pertamina

(Deep Learning as an Implementation of Mathematical Theory for Modeling Sentiment Dynamics: The Case of Pertamina's 'BBM Oplosan' Issue)

Andy Hermawan[a*], Adinda Prilly Cindana[b], Dian Margaretha Nainggolan[b], Rizky Jemal Safryan[b]

[a]*Teknik Informatika, FTIK, Universitas Indraprasta PGRI, Jakarta, Indonesia, 12530*
[b]*Purwadhika Digital Technology School, Jakarta, Indonesia, 12920*

[*]*Corresponding author: andyhermawan@unindra.ac.id*

Received 17-09-2025, Revised 17-10-2025, Accepted 18-10-2025, Published 18-10-2025

Keywords:

Deep Learning; Mathematical Modeling; Sentiment Analysis; Lexicon-Based Analysis; BBM Oplosan

ABSTRACT. Public sentiment dynamics provide a quantitative reflection of how societal trust and perception evolve during crises. This study implements mathematical theory through deep learning techniques to model changes in public sentiment surrounding Pertamina's "BBM Oplosan" (fuel adulteration) issue, which went viral in Indonesia in early 2025. Twitter (X) data containing the keyword "Pertamina" were collected across two temporal windows—before and after the issue's emergence. Sentiment was classified into positive, neutral, and negative categories using both lexicon-based analysis (InSet Lexicon) and deep learning architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a hybrid CNN-LSTM model. From a mathematical standpoint, deep learning serves as a functional approximation framework that minimizes loss through gradient-based optimization—an implementation of multivariable calculus and linear algebra principles. Results show that negative sentiment increased from 23.5% to 48.2%, while positive sentiment declined from 44.6% to 26.2%, indicating a significant erosion of public trust. The CNN model achieved the highest validation accuracy (~63%), though it exhibited signs of overfitting. This research demonstrates how mathematical models underlying deep learning can be effectively applied to analyze real-world social phenomena, offering a robust quantitative framework for monitoring and interpreting public opinion dynamics during corporate crises.

INTRODUCTION

Public sentiment plays a crucial role in shaping the reputation and credibility of institutions, particularly state-owned companies such as Pertamina, which is responsible for maintaining national energy stability. When allegations of "BBM oplosan" involving Pertamina emerged and went viral in early 2025, public discourse shifted significantly, underscoring how sensitive and volatile public trust can be during crises. Understanding these sentiment dynamics is therefore essential for developing effective communication strategies and crisis response frameworks.

Previous research has demonstrated that sentiment analysis is a powerful tool for monitoring shifts in public opinion across domains such as crisis communication, consumer behavior, and social media analytics [1], [2]. Studies employing lexicon-based approaches and deep learning architectures—such as CNNs, LSTMs, and hybrid models—have achieved strong performance in classifying sentiment from short-text social media data [3], [4]. However, most existing work focuses on general topics such as product reviews, tourism sentiment, and political discourse, while research specifically examining sentiment dynamics in the context of national energy crises remains limited.

This gap indicates the need for more domain-specific sentiment analysis to understand public reactions during critical events involving corporate accountability and socio-economic concerns. The novelty of this study lies in integrating the mathematical foundations of deep learning with sentiment analysis to examine how public perception of Pertamina evolves before and after the viral issue. Unlike prior studies that primarily emphasize model accuracy, this research highlights mathematical interpretability by framing deep learning as a functional approximation problem grounded in optimization, linear algebra, and multivariable calculus.

Based on this background, the research formulates three core problems: how public sentiment toward Pertamina changed before and after the viral "BBM oplosan" issue; which dominant narratives emerged in each period; and how effectively lexicon-based methods and deep learning models classify sentiment in this crisis context. In line with these problems, the study aims to analyze the dynamics of sentiment polarity before and after

the issue surfaced, identify shifts in narrative focus through unigram, bigram, and WordCloud analyses, and evaluate the performance of CNN, LSTM, and CNN–LSTM models relative to a lexicon-based baseline.

The results of this study are expected to contribute theoretically by demonstrating the mathematical principles underlying deep learning for sentiment analysis, and practically by providing a data-driven framework to support crisis communication and reputation management in the national energy sector.

RESEARCH METHOD

The methodological framework in this study integrates theoretical and technical foundations from recent advances in sentiment analysis research [1], [2]. The methodology is designed as a structured pipeline comprising data acquisition, text preprocessing, vectorized feature representation, exploratory data analysis, sentiment classification using both lexicon-based and deep learning approaches, and model evaluation.

The study begins with automated data collection from the social media platform X (formerly Twitter) using targeted web scraping. The keyword “Pertamina” is used to capture large-scale public discourse on the topic of interest. Web scraping is chosen to ensure high-volume, real-time acquisition of textual data, enabling longitudinal analysis of public sentiment trends [2]. To maintain data quality, duplicate tweets, retweets, and non-textual entries are filtered out before further processing.

The theoretical backbone of this study lies in Natural Language Processing (NLP) and text mining, which enable computational interpretation of human language [3], [4]. Sentiment analysis is a technique for detecting affective states and opinions embedded in text, and it has been effectively applied in crisis communication, marketing, and social behavior studies [5], [6].

Text preprocessing is crucial for ensuring textual data are clean, normalized, and suitable for downstream analysis. This stage includes lowercasing (case folding), character normalization, removal of punctuation, URLs, and special characters, slang normalization, tokenization, stopword removal, and stemming using the Sastrawi Stemmer [7], [8]. Each step is implemented to reduce linguistic variability and noise that could negatively impact model training [9]. The tokenized corpus is then standardized to maintain consistency across samples.

To convert text into numerical input suitable for machine learning, this study applies multiple vectorization methods to capture both syntactic and semantic properties of the corpus. Classical statistical approaches such as Bag-of-Words and TF-IDF provide frequency-based term weighting, enabling sparse but interpretable representations [4]. In parallel, distributed representations are generated using word embeddings such as Word2Vec [10], [11], GloVe, and FastText [12] to capture contextual and subword-level semantics, thereby improving model generalization [3].

Exploratory Data Analysis (EDA) is conducted to provide an empirical understanding of the textual data distribution. This includes frequency analysis, sentiment proportion visualization, and observation of topic trends over time [13]. WordCloud visualization is employed to display dominant terms, allowing rapid identification of emerging discourse patterns in public conversations [2], [14]. This step is essential in validating whether the collected dataset is representative and thematically coherent.

A lexicon-based method using the InSet Lexicon (Indonesian Sentiment Lexicon) serves as the baseline. The lexicon contains 3,609 positive and 6,609 negative entries, each assigned a sentiment polarity score ranging from −5 to +5. Tweets are classified into positive, negative, and neutral categories based on aggregated scores per token [6], [7]. Lexicon methods offer transparency in sentiment attribution and are valuable for comparison with more complex models [1].

To improve predictive performance, three deep learning architectures are evaluated: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a hybrid CNN–LSTM. CNNs are effective at identifying local n-gram patterns, while LSTMs capture long-term sequential dependencies in text [5], [15]. The hybrid CNN–LSTM model combines the advantages of both, allowing simultaneous extraction of local features and temporal context [8]. Attention mechanisms are integrated to enhance the model’s focus on salient parts of the text, improving interpretability and accuracy.

All models are trained using a stratified training set to ensure class balance. Hyperparameter tuning is performed through grid search to optimize learning rate, dropout rates, and embedding dimensions. Model evaluation is carried out using accuracy, precision, recall, and F1-score metrics [1], [6]. K-fold cross-validation is used to ensure robust results. Additionally, a comparative analysis of lexicon-based and neural models is conducted to highlight trade-offs between interpretability and predictive power.

From a mathematical perspective, each deep learning model can be viewed as a function approximation problem, aiming to find a mapping.

$$f\theta: Rn \rightarrow Rk \quad (1)$$

Parameterized by weights θ, which minimizes the loss function

$$L(theta) = \left(\frac{1}{N}\right) \sum_{\{i=1\}}^{\{N\}} loss\left(f_{theta(x_i)}, y_i\right) \quad (2)$$

With ℓ\ellℓ representing the cross-entropy loss between predicted and true sentiment labels. Optimization is achieved through **gradient descent**, updating parameters according to

$$\theta_{\{t+1\}} = \theta_t - \eta \nabla_\theta L(\theta_t) \quad (3)$$

Where η denotes the learning rate.

The CNN applies convolution operations.

$$h_j = \sigma\left(w_j x + b_j\right) \quad (4)$$

To extract local patterns, the LSTM employs **recurrent gates** that compute hidden states using.

$$h_t = \sigma\left(W_h h_{\{t-1\}} + W_x x_t + b\right) \quad (5)$$

Capturing sequential dependencies in text data. These formulations implement linear algebra (matrix multiplication), calculus (gradient-based optimization), and functional analysis (nonlinear activation functions).

To reduce overfitting and improve generalization, dropout regularization, early stopping, and limited epochs are applied. Despite these measures, short-text characteristics and class imbalance remained key challenges in model training.

By integrating mathematical theory with computational implementation, this methodology ensures a rigorous, structured approach to modeling public sentiment dynamics during a crisis. The mathematical formulations behind deep learning emphasize that these models are not merely heuristic but are grounded in formal optimization and functional representation theory.

RESULTS AND DISCUSSION

1. Sentiment Analysis Before and After the "BBM Oplosan" Issue

This sentiment analysis examined how public perceptions of Pertamina shifted before and after the emergence of the "BBM oplosan" (fuel adulteration) issue. Sentiment was categorized into three types—positive, neutral, and negative—representing the public's opinion trends toward the company.

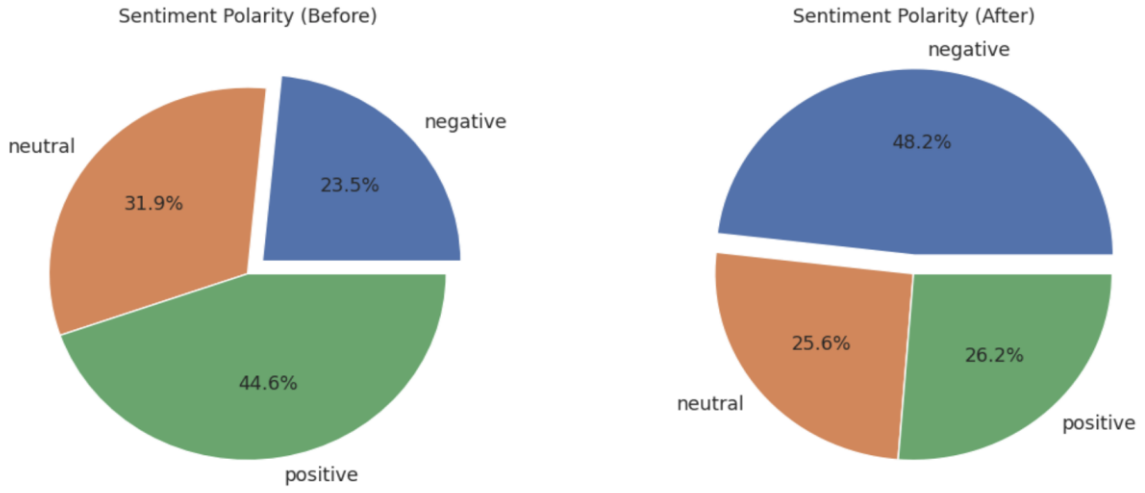

Figure 1. Sentiment Polarity Proportion Before the BBM Oplosan Issue and After the BBM Oplosan Issue.

Before the issue emerged, positive sentiment dominated at 44.6%, followed by neutral at 31.9% and negative at 23.5%. After the issue surfaced, public perception changed dramatically: negative sentiment surged to 48.2%, positive sentiment dropped sharply to 26.2%, and neutral sentiment decreased slightly to 25.6%.

This comparison indicates a substantial shift in public opinion—from a predominance of positive sentiment to a dominance of negative sentiment. The spike in negative sentiment reflects the strong impact of the fuel adulteration issue on Pertamina's image, signaling declining public trust and growing dissatisfaction following the scandal.

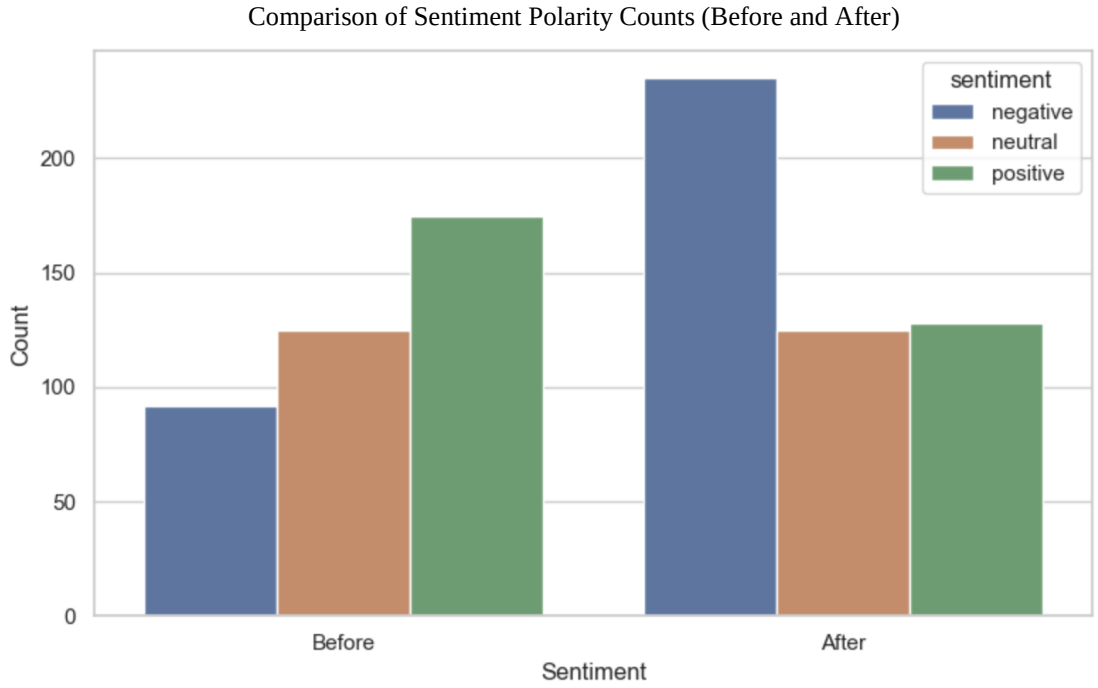

Figure 2. Sentiment Polarity Counts Before and After the BBM Oplosan Issue.

Before the issue, positive sentiment dominated, with around 180 tweets, while negative sentiment was about 90. Afterward, negative sentiment spiked to over 200 tweets, surpassing both positive and neutral sentiment.

This sharp rise in negative sentiment supports the earlier pie chart findings, which showed negative sentiment rising from 23.5% to 48.2%. It illustrates how the BBM oplosan issue significantly damaged Pertamina's public image, reflected in a steep decline in positive opinions and the dominance of negative opinions.

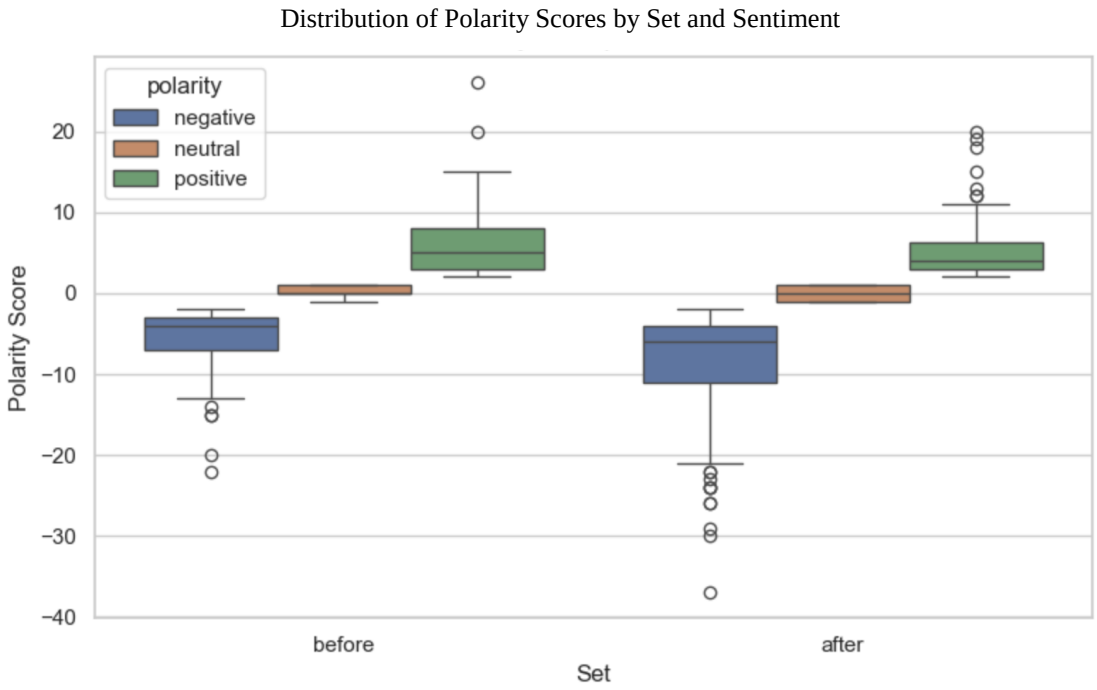

Figure 3. Sentiment Polarity Score Distribution.

The box plot shows a clear downward shift in negative sentiment scores after the issue emerged. The post-issue negative median is lower than the pre-issue median, indicating more intense negative expressions. More outliers appear after the issue, some below -30, signaling extreme negative reactions.

For positive sentiment, the median remains relatively stable but slightly declines, meaning positive opinions persisted but weakened in intensity. Neutral sentiment shows a narrow distribution with minimal variation, indicating stable neutral opinions.

The presence of extreme negative outliers reinforces the earlier findings: not only did negative sentiment increase, but its emotional intensity deepened, reflecting a surge of stronger public reactions after the scandal.

2. Analysis of Words Before and After the Viral Oplosan Pertamax Case

a) Unigram Frequency Analysis Before and After the Pertamax Adulteration Case Went Viral

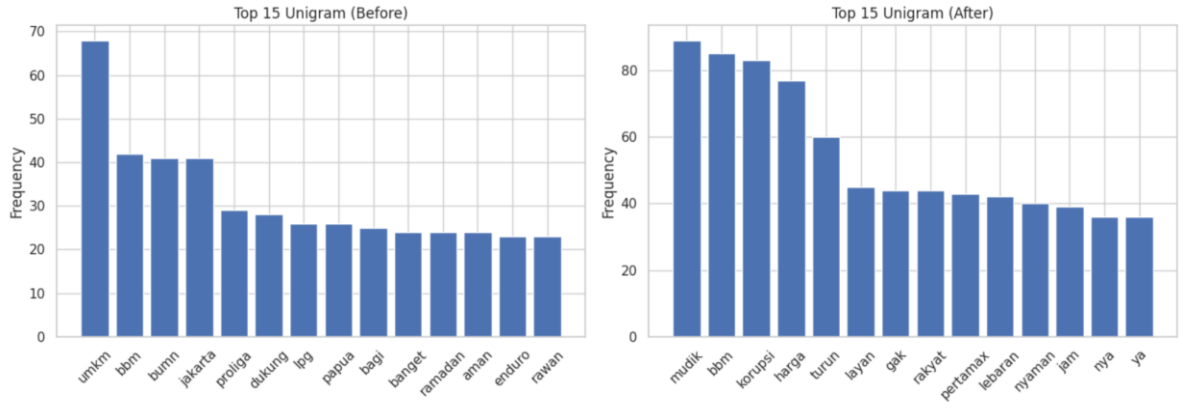

Figure 4. Top 15 Unigram Frequency Chart Before and After the Pertamax Adulteration Case.

Before the issue, dominant words such as “umkm,” “bbm,” “bumn,” and “jakarta” indicate conversations focused on corporate policy support and social contributions, particularly tied to government programs and economic activities. Words like “Ramadan” and “Aman” suggest a neutral to positive context centered on seasonal activities and fuel services.

After the issue went viral, dominant words shifted dramatically to “korupsi,” “harga,” “turun,” and “rakyat,” indicating rising public criticism regarding trust, transparency, and fuel pricing. Words like “pertamax” and “gak” further reveal increased negative perceptions and skepticism. Some neutral words, such as “mudik,” “lebaran,” and “nyaman,” remained, indicating ongoing functional or informational discussions, especially around holiday fuel availability.

Overall, unigram analysis reveals a discursive shift from institutional and supportive topics to criticism and public dissatisfaction, aligning with the sentiment analysis results, which show an increase in negative sentiment after the issue.

b) Bigram Frequency Analysis Before and After the Pertamax Adulteration Case Went Viral

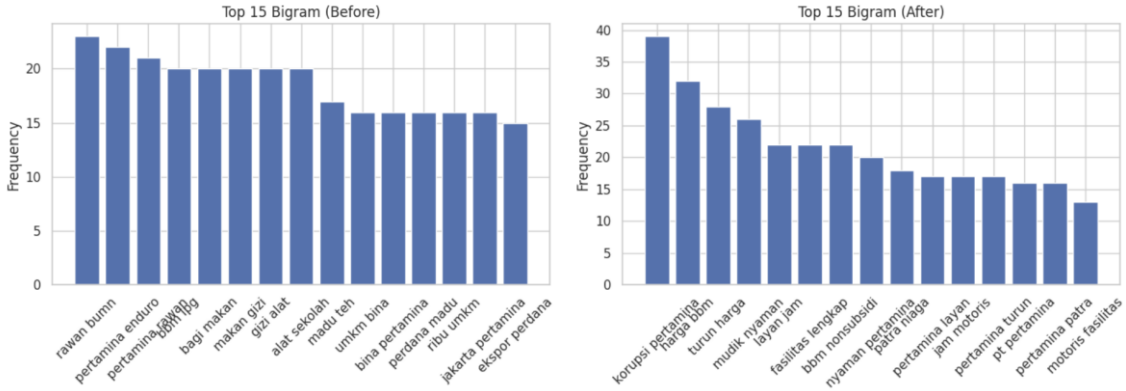

Figure 5. Top 15 Bigram Frequency Chart Before and After the Pertamax Adulteration Case.

Before the issue, bigrams like “rawan bumn,” “pertamina enduro,” “pertamina bumn,” and “bmh lpg” reflect diverse, mostly positive or neutral conversations that cover product promotions, corporate activities, and social responsibility programs.

After the issue, dominant bigrams shifted clearly toward negative and critical narratives: “korupsi pertamina,” “harga bbm,” and “turun harga” reveal more emotionally charged discourse. Phrases like “pertamina turun” and “pt pertamina” show how the corporate name became central to criticism. Some neutral bigrams, such as “mudik nyaman” and “fasilitas lengkap”, persisted, indicating residual neutral or positive conversations about operational services.

This bigram analysis confirms the unigram findings, showing a clear thematic shift from corporate and social narratives to issues of pricing and trust.

c) WordCloud Analysis of Negative Sentiment Before and After the Pertamax Adulteration Case Went Viral

Figure 6. WordCloud of Negative Sentiment Toward Pertamina Before and After the Pertamax Adulteration Case.

Before the issue, negative sentiment was weak and mostly associated with corporate activities like promotions or social programs. Words like “dukung,” “umkm,” “bbm,” and “lpg” dominated.

After the issue, the negative word frequency and character changed drastically. Words like “harga,” “bbm,” “pertamax,” “nonsubsidi,” “turun,” and “oplos” became dominant, along with “rakyat,” “korupsi,” and “subsidi,” showing a shift toward structural distrust and moral outrage.

The WordCloud confirms the CNN, LSTM, and CNN-LSTM sentiment analysis results: negative sentiment not only increased in volume but also deepened in emotional and thematic intensity.

3. Sentiment Model Analysis

a) CNN Model

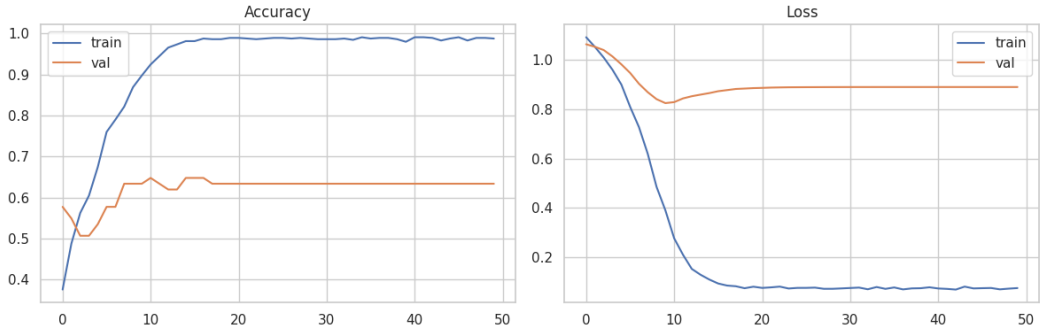

Figure 7. CNN Model Training and Validation Accuracy and Loss

Figure 7 depicts the training and validation performance of the CNN model. Training accuracy rapidly approaches 100% within the first 10 epochs, while validation accuracy plateaus at approximately 63% early on. The training loss drops sharply and stabilizes at a very low level, but the validation loss stagnates after an initial decline, indicating strong overfitting. This pattern suggests the model fits the training data almost perfectly but fails to generalize to unseen samples.

CNN’s ability to capture local n-gram patterns allows for fast convergence during training. However, its lack of temporal or contextual modeling, combined with limited regularization, makes it prone to memorizing the training set rather than learning robust sentiment features. The early saturation of validation performance implies that the model may be under-regularized relative to its capacity.

Table 1. CNN Confusion Matrix

	Pred Negative	Pred Neutral	Pred Positive
Actual Negative	40	21	4
Actual Neutral	12	36	2
Actual Positive	10	16	35

Table 1 provides a clearer picture of how this overfitting manifests in classification behavior. The model correctly classifies 40 of 65 negative samples, 36 of 50 neutral samples, and 35 of 61 positive samples. This indicates strong performance on the majority of samples, but the distribution of misclassifications reveals systematic tendencies:

1. Positive sentiment is mostly predicted correctly but shows low recall, with 12 and 10 positive samples being misclassified as negative and neutral, respectively.
2. Neutral sentiment exhibits high recall but low precision, as many misclassified samples from other classes are absorbed into the neutral category.
3. Negative sentiment shows moderate recall and precision, indicating more balanced performance.

The dominant misclassification occurs between neutral and positive, which often share overlapping lexical cues in short-text sentiment analysis.

b) LSTM Model

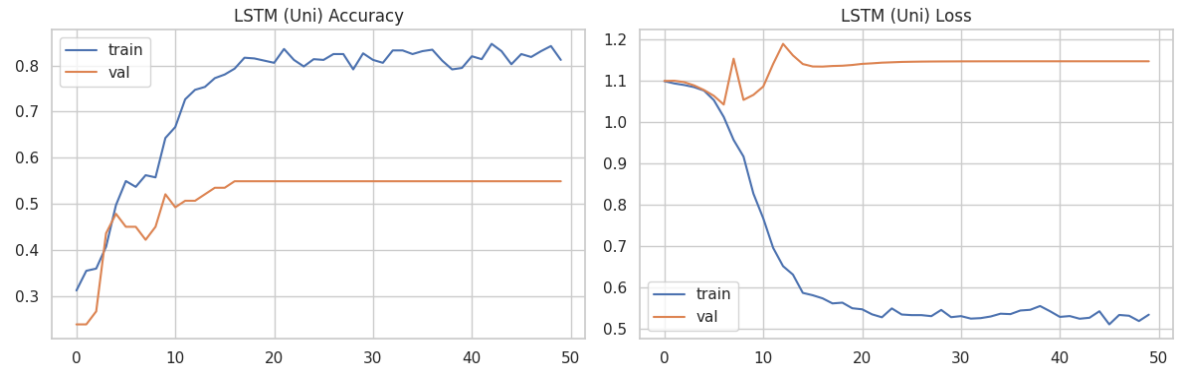

Figure 8. LSTM Model Training and Validation Accuracy and Loss

Figure 8 illustrates the training and validation performance of the LSTM model. Training accuracy steadily increases and surpasses 80% after approximately 20 epochs, while validation accuracy plateaus early at around 48%. This growing gap between the two curves is a clear indicator of severe overfitting. The validation loss curve diverges after the first few epochs and remains high. In contrast, the training loss continues to decline—further supporting the conclusion that the model fails to generalize beyond the training data.

This pattern reflects LSTM's sensitivity to limited and short-length input sequences. Although LSTM is designed to capture long-term dependencies, its representational power becomes underutilized when text spans are short, leading to rapid memorization of training patterns and weak generalization on unseen samples.

Table 2. LSTM Confusion Matrix

	Pred Negative	Pred Neutral	Pred Positive
Actual Negative	35	18	12
Actual Neutral	9	31	10
Actual Positive	10	32	19

Table 2 provides a detailed view of the LSTM model's classification behavior. The model correctly classifies 35 of 65 negative samples, 31 of 50 neutral samples, and only 19 of 61 positive samples. A considerable number of positive instances are misclassified as neutral (32 samples), which shows the model's difficulty in identifying positive sentiment.

Across all classes, the neutral label accounts for most misclassifications, indicating that the model tends to default to neutral when the sentiment signal is weak or ambiguous. This likely results from overlapping lexical or syntactic cues between neutral and positive language in the dataset. Negative sentiment, on the other hand, achieves better recall, likely due to stronger, more distinct linguistic markers in negative expressions.

c) CNN–LSTM Model

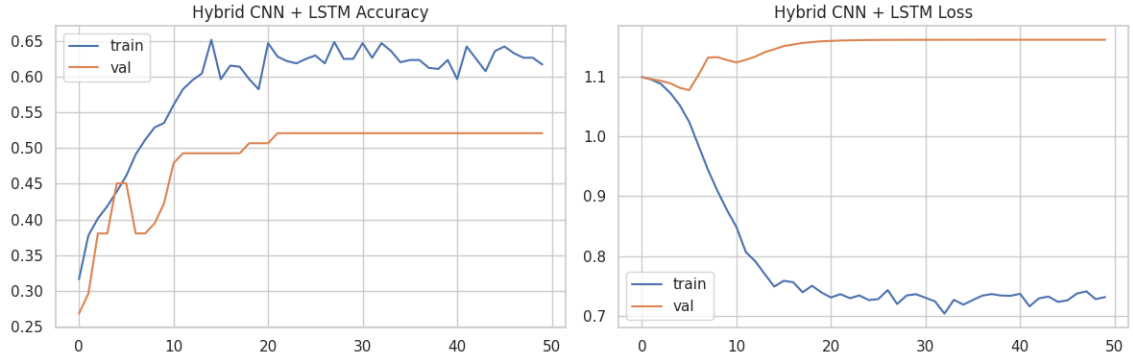

Figure 9. CNN–LSTM Model Training and Validation Accuracy and Loss

Figure 9 shows the training and validation curves for accuracy and loss of the hybrid CNN–LSTM model. The training accuracy gradually increases and stabilizes around 63%, while validation accuracy plateaus early at approximately 50%, indicating a notable generalization gap. The divergence between training and validation loss after roughly 10 epochs further confirms moderate overfitting. While the training loss continues to decrease steadily, the validation loss rises and stabilizes at a higher level, showing that the model fails to adapt to unseen data despite its more complex architecture.

This pattern suggests that the CNN–LSTM combination captures local patterns effectively but struggles to generalize across diverse short-text samples. The early plateau in validation accuracy is consistent with insufficient representational capacity for capturing nuanced sentiment, especially in imbalanced data distributions. The increasing validation loss also indicates that the model is beginning to memorize the training data rather than learning generalizable sentiment cues.

Table 3. CNN–LSTM Confusion Matrix

	Pred Negative	Pred Neutral	Pred Positive
Actual Negative	40	16	9
Actual Neutral	10	31	9
Actual Positive	21	23	17

Table 3 presents the confusion matrix for the CNN–LSTM model. The model correctly classified 40 out of 65 negative samples, 31 out of 50 neutral samples, and only 17 out of 61 positive samples. The recall for the positive class is particularly low, as a large proportion of positive instances were misclassified as either negative (21 samples) or neutral (23 samples). This imbalance reflects the model's difficulty in distinguishing positive sentiment, likely due to overlapping lexical patterns with neutral expressions in short texts.

A closer inspection reveals a systematic bias toward the negative and neutral classes. Negative predictions dominate across misclassified samples, suggesting that the decision boundary is skewed. This is consistent with the class imbalance often present in real-world sentiment datasets, where negative narratives tend to be more varied and expressive than positive ones.

CONCLUSION

The viral Pertamax adulteration issue substantially reversed sentiment toward Pertamina: positive dominance gave way to negative sentiment with deeper intensity. Narratives shifted from institutional/operational topics toward price transparency, fairness, and governance. Among the tested models, CNN was the most stable yet still overfitted on short texts with imbalanced labels. Future work should adopt pre-trained embeddings or Transformer architectures, expand data coverage, and balance classes to improve generalization.

REFERENCES

[1] A. Rajesh and T. Hiwarkar, "Sentiment analysis from textual data using multiple channels deep learning models," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, Nov. 2023, doi: 10.1186/s43067-023-00125-x.

[2] M. S. Islam *et al.*, "'Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach,'" *Artif Intell Rev*, vol. 57, no. 3, Mar. 2024, doi: 10.1007/s10462-023-10651-9.

[3] U. Krzeszewska, A. Poniszewska-Marańda, and J. Ochelska-Mierzejewska, "Systematic Comparison of Vectorization Methods in Classification Context," *Applied Sciences (Switzerland)*, vol. 12, no. 10, May 2022, doi: 10.3390/app12105119.

[4] H. D. Abubakar and M. Umar, "Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec," *SLU Journal of Science and Technology*, vol. 4, no. 1 & 2, pp. 27–33, Aug. 2022, doi: 10.56471/slujst.v4i.266.

[5] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models," Apr. 2019, doi: https://doi.org/10.48550/arXiv.1904.04206.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 2013, [Online]. Available: http://arxiv.org/abs/1301.3781

[7] M. E. Alzahrani, T. H. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," 2022, *Hindawi Limited*. Doi: 10.1155/2022/3840071.

[8] N. K. Gondhi, Chaahat, E. Sharma, A. H. Alharbi, R. Verma, and M. A. Shah, "Efficient Long Short-Term Memory-Based Sentiment Analysis of E-Commerce Reviews," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/3464524.

[9] H. Wu, Y. Gu, S. Sun, and X. Gu, *Aspect-based Opinion Summarization with Convolutional Neural Networks*. Vancouver: IEEE, 2016. Accessed: Oct. 22, 2025. [Online]. Available: https://doi.org/10.1109/IJCNN.2016.7727602

[10] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," Feb. 2014, [Online]. Available: http://arxiv.org/abs/1402.3722

[11] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1607.04606

[12] T. T. Allen, Z. Sui, and K. Akbari, "Exploratory text data analysis for quality hypothesis generation," *Qual Eng*, vol. 30, no. 4, pp. 701–712, Oct. 2018, doi: 10.1080/08982112.2018.1481216.

[13] A. Alasmari, N. Farooqi, and Y. Alotaibi, "Sentiment analysis of pilgrims using CNN-LSTM deep learning approach," *PeerJ Comput Sci*, vol. 10, pp. 1–47, 2024, doi: 10.7717/PEERJ-CS.2584.

[14] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, IEEE Computer Society, 2014, pp. 1833–1842. doi: 10.1109/HICSS.2014.231.

[15] L. Deng, T. Yin, Z. Li, and Q. Ge, "Analysis of the Effectiveness of CNN-LSTM Models Incorporating Bert and Attention Mechanisms in Sentiment Analysis of Data Reviews," 2024, pp. 821–829. doi: 10.2991/978-94-6463-238-5_106.