IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.x, No.x, July xxxx, pp. 1~5
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.xxxx

Multi-Domain Sentiment Analysis on Ibu Kota Nusantara (IKN) Tweets Using CNN-LSTM

Abstrak

Pembangunan Ibu Kota Nusantara (IKN) merupakan proyek nasional yang bertujuan untuk memindahkan ibu kota dari Jakarta ke Kalimantan Timur. Proyek ini memunculkan beragam opini pro dan kontra yang banyak disampaikan melalui media sosial seperti Twitter (sekarang dikenal sebagai X). Analisis sentimen terhadap opini ini menjadi penting untuk memahami persepsi publik terhadap proyek IKN. Namun, penelitian analisis sentimen terdahulu sering kali tidak mempertimbangkan variasi domain dalam data yang dianalisis, seperti ekonomi, lingkungan, dan politik, yang memiliki karakteristik bahasa yang berbeda.

Penelitian ini bertujuan untuk mengembangkan model analisis sentimen multi-domain dengan membandingkan tiga metode utama: CNN-LSTM, CNN, dan LSTM. Model multi-domain ini dirancang untuk mengatasi perbedaan karakteristik dari masing-masing domain dan meningkatkan kemampuan model dalam menangkap pola sentimen yang lebih kompleks. Selain itu, penelitian ini menerapkan dua pendekatan embedding, yaitu word embedding untuk memahami konteks secara luas dan keyword embedding untuk membantu model fokus pada kata kunci penting dalam setiap domain.

Hasil penelitian menunjukkan bahwa model multi-domain lebih unggul dibandingkan model single-domain, karena mampu meningkatkan performa klasifikasi dengan memanfaatkan informasi dari berbagai domain. CNN-LSTM terbukti menjadi model terbaik, dengan keseimbangan optimal antara Accuracy dan F1-Score dalam berbagai skenario. Penggunaan Keyword Embedding juga terbukti meningkatkan performa model, terutama pada LSTM, yang sebelumnya memiliki performa terendah. Selain itu, eksperimen menunjukkan bahwa Keyword Embedding dengan 5 keyword lebih optimal dibandingkan 10 keyword, karena memberikan hasil yang lebih stabil dan menghindari penurunan performa pada model CNN dan CNN-LSTM.

Kata kunci*— Analisis Sentimen Multi-Domain, Ibu Kota Nusantara (IKN), CNN-LSTM, Word Embedding, Keyword Embedding*

Abstract

The construction of Ibu Kota Nusantara (IKN) is a national project aimed at relocating Indonesia's capital from Jakarta to East Kalimantan. This project has sparked various public opinions, both in favor and against, which are widely expressed through social media platforms such as Twitter (now known as X). Sentiment analysis of these opinions is crucial for understanding public perception of the IKN project. However, previous sentiment analysis studies have often overlooked domain variations in the analyzed data, such as economy, environment, and politics, each of which has distinct linguistic characteristics.

This study aims to develop a multi-domain sentiment analysis model by comparing three main methods: CNN-LSTM, CNN, and LSTM. The multi-domain model is designed to address the differences in characteristics across domains and enhance the model's ability to capture more complex sentiment patterns. Additionally, this study implements two embedding approaches: word embedding for broader contextual understanding and keyword embedding to help the model focus on key terms specific to each domain.

Received June 1st,2012; Revised June 25th, 2012; Accepted July 10th, 2012

The results indicate that multi-domain models outperform single-domain models, as they improve classification performance by leveraging information from multiple domains. CNN-LSTM proved to be the best model, achieving the most balanced Accuracy and F1-Score across various scenarios. The use of Keyword Embedding also significantly enhances model performance, particularly benefiting LSTM, which initially had the lowest performance. Moreover, experiments demonstrate that Keyword Embedding with 5 keywords is more optimal than 10 keywords, as it provides more stable results and prevents performance degradation in CNN and CNN-LSTM models.

Keywords— *Multi-Domain Sentiment Analysis, Ibu Kota Nusantara (IKN), CNN-LSTM, Word Embedding, Keyword Embedding*

1. INTRODUCTION

The development of Ibu Kota Nusantara (IKN) aims to relocate Indonesia's capital from Jakarta to East Kalimantan, promoting regional development, economic growth, and national competitiveness. However, the project has sparked both support and opposition, with public opinions widely expressed on social media platforms like Twitter (now known as X). As a widely used platform, Twitter allows users to share messages or "tweets" with a 280-character limit [1]. Given its extensive use for public discourse, Twitter serves as a valuable source for sentiment analysis, enabling a deeper understanding of public perceptions regarding the IKN project.

Sentiment analysis is an automated process that involves understanding, extracting, and processing textual data to identify sentiment-related information within a statement or opinion. The goal of sentiment analysis is to determine whether a given opinion is positive or negative [2]. Various studies have explored sentiment analysis on IKN using machine learning and deep learning techniques, including Naïve Bayes [2], Neighbor Weighted K-Nearest Neighbor [3], Support Vector Machine [4], and Long Short-Term Memory (LSTM) [1]. However, these studies primarily focused on general sentiment classification without considering domain-specific variations, such as economy, environment, and politics. A key challenge in sentiment classification is domain dependency, where models trained on one domain may not perform well on others [5]. Multi-domain sentiment analysis addresses this issue by developing models that generalize across multiple domains while capturing domain-specific sentiment patterns.

Deep learning offers promising methods for multi-domain sentiment analysis, such as the Domain Attention Model that utilizes LSTM networks with an attention mechanism [6]. Similarly, the Collaborative Attention Neural Network (CANN) proposed by [7] employs BiLSTM-based self-attention and domain attention modules. The multi-domain approach has also been applied in sarcasm detection [8] and fake review detection [9], both using a hybrid CNN-LSTM architecture. The CNN-LSTM model leverages CNN for feature extraction, LSTM for long-term dependencies, and pooling layers for dimensionality reduction, making it suitable for handling multi-domain sentiment classification [8], [9].

This study proposes a CNN-LSTM-based multi-domain sentiment analysis model for Twitter discussions on IKN, focusing on economy, environment, and politics. The combination of CNN and LSTM enables efficient extraction of local patterns and long-term dependencies. To enhance model performance, word embedding provides broader contextual understanding, while keyword embedding emphasizes domain-specific terms. This integrated approach aims to improve sentiment classification accuracy while maintaining generalization across multiple domains.

2. METHODS

The research stages carried out can be seen in Figure 1.

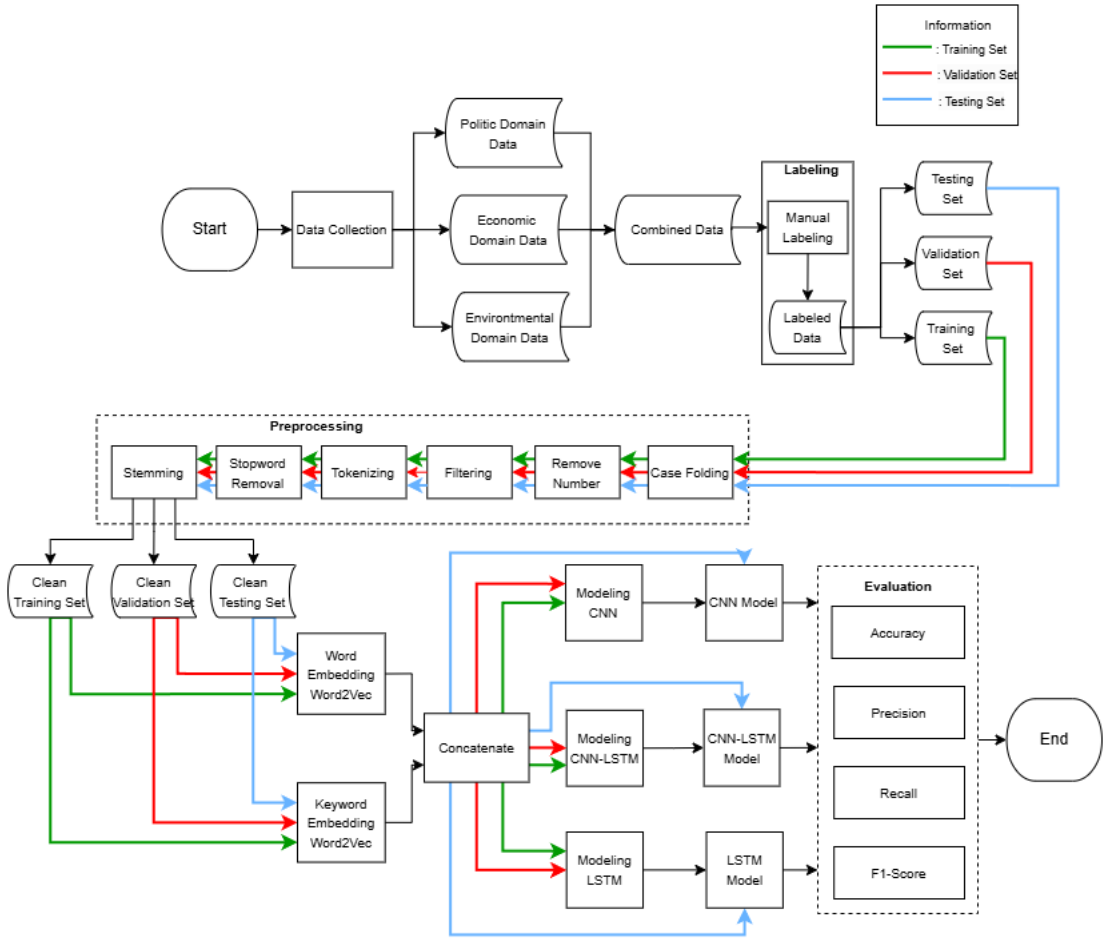

Figure 1 Research Stages

2.1 Data Collection

Data on tweets about the economy, environment, and politics of Ibu Kota Nusantara (IKN) was collected from Twitter/X between November 2023 and May 2024, covering the period before and after the election to ensure rich and relevant data. The data collection was conducted using a crawling technique, an automated method for gathering data from websites based on user-specified keywords [10], utilizing the tweet-harvest library. Crawling was performed periodically using the keywords listed in Table 1.

Table 1 Crawling process keywords

Domain	Keyword
Economy	"ekonomi ikn"
Environment	"lingkungan ikn"
Politic	"politik ikn"

Each keyword generated a dataset containing 1,500 tweets per domain, which were then saved in **.csv** format. The total collected dataset consists of 4,500 tweets, evenly distributed across the three domains (economy, environment, and politics).

2.2 Data Labeling

The collected dataset was then curated to check for duplicate tweets by directly comparing the text to ensure no repeated content. After curation, the final dataset ready for labeling consisted

of 1,339 tweets in the economy domain, 1,323 in the environment domain, and 1,450 in the politics domain, totaling 4,112 tweets for the multi-domain dataset.

Labeling was performed manually by categorizing tweets into positive and negative sentiment. Three Master's students in Artificial Intelligence at Universitas Gadjah Mada were involved in the labeling process to ensure consistency and reliability. Each annotator was responsible for labeling sentiment across all three domains. In cases of disagreement, majority voting was used to determine the final label. Examples of labeled data can be found in Table 2.

Table 2 Example of Data Labeling

Data	Annotator 1	Annotator 2	Annotator 3	Final Label
IKN Nusantara akan berperan sebagai pusat ekonomi baru dan simbol kemajuan teknologi masa depan bangsa. https://t.co/6YnV6gX0vx	Positive	Positive	Positive	Positive
IKN itu kota yg tiba2 dibangun dari nol beda dengan perkembangan kota pada umumnya kalo ga ada perkembangan industri atau ekonomi lainnya siap2 jadi kota mati	Negative	Negative	Negative	Negative

After the labeling process was completed, the next step was to perform validation using Kappa's Statistic, specifically Cohen's Kappa [11] and Fleiss' Kappa [12], as conducted by [13] in dataset creation, to evaluate the level of agreement among annotators. The levels of agreement used in Cohen's Kappa and Fleiss' Kappa can be seen in Table 3.

Table 3 Levels of Agreement in Cohen's Kappa and Fleiss' Kappa

Kappa	Level of Agreement
> 0.8	Almost Perfect
0.6 – 0.8	Substantial
0.4 – 0.6	Moderate
0.2 – 0.4	Fair
0 – 0.2	Slight
< 0	Poor

2.3 Data Splitting

This study uses a dataset split of 70% for training, 15% for validation, and 15% for testing. In the multi-domain scenario, the datasets from the three domains were first combined and then split using a 70:15:15 ratio. The same splitting method was applied in the single-domain scenario. Table 4 presents the dataset distribution for each domain and the multi-domain scenario.

Table 4 Split Dataset Distribution for Each Domain

Domain	Train	Validation	Test
Economy	937	201	201
Politic	1015	218	218
Environment	926	199	199
Total	2878	617	618

2.4 Preprocessing

In this research, preprocessing includes case folding, number removal, and filtering to clean text from punctuation, URLs, mentions, and irrelevant elements. Tokenization then splits the text into words, followed by stopword removal to eliminate insignificant words and stemming to convert words into their root forms.

2.5 Word Embedding

Word embedding is a feature learning method that maps each word or phrase in the vocabulary into a real-valued vector with a specific dimension [14]. This study utilizes the Word2Vec model [15]. Word2Vec generates word embeddings by mapping words into a continuous vector space. The pre-trained Word2Vec model used in this study was downloaded from the GitHub repository (https://github.com/deryrahman/word2vec-bahasa-indonesia). It was trained on the Indonesian Wikipedia corpus using Skip-gram, with a 300-dimensional vector size and a context window of 10 words. In this study, the word embeddings obtained from Word2Vec will be combined with keyword embeddings.

2.6 Keyword Embedding

The process of creating keyword embeddings begins with keyword extraction for each domain using the YAKE library. YAKE identifies the most relevant keywords or phrases statistically without requiring a large training dataset [16]. Keyword extraction is performed on the training data, selecting the top five keywords with the highest relevance scores for each domain. These keywords are then transformed into numerical vectors using a method similar to Word2Vec. Keywords2Vec modifies the Continuous Bag of Words (CBOW) architecture to represent each keyword as a single unit rather than individual words [17]. The resulting keyword embeddings have the same 300-dimensional structure as word embeddings, ensuring consistency and facilitating their combination.

2.7 Embedding Merging Process

The combination of word embedding and keyword embedding is performed using the concatenation method, where both embedding vectors are merged horizontally to create a higher-dimensional word representation. Each word in the text is represented by a 300-dimensional word embedding and a 300-dimensional keyword embedding. Through concatenation, these embeddings are combined into a single 600-dimensional vector for each word, as illustrated in Figure 2.

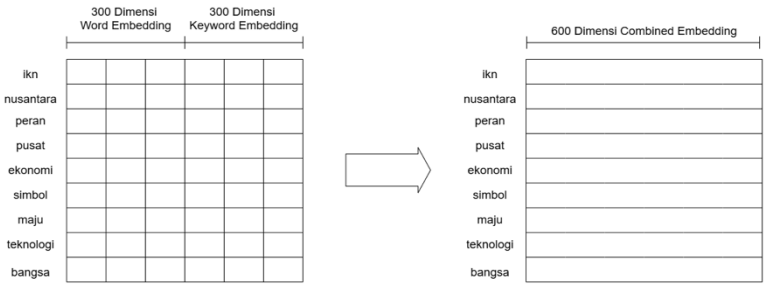

Figure 2 Illustration of the Embedding Merging Process

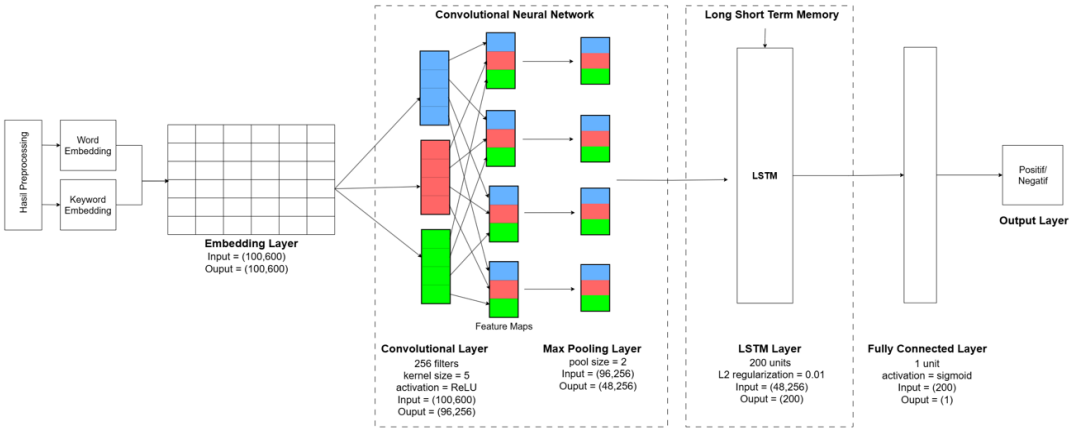

Figure 3 CNN-LSTM Architechture

2.8 Modeling

In this study, modeling was conducted using three methods: CNN-LSTM, CNN, and LSTM to compare their performance. The CNN-LSTM model architecture, illustrated in Figure 3, processes preprocessed tweets from economy, environment, and politics by converting them into embedding vectors using Word2Vec for word embedding and Keywords2Vec for keyword embedding, which are then combined in the embedding layer (600 dimensions). The CNN layer extracts textual patterns (e.g., n-grams) using ReLU activation, followed by a max pooling layer to reduce dimensionality while retaining key features. Extracted features are then passed to the LSTM layer, which captures contextual relationships and long-term dependencies in text sequences. A fully connected layer integrates these features before the sigmoid activation function in the output layer classifies tweets as positive or negative. The model's hyperparameters, obtained through grid search tuning, include 256 filters, a kernel size of 5, and ReLU activation for the convolutional layer. A max pooling layer with pool size 2 helps dimensionality reduction. The LSTM layer consists of 200 units with L2(0.01) regularization. A dropout layer with a rate of 0.7 is applied to prevent overfitting. The final fully connected layer has 1 unit with a sigmoid activation function for binary classification.

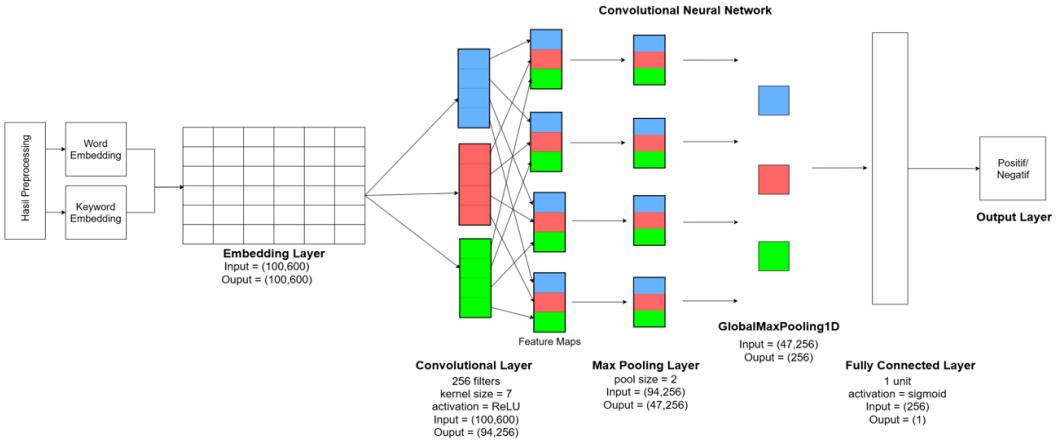

Figure 4 CNN Architechture

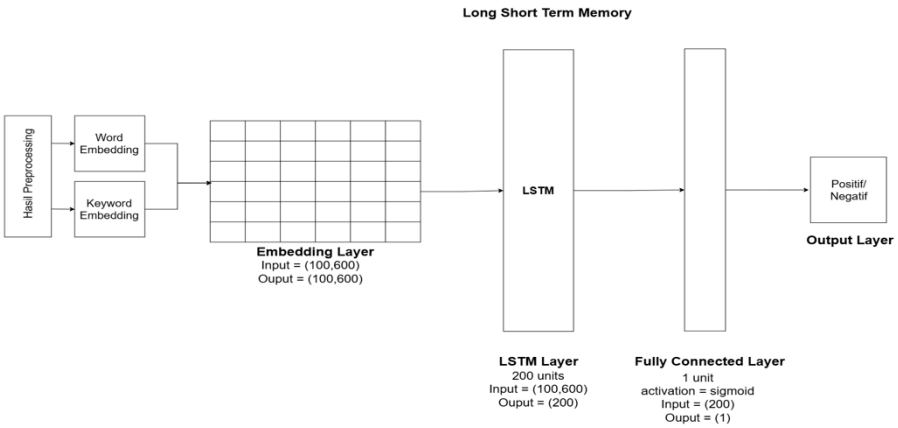

Figure 5 LSTM Architechture

One-dimensional CNN (1D CNN) is a variant of CNN that specifically handles sequential data or data that has one main dimension. In 1D CNN, the convolution operation is performed by shifting the convolution filter along one dimension of the data [18]. As shown in Figure 4, the model processes preprocessed data into 600-dimensional word and keyword embeddings. The convolutional layer, configured with 256 filters, a kernel size of 7, and ReLU activation, extracts key text patterns. Max pooling (pool size = 2) reduces dimensionality while retaining essential features, followed by a dropout rate of 0.3 to prevent overfitting. A GlobalMaxPooling1D layer

further compresses features, and a fully connected layer with a sigmoid activation function classifies sentiment as positive or negative. Hyperparameters, including filter size, kernel size, dropout rate, and batch size, are obtained using grid search to improve performance in multiple domains.

LSTM enhances Recurrent Neural Networks (RNN) by addressing the vanishing gradient problem, enabling better retention of long-term dependencies [1]. As shown in Figure 5, the model processes preprocessed data into word and keyword embeddings, which are combined in the embedding layer. The LSTM layer, optimized with 200 units through hyperparameter tuning, captures sequential dependencies in text. A dropout rate of 0.5 mitigates overfitting, while a fully connected layer with a sigmoid activation function classifies sentiment as positive or negative. These parameters were optimized using grid search to improve model performance.

2.9 Evaluation

The evaluation in this study consists of two scenarios: single-domain dataset and multi-domain dataset, both assessed using a confusion matrix to calculate accuracy, precision, recall, and F1-score.

2.9.1 Scenario 1: Single-Domain Dataset

Each domain (economy, environment, and politics) is evaluated separately. The dataset is split into 70% training, 15% validation, and 15% testing, undergoing preprocessing, word embedding, and keyword embedding using Word2Vec. CNN-LSTM, CNN, and LSTM models are trained independently per domain, and their performance is evaluated based on accuracy, precision, recall, and F1-score.

2.9.2 Scenario 2: Multi-Domain Dataset

In this scenario, data from all domains are combined into a single multi-domain dataset, split into 70% training, 15% validation, and 15% testing. The models are trained on this merged dataset to assess their ability to handle cross-domain variations. Evaluation is conducted in two settings:

(1) Overall evaluation – assessing model generalization on the entire multi-domain test set.
(2) Per-domain evaluation – testing the trained multi-domain model on each domain's test set.

Additional experiments analyze the impact of keyword embedding and keyword quantity:

(1) Without vs. With Keyword Embedding – compares model performance using only word embedding vs. a combination of word and keyword embedding.
(2) 5 vs. 10 Keywords – evaluates whether increasing keyword quantity improves contextual understanding or introduces redundancy.

All models are trained with consistent parameters, and performance is assessed using accuracy, precision, recall, and F1-score.

3. RESULTS AND DISCUSSION

3.1 Data Labeling Results

Based on the results in Table 5, the data annotation process for the politics domain, analyzed statistically, achieved a "Fair" agreement level for both methods, Cohen's Kappa and Fleiss' Kappa. The Cohen's Kappa value of 0.3593 and Fleiss' Kappa value of 0.3057 indicate a fair level of agreement, meaning that the agreement among annotators is adequate but still has interpretational differences that can be improved.

The results in Table 6 show that the annotation process for the economic domain achieved a "Moderate" level of agreement using both Cohen's Kappa and Fleiss' Kappa methods, with Cohen's Kappa at 0.5037 and Fleiss' Kappa at 0.4812. These values indicate a sufficiently adequate level of agreement among annotators in this domain.

Table 5 Politics Domain Kappa Score Results

Politics Domain Kappa Score	
Pair	Kappa Value
A1-A3	0.0756
A2-A1	0.1631
A2-A3	0.8393
Average Cohen's Kappa Value	0.3593
Fleiss's Kappa Value	0.3057

Table 6 Economic Domain Kappa Score Results

Economic Domain Kappa Score	
Pair	Kappa Value
A1-A3	0.2590
A2-A1	0.9920
A2-A3	0.2601
Average Cohen's Kappa Value	0.5037
Fleiss's Kappa Value	0.4812

In Table 7, the annotation results for the environmental domain indicate a "Moderate" level of agreement based on statistical evaluation using Cohen's Kappa and Fleiss' Kappa methods. Cohen's Kappa reached 0.4913, while Fleiss' Kappa was 0.4785, demonstrating a sufficient level of agreement among annotators in this domain.

Table 7 Environment Domain Kappa Score Results

Environment Domain Kappa Score	
Pair	Kappa Value
A1-A3	0.2821
A2-A1	0.7997
A2-A3	0.3921
Average Cohen's Kappa Value	0.4913
Fleiss's Kappa Value	0.4785

The economic domain data consists of 840 positive data (62.7%) and 499 negative data (37.3%). In the political domain, there were 562 positive data (41.7%) and 786 negative data (58.3%). Meanwhile, data on the domain environment includes 873 positive data (66%) and 450 negative data (34%). The distribution of labels for each domain can be seen in Figure 6.

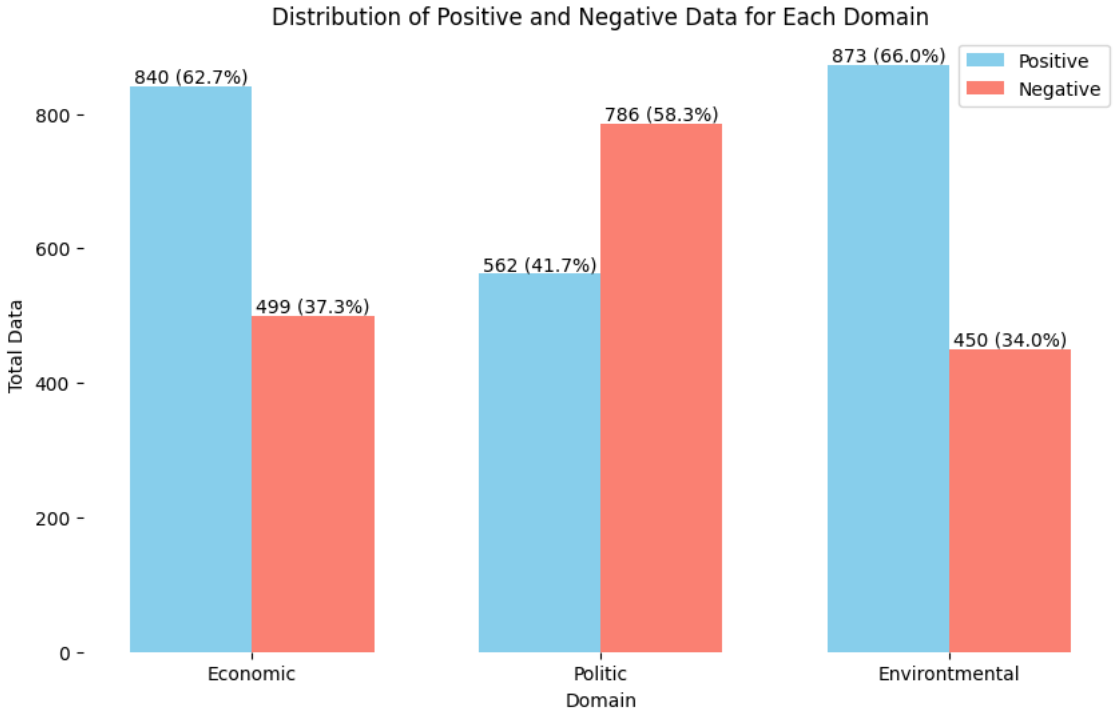

Figure 6 Label Distribution for each Domain

3.2 Evaluation Results

3.2.1 Scenario 1: Single-Domain Dataset

Based on Table 8, CNN-LSTM demonstrated the best performance across all domains. In the Economy domain, it achieved the highest Accuracy (83.58%) and F1-Score (83.20%), outperforming CNN, which showed slightly lower but stable results. LSTM had the lowest performance, struggling to recognize both classes effectively. In the Environment domain, CNN-LSTM again performed best, with Accuracy (87.43%) and F1-Score (85.77%), followed closely by CNN, while LSTM had the lowest Recall (85%), indicating more frequent misclassification of positive instances. In the Politics domain, CNN-LSTM had the highest Accuracy (83.94%), but lower Recall (58.18%) and F1-Score (59.76%), showing difficulty in detecting positive instances. CNN had high Precision (91.20%) but poor Recall (52.5%), while LSTM had the same Accuracy (82.56%) as CNN but a lower F1-Score (51.97%) and Recall (53.46%), struggling to identify positive class patterns.

Table 8 Comparison of Model Performance in Single Domain

Model	Domain	Accuracy	Precision	Recall	F1-Score
CNN-LSTM	Economic	83.58	82.96	84.72	83.20
	Environtment	87.43	85.15	86.54	85.77
	Politics	83.94	80.99	58.18	59.76
CNN	Economic	82.58	81.51	82.44	81.86
	Environtment	86.93	84.49	87.03	85.47
	Politics	82.56	91.20	52.5	49.93
LSTM	Economic	75.62	75.43	71.62	72.47
	Environtment	82.41	80.70	85	81.32
	Politics	82.56	78.85	53.46	51.97

3.2.2 Scenario 2: Multi-Domain Dataset

Based on Table 9, the performance of the multi-domain model in the Economy domain shows that CNN-LSTM demonstrated the best performance achieving the highest Accuracy (88.05%) and F1-Score (87.72%), indicating its strong ability to classify both positive and negative classes. CNN followed with slightly lower but balanced performance, while LSTM had the weakest results. In the Environment domain,

CNN achieved the highest Accuracy (88.44%) and F1-Score (87.28%), showing better balance in classification, with CNN-LSTM performing nearly as well. LSTM remained competitive but was less effective than the other models. In the Politics domain, CNN-LSTM and LSTM had equal Accuracy (84.40%), but CNN-LSTM performed better in capturing the positive class with higher Recall (64.28%) and F1-Score (67.14%), making it the most balanced model despite data imbalance. CNN had slightly lower Accuracy but outperformed LSTM in recognizing both classes. Overall, CNN-LSTM excelled in the Economy domain, CNN performed best in the Environment domain, and CNN-LSTM remained the most balanced model in the Politics domain.

Table 9 Comparison of Model Performance in Multi-Domain

Model	Domain	Accuracy	Precision	Recall	F1-Score
CNN-LSTM	Economics	88.05	87.28	89.09	87.72
	Environtment	87.93	85.58	88.62	86.68
	Politics	84.40	75.68	64.28	67.14
CNN	Economics	84.57	83.96	85.77	84.22
	Environtment	88.44	86.12	89.41	87.28
	Politics	83.94	77.05	60.12	62.41
LSTM	Economics	83.08	83.14	85.05	82.86
	Environtment	85.92	83.44	86.29	84.47
	Politics	84.40	78.36	61.37	64.06

The use of multi-domain models improves classification performance across domains, particularly in LSTM for Economy and Politics and F1-Score in Politics, demonstrating better recognition of complex patterns. In the Economy domain, CNN-LSTM showed the highest improvement (Accuracy: 83.58% → 88.05%, F1-Score: 83.20% → 87.72%), while LSTM had the most significant boost (Accuracy: 75.62% → 83.08%, F1-Score: 72.47% → 82.86%). The Environment domain saw smaller gains, with CNN remaining the best model (Accuracy: 86.93% → 88.44%, F1-Score: 85.47% → 87.28%). In the Politics domain, multi-domain models significantly improved F1-Score, with CNN-LSTM increasing from 59.76% to 67.14% and CNN from 49.93% to 62.41%, despite minor changes in Accuracy. These results indicate that multi-domain models are more effective in improving classification balance rather than just Accuracy, with CNN-LSTM excelling in Economy and Politics, while CNN remains the best in Environment.

a. Comparison of Multi-Domain Model Performance Without Keyword Embedding and With Keyword Embedding

The use of Keyword Embedding generally enhances the performance of the multi-domain model across all domains, with varying degrees of improvement. In the Economy domain, CNN-LSTM showed significant gains, with Accuracy increasing from 85.07% to 88.05% and F1-Score from 84.71% to 87.72%, followed by CNN and LSTM, which also improved. In the Environment domain, CNN remained the best model, while LSTM saw notable gains, indicating its reliance on additional contextual information. In the Politics domain, all models benefited significantly, with CNN-LSTM achieving a more balanced Precision-Recall tradeoff, CNN improving in class recognition, and LSTM showing the highest performance boost. These findings confirm that Keyword Embedding enhances classification accuracy, particularly for models requiring additional contextual understanding. The performance comparison is shown in Table 10.

Table 10 Comparison of Multi-Domain Model Performance Without Keyword Embedding and With Keyword Embedding

Scenario	Domain	CNN-LSTM				CNN				LSTM			
		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Without K.E	Eco	85.07	84.39	86.18	84.71	82.58	83.03	84.90	82.41	74.12	73.69	69.92	70.67
	Env	87.93	85.58	88.62	86.68	87.43	85.19	89.10	86.35	79.39	76.20	75.98	76.09
	Pol	83.48	73.30	61.78	64.20	82.56	71.82	56.37	57.07	80.73	64.67	58.16	59.41
With K.E	Eco	88.05	87.28	89.09	87.72	84.57	83.96	85.77	84.22	83.08	83.14	85.05	82.86
	Env	87.93	85.58	88.62	86.68	88.44	86.12	89.41	87.28	85.92	83.44	86.29	84.47
	Pol	84.40	75.68	64.28	67.14	83.94	77.05	60.12	62.41	84.40	78.36	61.37	64.06

b. Comparison of Multi-Domain Model Performance With 5 Keywords and 10 Keywords in Keyword Embedding

Using 10 keywords did not consistently improve model performance and in some cases led to a slight decline. In the Economy domain, CNN-LSTM and CNN experienced decreased Accuracy and F1-Score, indicating that additional keywords were not always beneficial, while LSTM showed improvement (Accuracy: 83.08% → 84.57%, F1-Score: 82.86% → 84.12%), suggesting better adaptability. In the Environment domain, increasing keywords had minimal impact, with slight declines in CNN-LSTM and CNN, while LSTM saw a marginal F1-Score improvement (84.47% → 85.47%). In the Politics domain, CNN-LSTM and CNN remained unchanged, but LSTM's performance dropped (Accuracy: 84.40% → 83.02%, F1-Score: 64.06% → 61.50%), indicating reduced model effectiveness with more keywords. Overall, 10 keywords did not always enhance classification performance and in some cases negatively affected model stability.The comparison of multi-domain model performance with 5 and 10 keyword embeddings is presented in Table 11.

Table 11 Comparison of Multi-Domain Model Performance with Keyword Embedding 5 Keywords and 10 Keywords

Scenario	Domain	CNN-LSTM				CNN				LSTM			
		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
5 Keyword	Eco	88.05	87.28	89.09	87.72	84.57	83.96	85.77	84.22	83.08	83.14	85.05	82.86
	Env	87.93	85.58	88.62	86.68	88.44	86.12	89.41	87.28	85.92	83.44	86.29	84.47
	Pol	84.40	75.68	64.28	67.14	83.94	77.05	60.12	62.41	84.40	78.36	61.37	64.06
10 Keyword	Eco	85.57	84.72	86.33	85.14	81.09	80	86.97	80.35	84.57	83.72	85.28	84.12
	Env	87.93	85.60	87.76	86.49	87.43	85.08	86.97	85.88	86.93	84.49	87.03	85.47
	Pol	84.40	75.68	64.28	67.14	83.94	77.05	60.12	62.41	83.02	72.36	59.56	61.50

4. CONCLUSIONS

The results of this study demonstrate that multi-domain models outperform single-domain models by leveraging cross-domain information to enhance classification performance. Among the evaluated models, CNN-LSTM consistently achieved the best balance between Accuracy and F1-Score, making it the most effective for sentiment classification across economy, environment, and politics domains. The use of Keyword Embedding significantly improved model performance, particularly for LSTM, which benefited the most from additional contextual information. However, experiments revealed that Keyword Embedding with 5 keywords was more effective than using 10 keywords, as excessive keywords led to redundancy and minor performance degradation, especially in CNN and CNN-LSTM models. Despite these improvements, the dataset used in this study had limitations, including class imbalance and the presence of irrelevant or ambiguous data, which may have affected model performance and generalization. While the proposed multi-domain CNN-LSTM model successfully enhanced sentiment classification, further improvements, such as better data curation, balancing techniques, larger datasets, alternative deep learning architectures (e.g., Transformer-based models or hybrid approaches with Attention Mechanisms), and the exploration of different embedding techniques such as FastText or GloVe, could further optimize multi-domain sentiment analysis in future studies.

REFERENCES

[1] Y. A. Pradana, I. Cholissodin, and D. Kurnianingtyas, "Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2389–2397, 2023.

[2] A. Sa'rony, P. P. Adikara, and R. C. Wihandika, "Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 10086–10094, 2019.

[3] M. I. D. Sakariana, Indriati, and C. Dewi, "Analisis Sentimen Pemindahan Ibu Kota Indonesia Dengan Pembobotan Term BM25 Dan Klasifikasi Neighbor Weighted K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 748–755, 2020.

[4] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[5] N. Lin, B. Chen, S. Fu, X. Lin, and S. Jiang, "Multi-domain Sentiment Classification on Self-constructed Indonesian Dataset," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12430 LNAI, no. March, pp. 789–801, 2020, doi: 10.1007/978-3-030-60450-9_62.

[6] Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, "Domain attention model for multi-domain sentiment classification," *Knowledge-Based Syst.*, vol. 155, no. May, pp. 1–10, 2018, doi: 10.1016/j.knosys.2018.05.004.

[7] C. Yue, H. Cao, G. Xu, and Y. Dong, "Collaborative attention neural network for multi-domain sentiment classification," *Appl. Intell.*, vol. 51, no. 6, pp. 3174–3188, 2021, doi: 10.1007/s10489-020-02021-7.

[8] R. Jamil, I. Ashraf, F. Rustam, E. Saad, A. Mehmood, and G. S. Choi, "Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/peerj-cs.645.

[9] S. N. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade, and T. H. H.

Aldhyani, "Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets," *Appl. Bionics Biomech.*, vol. 2021, 2021, doi: 10.1155/2021/5522574.

[10] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, "Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm," *J. Phys. Conf. Ser.*, vol. 1821, no. 1, 2021, doi: 10.1088/1742-6596/1821/1/012054.

[11] Cohen Jacob, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37-46 ST-A coefficient of agreement for nominal, 1960.

[12] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5. pp. 378–382, 1971, doi: 10.1037/h0031619.

[13] Riccosan, K. E. Saputra, G. D. Pratama, and A. Chowanda, "Emotion dataset from Indonesian public opinion," *Data Br.*, vol. 43, pp. 0–5, 2022, doi: 10.1016/j.dib.2022.108465.

[14] T. I. Z. M. Putra, S. Suprapto, and A. F. Bukhori, "Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia," *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022, doi: 10.35912/jisted.v1i1.1509.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, pp. 1–12, 2013.

[16] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, "YAKE! Keyword extraction from single documents using multiple local features," *Inf. Sci. (Ny).*, vol. 509, pp. 257–289, 2020, doi: 10.1016/j.ins.2019.09.013.

[17] J. Gabín, M. E. Ares, and J. Parapar, "Keyword Embeddings for Query Suggestion," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13980 LNCS, pp. 346–360, 2023, doi: 10.1007/978-3-031-28244-7_22.

[18] Y. Yuliska, D. H. Qudsi, J. H. Lubis, K. U. Syaliman, and N. F. Najwa, "Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, p. 1067, 2021, doi: 10.25126/jtiik.2021854842.