SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622

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

Sentiment Analysis of the Relocation of the National Capital on Social Media X

Yesi Ratna Dewi[1)*], Ni Wayan Sumartini Saraswati[2)], Maria Osmunda Eawea Monny[3)], Ida Bagus Gede Sarasvananda[4)], I Gede Andika[5)]
[1,2,3,4,5)]Institut Bisnis dan Teknologi Indonesia, Bali, Indonesia
[1)]yesiratna220203@gmail.com, [2)]sumartini.saraswati@instiki.ac.id, [3)]maria.monny@instiki.ac.id,
[4)]sarasvananda@gmail.com, [5)]gdandika@instiki.ac.id

Submitted :March 6, 2025 | **Accepted** : April 10, 2025 | **Published** : April 14, 2025

Abstract: The relocation of the national capital is a national strategic development project that seeks input from the public. This research analyzes public sentiment towards the relocation of the capital city using the Lexicon SVM method with data from X social media. The analysis was conducted in two languages, namely Indonesian and English, to find out how public opinion on the relocation of Indonesia's capital city at the global level. The sentiment classification results show that in Indonesian, public sentiment tends to be balanced with a model accuracy of 94%, where 51.55% is positive sentiment and 48.45% is negative. Meanwhile, in English, positive sentiment is more dominant with a model accuracy of 90%, where 87.63% is positive sentiment and 12.37% is negative sentiment. Evaluation using confusion matrix shows that this model provides good results, with high precision, recall, and F1-score values. Visualization using WordCloud and frequency analysis of unigrams, bigrams, and trigrams showed that positive sentiments mostly discussed the development aspects and government policies, while negative sentiments highlighted the social and economic impacts of the relocation. In addition, further analysis shows that public sentiment fluctuates based on important government announcements and major events related to the project. These findings demonstrate the importance of monitoring public opinion over time to understand shifts in perception. This research provides insights to the government and policymakers in understanding public opinion regarding the relocation of the nation's capital. By understanding sentiment patterns, more appropriate policies can be designed to increase public acceptance of the project and address public concerns effectively.

Keywords: Lexicon, Public Sentiment, Sentiment Analysis, Social Media, Support Vector Machine

INTRODUCTION

The relocation of the national capital announced by President Joko Widodo on August 26, 2019 from Jakarta to East Kalimantan is a major step and effort in encouraging equitable development in Indonesia (Nurdiyansyah, 2023). The selection of East Kalimantan as the new location of the national capital was based on various considerations, including its central geographical location and low risk of natural disasters, which support sustainable infrastructure development. The region also has a high level of accessibility, being adjacent to major cities such as balikpapan and samarinda, which already have major supporting infrastructure. A diverse population structure with low potential for conflict and the availability of critical infrastructure such as airports, ports and clean water sources support the relocation of the National Capital (M. K. Saraswati et al., 2022). Jakarta, which has long served as the core of Indonesian government and industry, is now facing significant challenges. The Central Bureau of Statistics reported that the population of Jakarta in September 2020 reached 10.56 million people, Since 2010, there has been a growth of around 956 thousand persons, or an average of 88 thousand people and is expected to continue to increase. This rapid population growth exacerbates the problems of traffic congestion, air pollution, and the risk of flooding, which is increasing due to land subsidence and climate change (Afifah et al., 2024).

The decision to relocate the country's capital city triggered a variety of opinions among the public, both domestic and foreign. Some are in favor because they see the positive potential for equitable development, the creation of new jobs, and increasing Indonesia's competitiveness abroad. However, there are also those who are concerned about the negative impacts, such as environmental issues, displacement of local residents, and the enormous cost

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

burden that must be incurred by the state which is the main issue. Some critics also cite the need for more in-depth research to ensure that the project truly provides long-term benefits for all Indonesians (Mega Hendarto, 2022). With the existence of diverse public responses, it is very important to analyze public sentiment towards this policy. Therefore, this research focuses on how to analyze public sentiment on social media X regarding the relocation of the National Capital City (IKN) using the Lexicon-SVM method?. According to data from goodstats on June 19, 2024 Indonesia occupies the 4th (fourth) position in the world as the most X users, this research utilizes data from social media X to analyze public sentiment, providing insight into opinion trends and public feelings on hotly discussed issues.

Utilizing the SVM lexicon method is one of the effective combinations in sentiment analysis. The lexicon-based approach works by utilizing a dictionary of words that have been assigned sentiment values to identify and classify emotions in text. This technique has the advantage of interpretability because it can immediately show words that have positive, negative, or neutral sentiment polarity (Mahajani et al., n.d.). Meanwhile, SVM as a machine learning algorithm is known for its reliable performance in handling text classification with maximum separation margin (Wahba et al., n.d.). By using features extracted from lexicon analysis, SVM is able to build an accurate classification model despite the high dimensionality of the data. The combination of these two methods not only improves sentiment classification accuracy but can also overcome the challenges of complex and diverse text data.

On the other hand, the Support Vector Machine (SVM) method is well suited for sentiment analysis due to its accurate classification and ability to handle a large number of features (Ahmed Khan et al., 2024). Such as research by (Liu, 2020) SVM is proven to be more effective in predicting sentiment than more complex models, such as Gradient Boosting, LSTM, and BERT. The second study by (Nurlaely et al., 2023) with an accuracy of 92%. Based on the research that has been done, the SVM method gets the highest accuracy and is superior to other methods, therefore the authors propose the SVM method with a lexicon-based approach to conduct sentiment analysis on public responses to the relocation of the national capital through comments uploaded on X social media. Sentiment analysis makes it easier to understand the views of the public regarding the relocation of the national capital and public perceptions of the policy.

LITERATURE REVIEW

Moving the capital of the country is a topic that has generated a wide range of reactions and public opinions, especially on X social media. Sentiment analysis is a widely used method to understand public trends and responses to this policy. Several previous studies have examined public sentiment using various analysis methods, including lexicon-based approaches and support vector machines. Such as the research conducted by (Nurcahyawati et al., 2023) This study evaluates the difference between manual and automatic sentiment annotation using Vader Lexicon, and combines Vader with Support Vector Machine (SVM) algorithm to improve classification accuracy. Results show that the automatic use of Vader Lexicon can improve sentiment annotation accuracy from 86% manually to 88.57%. And the integration of Vader and SVM produces the same classification accuracy, which is 88.57%. In addition, research by (Hendrastuty et al., 2021) evaluation using Confusion Matrix shows that the linear SVM kernel performs better than the RBF kernel, with 98.67% accuracy, 98% precision, 99% recall, and 98% F1-score. In contrast, the RBF kernel recorded an accuracy of 98.34%. Based on the analysis results, the majority of public sentiment towards this program is neutral, which is 98.34%. The third research by (H. Setiawan et al., 2021) The results showed that SVM algorithm has higher accuracy (85%) compared to Naive Bayes (81.20%), with precision of 83.60% for SVM and 79.40% for Naive Bayes. Recall of the SVM algorithm reached 84%, superior to Naive Bayes which amounted to 79.60%. This research emphasizes the superiority of SVM in higher result accuracy. The fourth research by (Nurdiyansyah, 2023) analyzes the sentiment towards moving the national capital from DKI Jakarta to Kalimantan using data from the TikTok application and the Long Short Term Memory (LSTM) method. The research was conducted to understand public opinion about whether this policy will bring benefits or not. The fifth research by (A. Setiawan & Suryono, 2024)analyzed public sentiment related to the transfer of IKN using SVM and Naïve Bayes. As a result, SVM excels with 84% accuracy, while Naïve Bayes is only 71%, showing that SVM is more effective in sentiment classification. Last by (Manik et al., 2021) this study shows that the SVM method with a linear kernel is effective in analyzing the sentiment of food e-commerce user reviews. Sayurbox achieved 91.4% accuracy with 70% positive sentiment (738 reviews), while Tanihub had 88.8% accuracy with 65% positive sentiment (348 reviews). This result proves that the majority of reviews on both platforms have a positive sentiment.

Some previous studies have used SVM methods for sentiment analysis, but they still have some limitations. The main difference in this study lies in three aspects. First, this study uses a more optimal percentage of data, so it is expected to increase the accuracy of the model compared to previous studies. Second, the analysis is conducted in two languages, namely Indonesian sentiment and English sentiment, while previous studies have only focused on one language. This allows for a broader understanding of public opinion. Third, this research combines Lexicon and SVM methods, instead of using SVM alone, to improve accuracy in sentiment classification. With these

*name of corresponding author

differences, this research is expected to provide more accurate and comprehensive results in analyzing public sentiment towards the relocation of the capital city.

METHOD

There are numerous stages used in this research to acquire accurate results and conclusions. Each stage of the research was carefully designed to obtain relevant data, apply appropriate methods, and produce well-founded results that can serve as a reference for further analysis. The research process includes various phases, ranging from data collection, data processing, to analysis and evaluation of the results. These stages aim to provide a deep and comprehensive understanding of public sentiment toward the relocation of the National Capital City (IKN). Data for this study was collected over a significant period, from January 2019 to January 2025, to capture evolving public opinions and sentiments surrounding the IKN relocation. The following is the research flow that will be carried out:

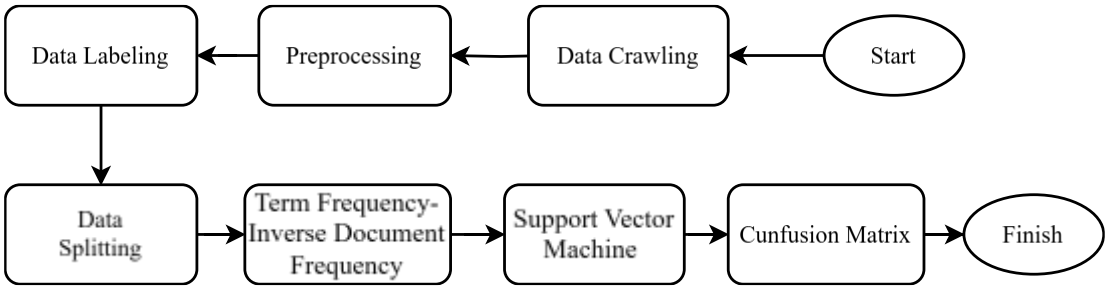

Fig. 1 Research flow

This research uses two languages, namely sentiment in Indonesian and English, by starting the data crawling process from platform X (Twitter). The data collected comes from various writings that talk about the relocation of the National Capital City (IKN), reflecting public opinions from various backgrounds. Following data collection, a text preprocessing stage is performed to increase data quality, which includes text cleaning to remove irrelevant characters, case folding to homogenize letters, stopword removal so that only meaningful words remain, tokenization to break text into word or phrase units, and lemmatization to convert words to their basic form. Next, the data is labeled with sentiment, categorized into two main classes, namely positive and negative sentiment, using a lexicon approach. This approach allows for a more systematic analysis of public opinion. According to research conducted by (Prayoga Siswono et al., 2024) the use of two-class classification in sentiment analysis can improve model accuracy and provide a clearer picture of the public's response to certain issues.

After the sentiment labeling process is finished, the data is separated into two major sections, 50% for training data in the lexicon formation stage and 50% for testing data that will be further processed using the SVM algorithm. This split tries to prevent overfitting and guarantee that the model can learn from a wide range of instances, allowing it to accurately generalize patterns to new data. This approach is particularly important in sentiment analysis, given the high variation in data, both in sentence structure, word choice, and the context of the opinions expressed. Therefore, choosing the right proportion of data sharing is expected to improve the model's performance in classifying sentiment more accurately (N. W. S. Saraswati et al., 2024). 10% of the y-test data was used for manual sampling to assess the adequacy of the model prediction results with manual analysis. This stage acts as extra validation in determining the model's performance and assuring that the method utilized can produce more accurate and consistent outcomes.

After data collection, the text is converted into numerical features by applying the Term Frequency-Inverse Document Frequency (TF-IDF) method to enable more detailed classified. In the classification stage, the Support Vector Machine (SVM) algorithm is used to predict the sentiment towards IKN based on the converted data. The performance of the model is then assessed using confusion matrix, which calculates the number of true positives, true negatives, false positives, and false negatives. The confusion matrix findings are used to calculate evaluation measures such as precision, recall, f1-score, and accuracy, which are used to determine how well the model classifies the signal.

Data Crawling

The initial step to retrieve X data is using Google Colab, After that, the command for crawling data is executed by including the token obtained from X. The collected data was then saved in CSV format. During the crawling process, a total of 8,357 Indonesian tweets and 4,781 English tweets were successfully obtained using keywords

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

that correspond to certain aspects. However, the data obtained is still raw data containing duplication and symbols that are not relevant to the research.

Table 1. Data Crawling Result

	Crawling Data	Data After Cleaning
Indonesian sentiment	8.357	6.339
English sentiment	4.781	2.415

Text Preprocessing

Preprocessing is an important stage in data mining that aims to transform raw data into a more structured and meaningful form. This process ensures the data is ready for further analysis, such as classification or prediction, and improves the accuracy of the model (Fitriyana et al., 2023). Text preprocessing also aims to clean and prepare the text for more accurate analysis. This process includes data cleaning to remove irrelevant elements, case folding to equalize lowercase letters, stopwords removal to remove common words that are not meaningful, tokenization to break the text into small units, and lemmatization to return words to their basic form. With these steps, the data becomes cleaner and ready for further analysis.

Data Labeling

The next step after preparing the data is to label it. At this point, the author uses InSet Lexicon for Indonesian language labeling and Vader Lexicon for English language labelling. InSet Lexicon is an Indonesian sentiment dictionary that includes 3,609 positive and 6,609 negative terms with polarity scores ranging from -5 to +5. This vocabulary is designed to categorize sentiment and better understand public opinion on diverse topics (Musfiroh et al., 2021).

VADER is a lexicon-based sentiment analysis tool used to analyze public opinion on moving the nation's capital. The process generates four sentiment scores, with the compound score as the main determinant: positive (≥ 0.05), negative (≤ -0.05), and neutral (-0.05 to 0.05). In this study, neutral sentiment is categorized as positive to simplify analysis (predictivehacks, 2020).

The labeling findings reveal the dispersion of attitude in Indonesian and English about the moving of the national capital. In Indonesian, public opinion is quite balanced with 51.3% positive and 48.7% negative, showing a small percentage difference. Meanwhile, in English, positive sentiment is much more dominant, reaching 83.3%, compared to 16.7% negative sentiment.

Table 2. Labeling Result Example

Indonesian Full_Text	Sentiment
"presiden indonesia joko widodo bangun training center kota negara ikn nusantara dukung maju persepakbolaan indonesia rencana training center delapan lapang sepak bola"	Positive
"sejarah dunia kota negara republik indonesia pindah jakarta kota nusantara kalimantan timur akibat sebar bau ketek erina"	Negative
"lucu komentar netizen berita gempa ikn nusantara anggap batal proyek lintas lintas api gunung aktif ya gempa jarang gempa lintas terjadi"	Negative
English Full_Text	**Sentiment**
"im happy living archipelago capital city close everywhere move little pasput move little partner move little many new place opening near downside really congested"	Positive
"related sound diri forerunner nagari would later become archipelago namely diri kingdom often said medang kamulang kingdom moved east java starting bank branta river later became diri kingdom"	Positive
"head deputy head indonesia new capital city authority resign mister say full story"	Negative

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data Splitting

After performing word labeling, in this study the data is divided using a 50:50 ratio, where 50% is used for training and 50% for testing. This split aims to prevent overfitting, ensure the model learns from enough examples, and is capable of effectively generalizing patterns to fresh data. This approach is important in sentiment analysis, given the large variation in sentiment data and the need for models to accurately recognize relevant patterns. However, because the accuracy results obtained from using the lexicon are not sufficient, 10% of the test data is taken as manual sampling. This manual sampling aims to improve the quality of the test data with a more accurate classification based on knowledge or manual interpretation. The manually classified data is then used to further evaluate whether the model has made a mistake in predicting sentiment on certain data.

In addition, it should be emphasized that the training data in this study is not manually annotated, but rather uses automated lexicon-based approaches, namely VADER Lexicon and Inset Lexicon. Although this approach is quite efficient, based on previous research, the performance of this lexicon method only reaches around 65% (Murti, n.d.). To ensure that the model continues to perform well, we evaluated the manually annotated test data using confusion matrix. The results of this evaluation showed that even though the training data came from lexicon annotations and was limited in number, the model was still able to provide good classification results, thus proving the effectiveness of the evaluation approach used in this study.

Term frequency inverse document frequency (TF-IDF)

Text analysis is a way to extract information from unstructured text data. Because computers cannot process data in text form, word weighting is a key step in converting text to numerical format. One of the strategies used in this study to express text data in numerical form is Term Frequency-Inverse Document Frequency (TF-IDF) (Harishamzah, 2020). TF-IDF is a weighting method used in text processing to estimate the value of a word in a document in relation to a collection of other documents. The TF-IDF technique multiplies the TF and IDF values, providing a term weight for each word in the document (Septiani & Isabela, 2022). The following are some results of TF-IDF weighting that show how important a word is in a document compared to all other documents. Words with higher TF-IDF values tend to have more specific and relevant meanings in the context of the document.

Table 3. The result of tf-idf weighting

Indonesian sentiment	English sentiment
bangun: 0.1154	good: 0.4105
biaya: 0.2202	capital: 0.0897
investasi: 0.3536	indonesian: 0.2492
maju: 0.1290	government: 0.2438
kawasan: 0.2202	moving: 0.2849
upaya: 0.2748	island: 0.1565
langkah: 0.2748	president: 0.2386
nusantara: 0.1108	kalimantan: 0.2635

RESULT

The classification results based on the initial labeling using the lexicon method on Indonesian data show that the proportion of positive sentiment is 51.3% and negative is 48.7%. After processing using the SVM model, the sentiment distribution shows a slight increase in the positive class to 51.55%, while the negative class becomes 48.45%. Meanwhile, on English data, lexicon labeling resulted in a positive sentiment dominance of 83.3% and negative of 16.7%. The SVM model then strengthens these results with a final prediction of positive by 87.63% and negative by 12.37%. These results show that both the initial labeling and model predictions tend to be consistent, with a higher proportion of positive sentiments especially in the English data, indicating a more optimistic perception of the analyzed topic.

After manual sampling, the model is evaluated to determine the predicted performance of the model. The performance evaluation of the SVM classification model on the test data is performed using some commonly used evaluation metrics, namely accuracy, precision, recall, and F1-score, to assess how well the model classifies the data correctly. In addition to using evaluation metrics, this analysis is also complemented by visualization of Word Cloud and word frequency for each sentiment class (positive and negative). Word Cloud is used to visually display the most prominent words based on their occurrence, while word frequency is presented to show how often certain words appear in each class. This approach not only provides a numerical overview through evaluation metrics, but also reveals commonly used language patterns. Thus, this analysis provides a more comprehensive understanding

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

of the sentiment classification results as well as the context of the analyzed data. The results can be seen in Table 4.

Table 4. Model Testing Results

	Accuracy	Precision	Recall	F1-Score
Indonesian sentiment	0.943218	0.970874	0.943396	0.956938
English sentiment	0.900000	0.926829	0.950000	0.938272

In this research using Visualization in sentiment analysis, such as WordCloud, helps understand data patterns by displaying words according to their frequency. The approach used can be unigram (single word), bigram (two words), or trigram (three words) to provide more context. In addition, the frequency-based word list also supports quantitative analysis, making it easier to understand the key words that influence sentiment.

The following are the results of Indonesian sentiment visualization, namely, unigram, bigram and trigram along with the frequency of words shown in Figure 2 and Table 5.

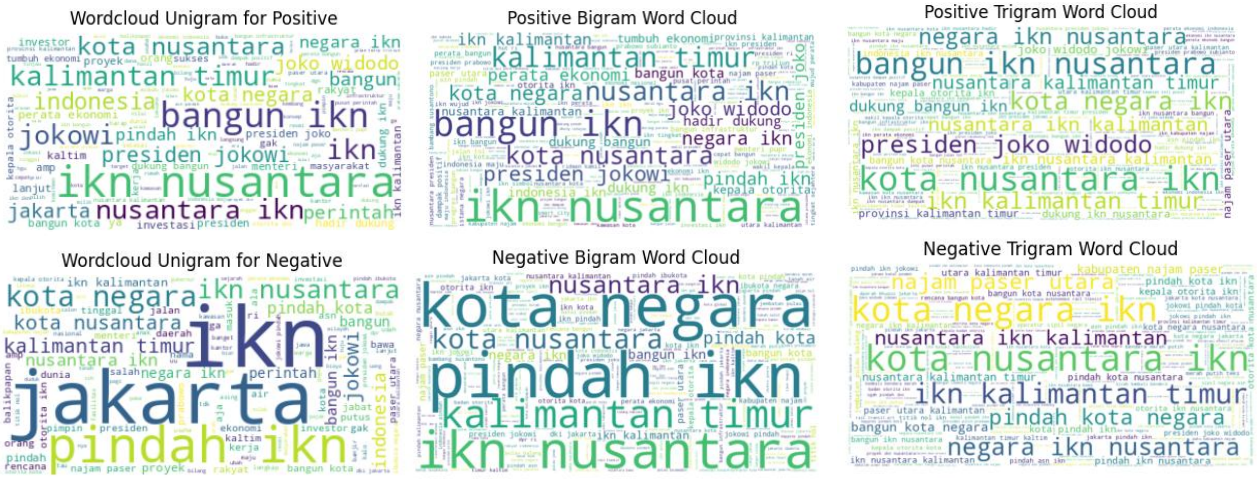

Fig. 2 Indonesian sentiment visualization results

Table 5. Results Frequency of Indonesian sentiment words

Unigram Positive Word Frequency		**Unigram Negative Word Frequency**	
ikn	2293	ikn	953
nusantara	1078	kota	495
bangun	904	pindah	388
kota	502	nusantara	352
jokowi	445	negara	277
Bigram Positive Word Frequency		**Bigram Negative Word Frequency**	
ikn, nusantara	740	pindah, ikn	163
bangun, ikn	445	kota, negara	162
kalimantan, timur	260	ikn, nusantara	155
nusantara, ikn	230	kalimantan, timur	140
kota, nusantara	204	kota, nusantara	132
Trigram Positive Word Frequency		**Trigram Negative Word Frequency**	
kota, nusantara, ikn	176	kota, nusantara, ikn	92

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

bangun, ikn, nusantara	142	kota, negara, ikn	83
kota, negara, ikn	92	ikn, kalimantan, timur	54
presiden, joko, widodo	87	negara, ikn, nusantara	41
ikn, kalimantan, timur	80	pindah, kota, negara	35

The following are the results of English sentiment visualization, namely unigram, bigram, and trigram along with word frequency shown in Figure 3 and Table 6.

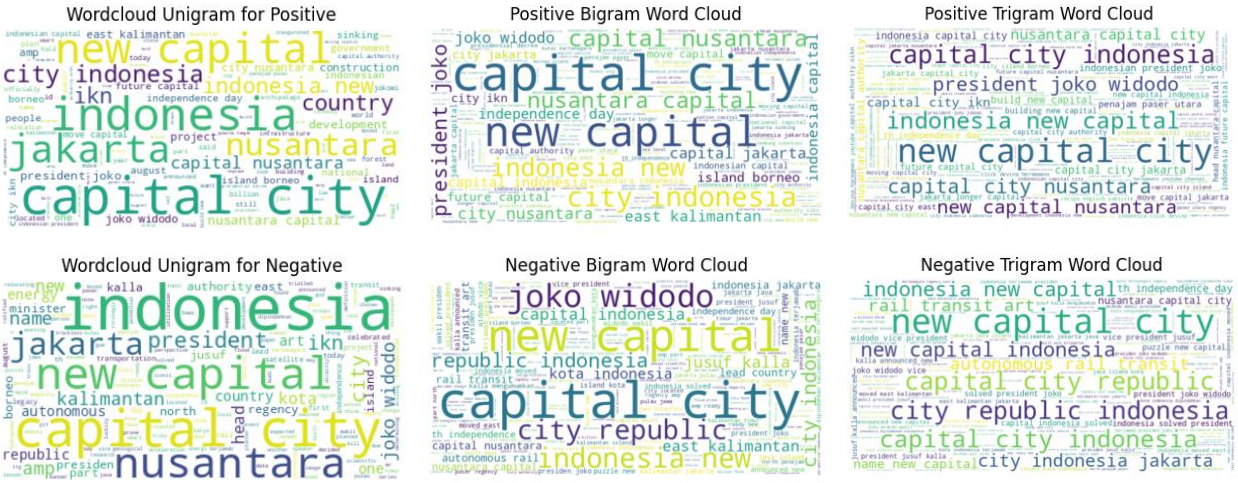

Fig. 3 English sentiment visualization results

Table 6. Results Frequency of English sentiment words

Unigram Positive Word Frequency		**Unigram Negative Word Frequency**	
capital	569	indoensia	20
city	410	capital	20
indonesia	355	city	18
nusantara	256	new	13
new	244	nusantara	9
Bigram Positive Word Frequency		**Bigram Negative Word Frequency**	
capital, city	284	capital, city	13
new, capital	197	new, capital	10
city, indonesia	61	joko, widodo	4
indonesia, new	53	city, indonesia	3
capital, nusantara	49	city, republic	3
Trigram Positive Word Frequency		**Trigram Negative Word Frequency**	
new, capital, city	106	new, capital, city	6
capital, city, indonesia	53	capital, city, indonesia	3
indonesia, new, capital	47	city, republic, indonesia	3
capital, city, nusantara	34	capital, city, republic,	3
president, joko,widodo	33	rail, transit, art	2

DISCUSSIONS

As shown in Table 4, the Support Vector Machine (SVM) model performed very well in the sentiment classification task for both Indonesian and English data. For the Indonesian sentiment data, the model recorded an accuracy value of 94.32%, precision of 97.09%, recall of 94.34%, and F1-score of 95.69%. These values show that the model is not only accurate, but also consistent in detecting positive and negative sentiments equally.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Meanwhile, for English data, the model also showed strong performance with an accuracy of 90%, precision of 92.68%, recall of 95%, and F1-score of 93.83%. These results indicate that the model has good generalization ability to English text, with high precision and recall. Overall, the model's performance in both languages reflects a high level of reliability in detecting sentiment, and supports consistent prediction results as shown in the sentiment distributions described earlier.

The results between the two languages in Table 4 show that the SVM model is slightly more optimal in handling English text. This may be due to the completeness of the linguistic resources or the more consistent language structure in the English data. Nonetheless, the high precision values in both languages indicate that the model's predictions are quite accurate in identifying the correct sentiment.

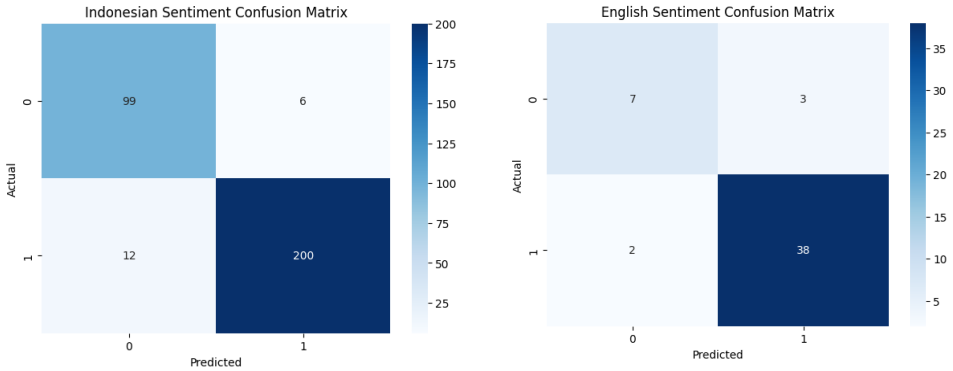

Fig. 4 Confusion Matrix Results

The confusion matrix in Figure 4 provides a detailed overview of the performance of the Support Vector Machine (SVM) model in analyzing Indonesian and English sentiment. This matrix categorizes the classification results into four main categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). For Indonesian data, the model correctly classified 99 negative data (TN) and 200 positive data (TP), and generated 6 false positives (FP) and 12 false negatives (FN). Meanwhile, on the English data, the model showed excellent performance with 38 TP, 7 TN, only 3 FP, and 2 FN. These results show that the model is able to perform accurate sentiment classification in both languages, with excellent positive sentiment detection especially in the English data.

These results are also in line with research from (Wahyudi et al., 2024) who used a similar approach to analyze public sentiment towards the relocation of the National Capital City (IKN). In this study, TF-IDF feature extraction and Support Vector Machine (SVM) method were used to classify public sentiment based on Twitter data. The research showed that the majority of sentiments were negative, reflecting the public's concerns and doubts about the policy. In this study, the approach is extended by involving data in two languages, namely Bahasa Indonesia and English, to capture more diverse public opinions. The use of two languages provides a wealth of linguistic and contextual features, which strengthens the model's ability to understand complex sentiment patterns. For example, the Indonesian and English sentiments shown in Figures 2 and 3, in addition, the word frequency sentiments in Tables 5 and 6 can be seen for positive and negative sentiments based on how much the word frequency is.

Based on the results of WordCloud visualization and Indonesian sentiment frequency analysis, the topic of the National Capital City (IKN) is the dominant topic discussed by the public, with different perspectives between positive and negative sentiments. In positive sentiments, the discussion focuses on development and support for the government. This can be seen from the occurrence of words such as “ikn”, ‘build’, and “jokowi”. Phrases like “ikn nusantara” and “bangun ikn” show enthusiasm for the development project, while phrases like “president joko widodo” emphasize the leadership role in the project. Meanwhile, negative sentiments mostly addressed the issue of evictions and their impact. Words like “move” and “country” reflect people's concerns. Phrases such as “move ikn” and “state city” appeared frequently, indicating a focus on the uncertainty of the relocation process and its consequences. In general, positive sentiments reflect hope for the development of IKN, while negative sentiments reflect concerns about the implementation and impact of moving the capital.

In addition, based on the results of word frequency analysis in English sentiment, the topic of IKN is widely discussed in positive sentiment. Words such as capital, city, indonesia, and nusantara appear frequently, signaling a focus on the construction of the new capital city and hope for the future. In bigrams, phrases such as capital city and new capital city reinforce the optimistic narrative towards the project. Trigrams such as ibu kota baru and ibu kota indonesia also reflect support for national identity and development. In contrast, negative sentiments appear with less intensity. While words such as capital and city still appeared, phrases such as city republic and joko

*name of corresponding author

widodo reflected more concern about the government's policies and direction. Overall, opinions in English towards IKN tended to be positive, with an emphasis on development and hope for the future of the new capital city.

CONCLUSION

This research shows that the Support Vector Machine (SVM) model is able to classify public sentiment towards the relocation of the National Capital with excellent performance. For Indonesian data, the SVM model achieved 94.32% accuracy, 97.09% precision, 94.34% recall, and 95.69% F1-score. As for the English data, the model recorded 90% accuracy, 92.68% precision, 95% recall, and 93.83% F1-score. These results show that the model has very stable performance in both languages, with high precision and detectability. Additional analysis such as WordCloud and n-gram identified "IKN" as the main keyword in the discussion. Positive sentiments were dominated by the themes of development and leadership, while negative sentiments were mostly related to the issues of eviction and project sustainability. These findings can be an important reference for policymakers to respond to public opinion and potential challenges in the process of relocating the National Capital.

REFERENCES

Afifah, D., Aristias, A., Manullang, I. A., Sukma, N. F., & Prasetyo, H. (2024). Kepastian Hukum Terkait Hak-Hak Masyarakat Konservatif Dalam Pembangunan Ibu Kota Nusantara. *Jurnal Penelitian Ilmu-Ilmu Sosial*, *1*(11). https://doi.org/10.5281/zenodo.11658207

Ahmed Khan, T., 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

Fitriyana, V., Hakim, L., Candra Rini Novitasari, D., Hanif Asyhar, A., Studi Matematika, P., Sains Dan Teknologi, F., Sunan Ampel Surabaya, U., & Timur, J. (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. In *Jurnal Buana Informatika* (Vol. 14, Issue 1).

Harishamzah. (2020). *Perbandingan Perhitungan Bobot TF-IDF secara Manual dan Menggunakan Python _ by Harishamzah _ BISA.AI _ Medium*. https://medium.com/bisa-ai/perbandingan-perhitungan-bobot-tf-idf-secara-manual-dan-menggunakan-python-377392a165c6

Hendrastuty, N., Rahman Isnain, A., & Yanti Rahmadhani, A. (2021). *Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine*. *6*(3). http://situs.com

Liu, S. (2020). *Sentiment Analysis of Yelp Reviews: A Comparison of Techniques and Models*. http://arxiv.org/abs/2004.13851

Mahajani, S. J., Srivastava, S., & Smeaton, A. F. (n.d.). *A Comparison of Lexicon-Based and ML-Based Sentiment Analysis: Are There Outlier Words?*

Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). In *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*.

Mega Hendarto, Y. (2022). *Isu Pemindahan Ibu Kota Negara, dari Perdebatan hingga Pengalihan - Kompas.id*. https://www.kompas.id/baca/riset/2022/02/01/isu-pemindahan-ibu-kota-negara-dari-perdebatan-hingga-pengalihan/

Murti, N. R. (n.d.). *Analisis Sentimen Pengguna Aplikasi Female Daily Menggunakan Support Vector Machine: Evaluasi Leksikon Sentistrenghtid dan InSet*.

Musfiroh, D., Khaira, U., Eko, P., Utomo, P., Suratno, T., Studi, P., Informasi, S., Sains, F., & Teknologi, D. (2021). *MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon*. *1*, 24–33.

Nurcahyawati, V., Mustaffa, Z., Raya Kedung Baruk, J., & Persiaran Tun Khalil Yaakob, L. (2023). Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation. *Journal of Information Systems Engineering and Business Intelligence*, *9*(1). https://doi.org/10.20473/jisebi.9.1.109-119

Nurdiyansyah, F. , P. F. U. (2023). *Analisis sentimen perpindahan ibu kota negara pada aplikasi Tiktok menggunakan metode LSTM*.

Nurlaely, R., Sartika Simatupang, D., & Lucia Kharisma, I. (2023). *Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM)*. *4*(2), 376–384. https://doi.org/10.37859/coscitech.v4i2.5161

Prayoga Siswono, A., Fauzi, S., Zalva Surya Hermawan, A., Riyandi, A., Panjaitan No, J. DI, Kidul, P., Purwokerto Selatan, K., & Banyumas, K. (2024). *Analisis Sentimen Pelantikan Presiden Indonesia 2024 Menggunakan Model Klasifikasi dan Algoritma Naive Bayes* (Vol. 4, Issue 1).

predictivehacks. (2020). *How to Run Sentiment Analysis in Python using VADER – Predictive Hacks*. https://predictivehacks.com/how-to-run-sentiment-analysis-in-python-using-vader/

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14622
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Saraswati, M. K., Ariananto, E., Adi, W., Perhubungan, K. S., Sekretariat, U., Ri, K., Hukum, A., Subbidang, S., Iklim, P., Lingkungan, P., & Sekretariat, H. (2022). Pemindahan Ibu Kota Negara Ke Provinsi Kalimantan Timur Berdasarkan Analisis Swot. *Jurnal Ilmu Sosial Dan Pendidikan (JISIP)*, *6*(2), 2598–9944. https://doi.org/10.36312/jisip.v6i1.3086/http

Saraswati, N. W. S., Putra, I. K. G. D., Sudarma, M., & Sukarsa, I. M. (2024). Enhance sentiment analysis in big data tourism using hybrid lexicon and active learning support vector machine. *Bulletin of Electrical Engineering and Informatics*, *13*(5), 3663–3674. https://doi.org/10.11591/eei.v13i5.7807

Septiani, D., & Isabela, I. (2022). *SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks*.

Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 183–192. https://doi.org/10.29408/edumatic.v8i1.25667

Setiawan, H., Utami, E., & Sudarmawan, S. (2021). Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, *5*(1), 43–51. https://doi.org/10.31603/komtika.v5i1.5189

Wahba, Y., Madhavji, N., & Steinbacher, J. (n.d.). *A Comparison of SVM against Pre-trained Language Models (PLMs) for Text Classification Tasks*.

Wahyudi, T., Azmi Verdikha, N., Muhammadiyah Kalimantan Timur, U., Ir Juanda No, J. H., Samarinda Ulu, K., Samarinda, K., & Timur, K. (2024). Klasifikasi Sentimen X-TWITTER Perihal Pemindahan Ibu Kota Indonesia Menggunakan Ekstraksi Fitur TF-IDF Dan Metode Support Vector Machine (SVM). *Jurnal Teknologi Informasi*, *18*(2). https://doi.org/10.47111/JTI

*name of corresponding author