Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 1708~1721
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp1708-1721

Nusantara capital city sentiment analysis using support vector machine and logistic regression

Valencia Eurelia Angelie Tania, Raymond Sunardi Oetama
Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Info

Article history:

Received Jun 28, 2024
Revised Nov 18, 2024
Accepted Nov 24, 2024

Keywords:

Capital city of the nusantara
IKN
Logistic regression
Sentiment analysis
Support vector machine

ABSTRACT

The decision to move position the capital city of Indonesia to East Kalimantan has drawn people's opinions, both pro and con, among the public, especially ahead of the presidential and vice-presidential elections. Discussions relevant to the relocation and construction of the capital city are increasingly crowded on social media, especially Twitter or X. This research aims to determine public sentiment regarding the development of the national capital to help the government and policymakers improve communication strategies, evaluate existing policies, and make more informed decisions based on public feedback. Public sentiment related to developing the Capital city of the Nusantara, including the presidential palace, toll road, and government offices, is analyzed. Support vector machine (SVM) and logistic regression (LR) algorithms are utilized for the sentiment classification. The results reveal that the SVM performs better in classifying sentiments in X data relevant to developing the Capital city of Nusantara, achieving an average accuracy of 91.97%.

Corresponding Author:

Raymond Sunardi Oetama
Department of Information Systems, Faculty of Engineering and Informatics
Universitas Multimedia Nusantara
Tangerang, Indonesia
Email: raymond@umn.ac.id

1. INTRODUCTION

Indonesia's capital relocation from Jakarta to East Kalimantan is a strategic policy announced by President Joko Widodo on April 26, 2019, and formalized through Law Number 3 of 2022 regarding the National Capital [1]. This decision addresses various challenges facing Jakarta, such as overcrowding that causes traffic congestion, air pollution, and a clean water crisis [2]. The island of Java houses 56.1% of Indonesia's total population [3], thus contributing 56.55% to the national gross domestic product (GDP) [4]. This fact shows the inequality of development between regions in Indonesia. The movement of the capital city to East Kalimantan, covering the areas of Penajam Paser Utara and Kutai Kartanegara, is expected to reduce the burden on Jakarta and encourage more equitable economic growth throughout Indonesia [5].

However, the development and relocation of the Capital city of the Nusantara (IKN) led to various opinions in the community, both pros and cons. Some see this as a strategic step to reduce population density in Jakarta and encourage equitable economic growth [6]. However, there are concerns about environmental damage in Kalimantan [7]. Additionally, government estimates suggest that relocating the capital city will cost Rp466 trillion [8], a large budget many consider could be better spent on more urgent matters. The government can consider the pros and cons related to the development of IKN when formulating further policies and overcoming existing concerns. In the run-up to the presidential and vice presidential

elections, discussions about relocating and developing the capital city increased on social media [9]. With 14.8 million active X users in Indonesia [10], Twitter is now X, a medium for people to express their opinions, and it often functions as a data source for public opinion analysis [11].

The algorithms used to analyze sentiment in this research are support vector machine (SVM) and logistic regression (LR). The selection of SVM is due to its effective handling of data-rich environments such as text and its ability to manage different data types through various kernel functions [12]. LR was selected because it can generate probabilities that indicate sentiment categories and coefficients that assess feature importance within the model [13]. Previous research shows that SVM is highly effective in tweet sentiment analysis, consistently achieving accuracies above 80% [14]. Another study demonstrates that SVM outperforms LR in sentiment analysis [15]. However, some prefer LR for text classification, as findings suggest that LR surpasses SVM [13]. Furthermore, research [16] classified public sentiment using LR, Naïve Bayes, support vector classifier, and stochastic gradient descent, revealing that LR produced the highest accuracy of 81%. By using feature extraction, the accuracy of logistic regression can even reach around 87% [17]. Therefore, this study aims to analyze the accuracy of the SVM and LR algorithms in the sentiment analysis of IKN development. public sentiment related to the development of IKN is examined, including the IKN presidential palace, IKN toll road, and IKN government offices, and it also compares the accuracy of SVM and LR algorithms. Several significant insights into public sentiment regarding the development of Ibu Kota Nusantara is provided. The findings on the most frequently mentioned words help identify the main issues discussed by the public, thereby offering valuable input for the government's future communication strategies, evaluate existing policies, and make more informed decisions based on public feedback.

2. METHOD

This research aims to analyze and compare the effectiveness of both SVM and LR to determine positive and negative sentiments related to developing the Capital city of Nusantara (IKN). Figure 1 shows this research workflow applies using the cross-industry standard process for data mining (CRISP-DM) approach.

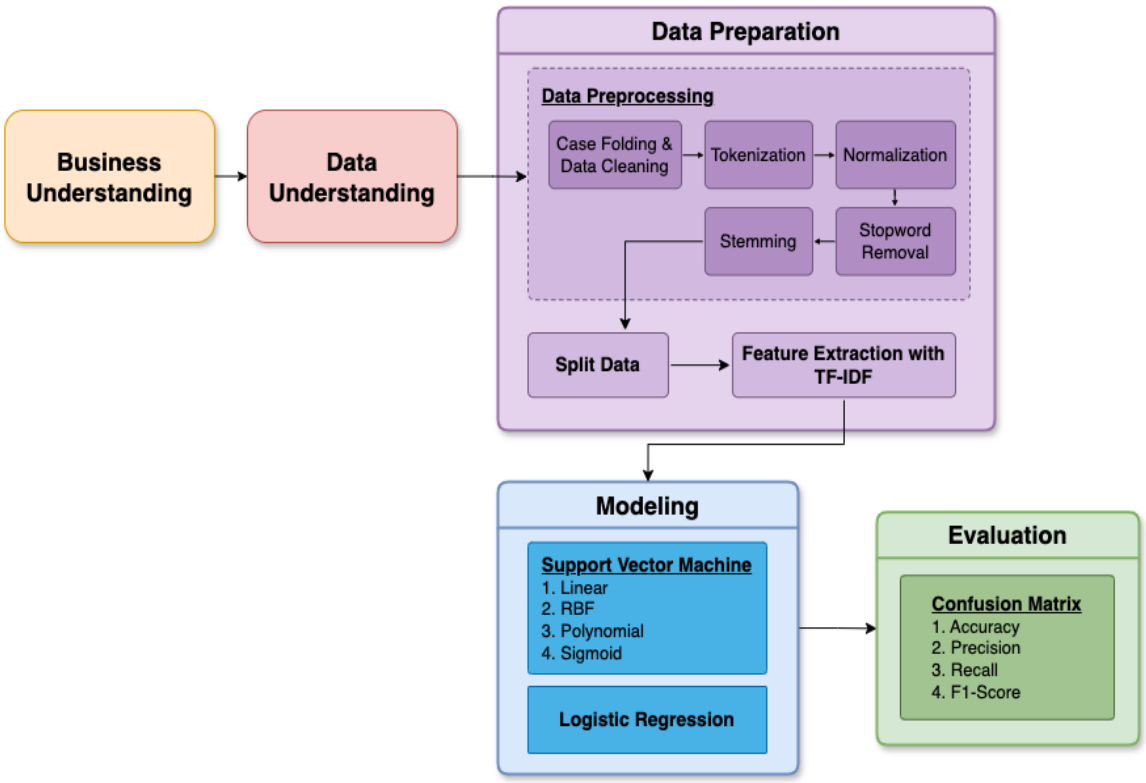

Figure 1. Research workflow

2.1. Business understanding

Business understanding focuses on understanding the business problem and the purpose of the data mining project. This research will analyze public sentiment towards the ongoing development of IKN Nusantara, especially the construction of the presidential palace, toll roads, and government offices. In addition, this research aims to identify influential accounts that form public opinion and spread information related to the construction of IKN.

2.2. Data understanding

Data understanding involves data collection, labeling, and exploration to understand data characteristics and quality. Data collection is performed using NodeXL, while data labeling is done manually. In this research, the dataset used is data from public tweets relevant to the IKN Nusantara development. Data was collected from November 1, 2023, to January 31, 2024, with the keywords "*Pembangunan IKN*" (Development of IKN), "*Istana Kepresidenan IKN*" (IKN presidential palace), "*Jalan Tol IKN*" (IKN toll road), and "*Perkantoran Pemerintah IKN*" *(*IKN government office). Data contains information about interactions between X users, including who initiated the interaction, the target of the interaction, types of interaction, dates and times, and the content of the tweets.

2.3. Data preparation

Every step is necessary to make the final data set for the modeling phases covered in data preparation. At this stage, data preparation includes data preprocessing, splitting into training and testing data, and feature extraction.

2.3.1. Data preprocessing

Data preprocessing involves transforming and selecting data to make it more structured and understandable. The main goal is to optimize text data to be analyzed effectively in the sentiment classification process [18]. Data preprocessing consists of:

a) Case folding: transforms all texts into lowercase.
b) Data cleaning: removes unnecessary elements from the text, such as special or non-alphabetic characters, symbols (!@#$%^&*+_/ and others), emoticons, hyperlinks, and others.
c) Tokenization: breaks sentences into chunks of words called tokens.
d) Normalization: corrects writing and spelling errors in the text and converts abbreviations or slang into standard forms.
e) Stopword removal: removes common words that often appear in text but have no meaning and do not have a significant effect.
f) Stemming: transforms words by converting affixed words into basic words that match the language structure using Sastrawi, a famous Indonesian stemming library.

Table 1 demonstrates the step-by-step preprocessing and its changes in tweet data, showing how the original tweet is transformed. Some rows may contain empty lists after data preprocessing, such as cleaning, tokenization, stopword removal, normalization, and stemming. It is because the words in those rows are deleted during preprocessing. Therefore, rows with empty lists will be removed, and the index will be reorganized to maintain data consistency for further analysis.

2.3.2. Split data

The splitting data stage in sentiment analysis breaks down the dataset into training and testing data. This process is important in building machine learning models, including sentiment analysis. The model learns from training data and then tests its accuracy using testing data to evaluate its performance. The data will be divided into a split to 80% of the data as training data and 20% as testing data.

2.3.3. Feature extraction

Term frequency-inverse document frequency (TF-IDF) is a weighting technique used to find important values in documents, which helps process text more accurately. The weight of each word is computed in a text [19]. The score is calculated based on the frequency at which the word appears in the document (TF) and how important the word is in all documents (IDF) [20]. In this research, the TF-IDF process is performed using the TfidfVectorizer() module of the scikit-learn library in Python to calculate TF and IDF values and convert text into vectors.

2.4. Modeling

The modeling stage involves using techniques to develop models that can accurately predict or classify based on existing data. In this research, the algorithms used are SVM and LR. This research optimizes the SVM algorithm by testing the cost (C) parameter value on linear, RBF, polynomial, and sigmoid kernels.

Table 1. Data preprocessing results

No	Stages	Tweet	Changes
1	Original Tweet	*Kepikiran, lah iya juga ya. tol kaltim dibangun juga pas zaman pak jokowi* 🤔*\nmungkin kalo bkn pak jokowi, nih kaltim bakal gini gini aja gada kemajuan pembangunan y.\nsampe akhirnya ada ikn wuw*	original tweet written by the user.
2	Case folding and data cleaning	***kepikiran*** *lah iya juga ya tol kaltim dibangun juga pas zaman pak jokowi mungkin kalo bkn pak jokowi nih kaltim bakal gini gini aja gada kemajuan pembangunan y sampe akhirnya ada ikn wuw*	*Kepikiran* becomes *kepikiran* (thinking). All punctuations and special characters (e.g., 🤔) are removed.
3	Tokenization	*[kepikiran, lah, iya, juga, ya, tol, kaltim, dibangun, juga, pas, zaman, pak, jokowi, mungkin, kalo, bkn, pak, jokowi, nih, kaltim, bakal, gini, gini, aja, gada, kemajuan, pembangunan, y, sampe, akhirnya, ada, ikn, wuw]*	Text is split into individual tokens.
4	Normalization	*[**terpikirkan**, lah, iya, juga, ya, tol, kaltim, dibangun, juga, **saat**, zaman, pak, jokowi, mungkin, kalo, bkn pak, jokowi, nih, kaltim, bakal, gini, gini, aja, tidak, kemajuan, pembangunan, y, sampe, akhirnya, ada, ikn, wuw]*	*kepikiran* (thinking). becomes *terpikirkan* (thought of), *pas* (just in time) become saat (when).
5	Stopword removal	*[terpikirkan, tol, kaltim, dibangun, zaman, jokowi, jokowi, kaltim, tidak, kemajuan, pembangunan, ikn]*	*Removing lah* (expression), *iya* (yes), *juga* (also), *ya* (yes), *saat* (when), *pak* (salutation), *mungkin* (maybe), *kalo* (if), *bkn* (undetectable word), *nih* (here), *bakal* (will), *gini* (like this), *aja* (just), *y* (undetactable word), *sampe (untill), akhirnya (finally), ada* (there is), wuw (expression).
6	Stemming	*[**pikir**, tol, kaltim, **bangun**, zaman, jokowi, jokowi, kaltim, tidak, **maju**, **bangun**, ikn]*	*terpikirkan* (thought of) becomes *pikir* (think), *dibangun* (built) becomes *bangun* (build), *kemajuan* (progress) becomes *maju* (progress), pembangunan (construction) becomes *bangun* (build).

2.4.1. Support vector machine

SVM is a part of the machine learning algorithms that use a supervised learning category for classifying and regressing tasks. It seeks the best hyperplane by maximizing the distance between data groups (margin). The hyperplane function separates different groups of data. The margin is applied to separate the data groups; in this case, positive data (+1) from negative data (-1) makes it more likely that the algorithm can classify the data into the correct groups more accurately [21]. A SVM ‘s main advantage is its ability to analyze various data types. It is facilitated by kernel functions, which allow SVM to operate in higher dimensional spaces without explicit calculations. This feature enables SVM to manage non-linear data effectively. SVM kernel functions include linear, polynomial, gaussian radial basis function (RBF), and sigmoid with the following functional equations [22].

a) Linear Kernel

$$K(x_i, z) = x_i^T z \quad (1)$$

b) RBF Kernel

$$K(x_i, z) = exp[-\gamma \|x_i - z\|^2], \gamma > 0 \quad (2)$$

c) Polynomial Kernel

$$K(x_i, z) = (1 + x_i^T z)^d \quad (3)$$

d) Sigmoid Kernel

$$K(x_i, z) = tanh\ (\gamma x_i^T z + r) \quad (4)$$

Where:
x_i : Feature vector from the training dataset.
x : Feature vector from the testing dataset.
γ : Scalar factor to adjust sensitivity model to differences in feature space.
p : Degree of the polynomial.
r : Constant term.

2.4.2. Logistic regression

LR applies supervised learning to classifying a dependent variable's probability based on independent variables [23]. LR, developed from linear regression, is used for binary classification, where the dependent variable is discrete (e.g., 0 or 1, yes or no). LR uses sigmoid or logistic functions to convert linear inputs into values between 0 and 1, ensuring the prediction results are always within this range [24]. In sentiment analysis, LR classifies texts, such as tweets or reviews, into positive or negative sentiment categories by relating text features to sentiment categories [25].

2.5. Evaluation

Evaluation is the stage of model performance assessment. At this stage, the model's sentiment classification is assessed to determine the level of accuracy produced by each model. The algorithm's performance is evaluated using the confusion matrix [26]. It includes four key parameters: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). True positive refers to the number of positive data accurately classified. False positive refers to the number of negative data incorrectly classified as positive. Meanwhile, true negative refers to the number of negative data correctly classified, while false negative is positive data incorrectly predicted as negative. Accuracy, precision, recall, and the F1-score are calculated from these values, which help examine the classification model's effectiveness [26].

3. RESULTS AND DISCUSSION

The overall findings of this study indicate that public sentiment towards the IKN is generally positive. Broad public support for the development of the IKN is reflected in their confidence in the project's progress and commitment to its successful completion. The public also appreciates the architectural design of the IKN presidential palace, especially the integration of green spaces and Indonesian cultural elements. Public trust in the project's progress and government leadership is evident, especially in achieving key development goals. The public is optimistic that the IKN toll road will contribute to overall infrastructure improvements and achieve national development targets effectively. Positive sentiment towards the IKN government building reflects public support for its architectural quality, design, and vision. Another major study finding is that the SVM algorithm consistently performs better than other models like LR, with accuracy rates as high as 100% for the IKN presidential palace dataset and an overall average of 91.97%.

The distribution of data after preprocessing and the sentiment categories manually labeled are shown in Table 2. Generally, they exhibit more positive sentiments than negative sentiments for all datasets. The development of IKN has the highest number of tweets, while only a small amount of data contains discussions about IKN president palace.

Table 2. Number of sentiment categories

Dataset	Positive	Negative	Total Tweets
Development of IKN	15,214 (53%)	13,364 (47%)	28,578
IKN presidential palace	34 (62%)	21 (38%)	55
IKN toll road	577 (61%)	375 (39%)	952
IKN government offices	476 (55%)	391 (45%)	867
Total	16,301 (54%)	14,151 (46%)	30,452

3.1. Development of IKN

Figure 2 shows the words frequently appearing in tweets about the development of IKN in both positive and negative sentiments. The words '*ikn*' and '*bangun*' dominate both sentiments, indicating that these words are the main topics in discussions related to IKN development. However, their frequency is

higher in the positive context, which could indicate a more robust response or more support for the idea of IKN development. In positive sentiment, the words '*dukung*' (support) and '*komitmen*' (commitment) appear, showing the public's confidence and encouragement for the continuation and success of IKN development. Meanwhile, words like '*rakyat*' (people) and '*paslon*' (presidential candidates) appear in a negative context.

Table 3 details an analysis outlining effective strategies to utilize this sentiment data to address the public's stance on IKN development. Public responses highlighted several issues related to funding IKN projects. The government must clarify the allocation of these funds and use them to respond to public concerns. It is important to emphasize that the development of IKN involves not only the state budget (APBN) but also other funding sources such as public-private partnerships (PPP) and private investment [7]. The government must also show how the development of IKN will benefit the national economy, not just limited to the IKN area. Furthermore, the positive sentiments reflected in keywords such as 'support,' 'complete,' and 'commitment' suggest a robust public backing for the IKN initiative. To sustain and possibly increase this support, the government must enhance its communication strategies, focus diligently on the timely completion of projects, and reaffirm its commitment to meeting the goals of the IKN development. Such proactive measures are essential to aligning project outcomes with public expectations and ensuring the project's success in fostering economic growth.

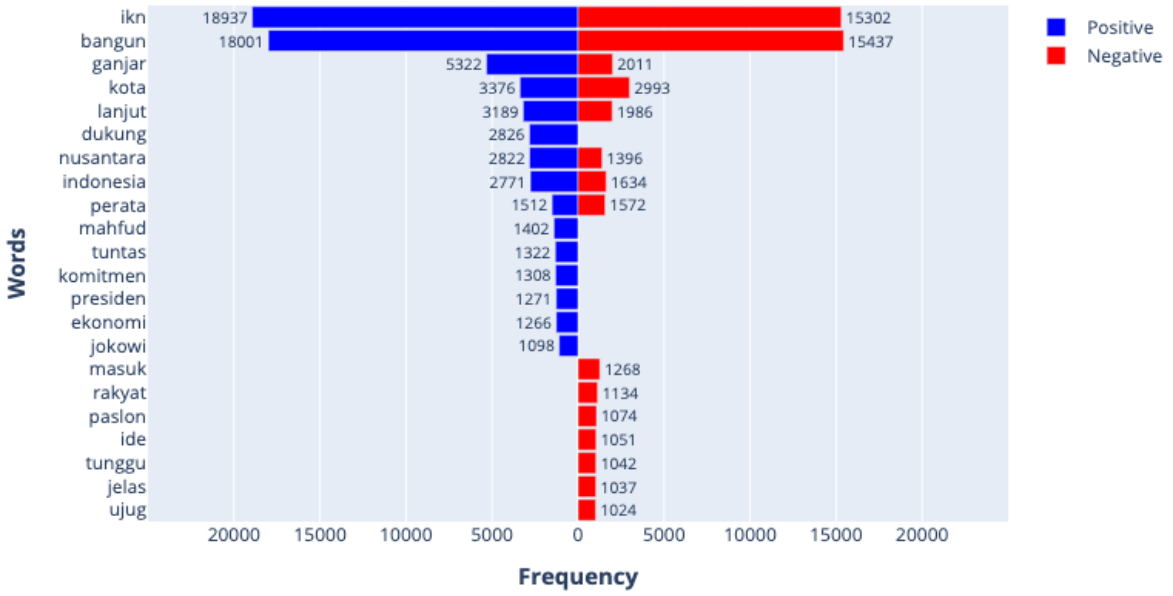

Figure 2. 15 most frequent words based on sentiment in the IKN development discussion

Table 3. Discussion on IKN development based on public response

Keyword	Tweet representation	Interpretation/Suggestion
Enter	Where is the logic in equitable development if only the IKN is being built, and significant national budget funds enter into IKN exclusively?	Clarify funding sources and distribution plans.
Citizen	So, what you're saying about the IKN development already being a certain percentage complete—where is the money coming from? Is it state funds? Does this mean you are quietly burdening the citizens again with your ambition to build the IKN? How can someone like you exist?	
Candidates	Many financial experts have rejected using the national budget for IKN development, and some candidates oppose it. It suggests indications of unpreparedness.	
Support	Please continue to support the IKN to ensure more balanced development and a broader vision for Indonesia while AMIN follows its path... Let those who are good for the Indonesian people continue their struggle.	Enhance communication, ensure project completion, and reaffirm commitment to IKN goals.
Complete	The IKN is a legislative product that must be completed despite the controversies and debates. Like it or not, the development of the IKN must continue. Ganjar is committed to continuing the IKN as he is a person who adheres to the law (constitution)	
Commitment	Ganjar's commitment to continue the IKN has been there from the start and has never shifted. According to him, the continuation of IKN development has become a duty that must be fulfilled.	

3.2. IKN presidential palace

Figure 3 shows the words frequently appearing in tweets about the IKN presidential palace in both positive and negative sentiments. Words such as "*presiden*" (President), "*ikn*," "*istana*" (palace), "*bangun*" (build), and "*jokowi*" often appear in both sentiment categories, but their frequency is higher in the positive context. In the positive sentiment category, words like "*ridwan*" and "*kamil*" appear, referring to public figures associated with positive perceptions in the ongoing discussions. Conversely, in the graph depicting negative sentiment, the word "*bogor*" appears, referring to a specific location, and "*anies*" refers to another public figure.

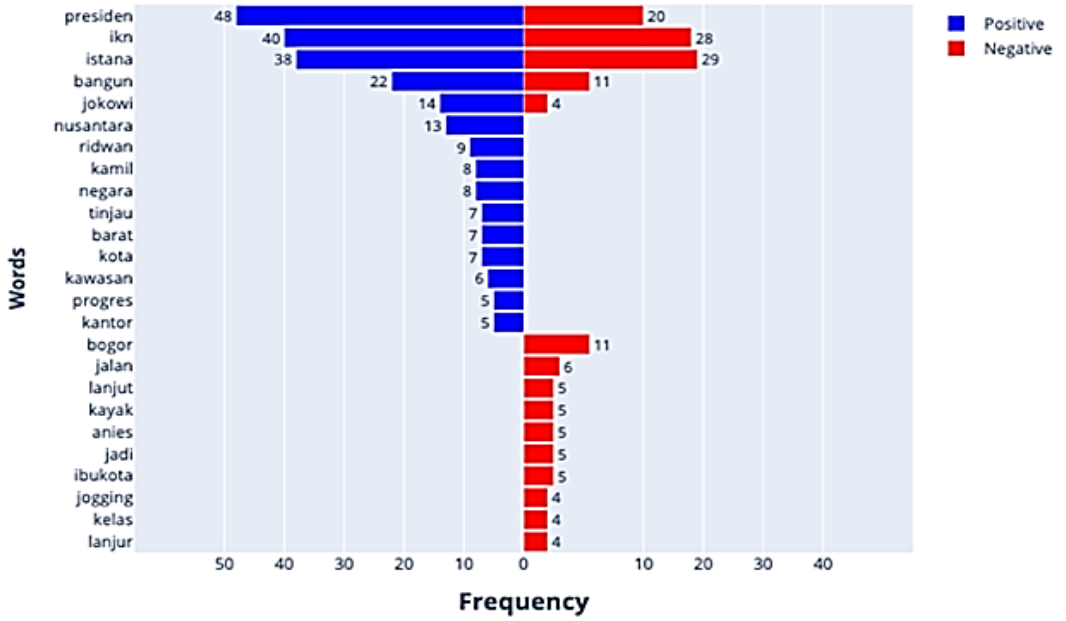

Figure 3. The 15 most frequent words based on sentiment in the IKN presidential palace discussion

Table 4 outlines development suggestions that include adopting the Bogor Palace's architectural and functional aspects, known for its expansive parks and gardens that create a beautiful and comfortable atmosphere [27]. The recommendations also propose that the IKN presidential palace incorporate green spaces for social and sporting activities, blending modern architectural designs with elements of Indonesian culture to replicate the ambiance of Bogor Palace. Additionally, Anies Baswedan's comments highlight the critical need for a budget that supports infrastructure development and other essential public services [28]. Positive reactions to the palace's development, marked by "review" and "*progress*," indicate thorough oversight and steady advancement. The word "*ridwan*" frequently relates to his role as a curator, emphasizing his contribution to ensuring that the palace's design is modern, functional, and culturally representative of Indonesia's diversity [29].

Table 4. Discussion on IKN presidential palace based on public response

Keyword	Tweet representation	Interpretation/Suggestion
Bogor	The construction of the presidential palace at IKN will be completed, and it will function like the Bogor Palace or the Tampak Siring Palace.	Use Bogor Palace as a spatial and architectural reference.
Anies	Presidential candidate number 1, Anies Baswedan, stated that there are other more pressing needs than the construction of the presidential palace at IKN Nusantara. One of these is affordable fertilizer for farmers.	Balance infrastructure development with other critical needs.
Jogging	JIS is enjoyed by all segments of society, suitable for jogging. Will IKN offer jogging opportunities at the presidential palace office?	Ensure accessible open spaces for all communities.
Review	President Jokowi, accompanied by several other officials, continues to review the progress of the presidential palace complex construction at IKN to witness the results that will bring positive changes to Indonesia.	Enhance oversight and integrate cultural values in designs.
Progress	New Year 2024, are you ready to move to IKN yet? The construction progress of the national palace building and the ceremonial field at the presidential palace area in the national capital is already 42.01%.	
Ridwan	The presidential palace at IKN will showcase the cultural diversity of the archipelago, says Ridwan, the IKN curator.	

3.3. IKN toll road

Figure 4 shows the words frequently appearing in tweets about the IKN Toll Road in both positive and negative sentiments. Words such as "*ikn*," "*jalan*" (road), "*tol*" (toll), and "*bangun*" (build) often appear in both sentiment categories, indicating that these words are the main topics in discussions related to the IKN toll road. However, their frequency is higher in the positive context. In the positive sentiment category, words like "*progress*" (progress), "*lanjut*" (continue), "*target*" (target), and "*jokowi*" (referring to the President) appear, highlighting the public's focus on progress and the President's influence on the development. Conversely, in the graph depicting negative sentiment, words like "*perintah*" (order), "*anies*" (referring to another public figure), "*cepat*" (fast), and "*rakyat*" (people) appear.

Discussion on IKN toll road is shown in Table 5. The analysis indicates public concerns about the allocation of funds and the overall benefits of the IKN development. Therefore, it is also essential for the government to balance infrastructure development, focusing on IKN and other areas that need similar development. The fulfillment of urgent community needs such as education, health, and other facilities must go hand in hand. This approach ensures that development supports sustainable growth and meets the community's needs. On the other hand, positive sentiments towards the development of the IKN toll road include the words "continue", "*target*", and "*jokowi*", which shows public support and trust in the President's progress and involvement, adding to public optimism for the achievement of set development targets. To maintain this support, the government needs to improve transparent communication on development progress and ensure that infrastructure policies and projects, such as the IKN toll road, align with people's needs and aspirations and are executed with efficiency and quality.

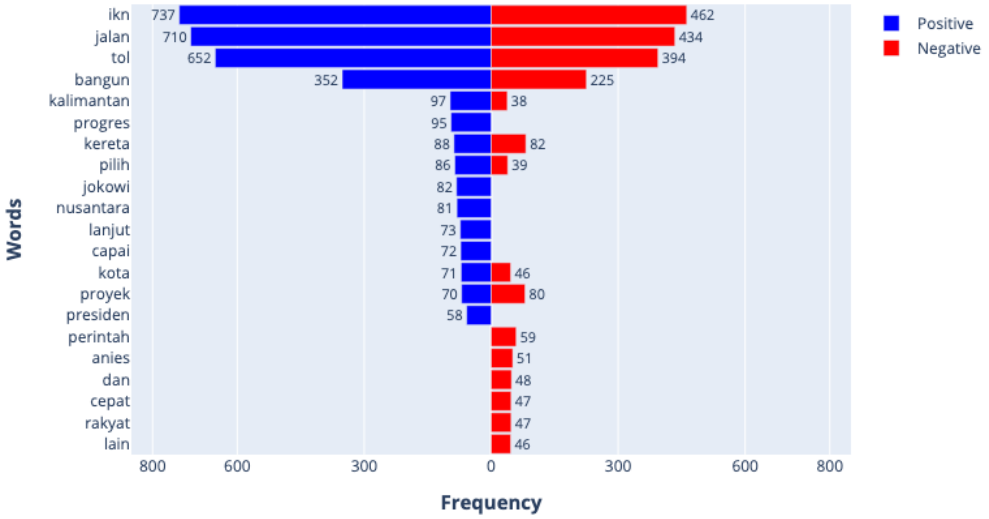

Figure 4. The 15 most common words from tweets about the IKN toll road

Table 5. Discussion on IKN toll road based on public response

Keyword	Tweet representation	Interpretation/Suggestion
Government	By 2024, village roads like this will still exist in one of the villages in West Sumatra, while the central government heavily borrows for high-speed trains, tolls, and IKN, which only a handful of rich people and ministry officials will enjoy.	Expand infrastructure initiatives to include other underdeveloped regions.
Anies	@tvOneNews Mr. Anies appreciates the toll road and train projects by Mr. Jokowi but disagrees with the IKN project as many cities are still unbalanced, especially with the IKN budget being 460T. It would be better to allocate those funds for upgrading contract teachers, building clinics, and raising police salaries	Address critical needs in education, healthcare, and public welfare.
Citizen	Enough, Mr. President, stop pretending to care. Unemployment is rampant, high school graduates cannot find jobs, and many Indonesian children drop out of school. Focus on building IKN and constructing toll roads for the rich without caring for your people.	
Continue	Many people still want President Jokowi's programs to continue. Not everyone wants to start from scratch again. Many still want the IKN to continue, along with the development of toll roads and other facilities. Some choices lean towards change, while others prefer continuity.	Transparent communication and ensuring that policies and infrastructure projects align with the public's needs and aspirations.
Target	Hopefully, the progress of the toll road can quickly meet the set targets to accelerate the development of IKN Nusantara.	
Jokowi	Indonesia's President, Joko Widodo, stated that the progress of the toll road construction to support access to the National Capital, IKN Nusantara, has reached 55%.	

3.4. IKN government office

Figure 5 shows the frequency of words frequently appearing in tweets about IKN government offices in positive and negative sentiments. "*ikn*," "*gedung*" (building), and "*bangun*" (build) is the most commonly mentioned, predominantly in a positive context. Meanwhile, words like "*desain*" (design) and "*kantor*" (office) also have high frequencies in both sentiments, indicating discussions about the physical aspects of the IKN offices. In the positive sentiment category, words such as "*menang*" (win), "*juara*" (champion), "*kawasan*" (area), and "*ikut*" (participate) appear. Conversely, in the graph depicting negative sentiment, words like "*dpr*" (house of representatives), "*investor*," "*pindah*" (move), and "*lembaga*" (institution) are present.

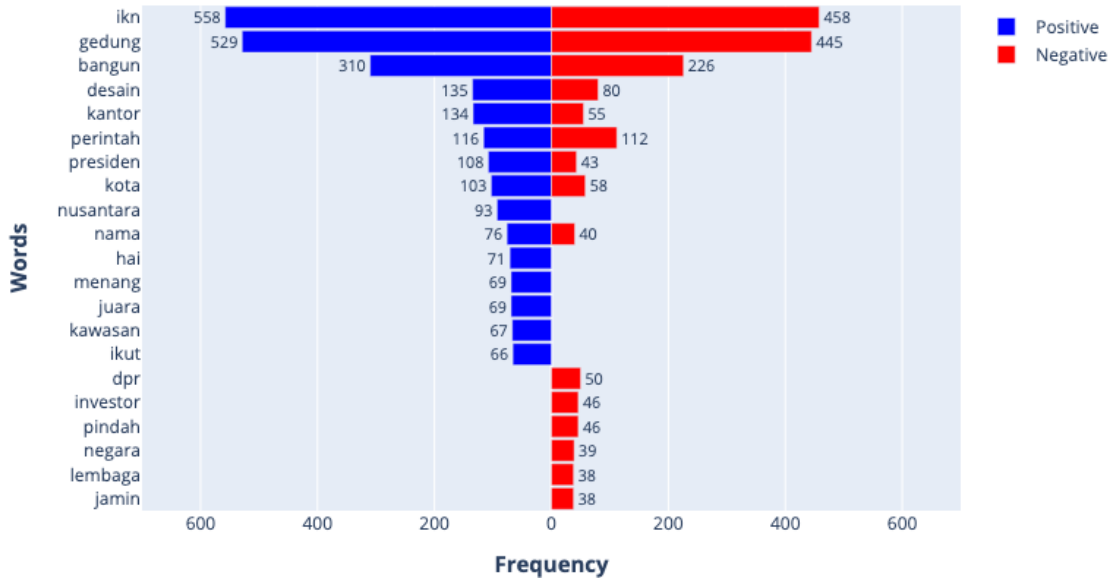

Figure 5. The 15 most frequent words based on sentiment in the IKN government office discussion

Based on Table 6, public responses highlighted concerns and questions about using abandoned government buildings and assets. Abandoned government buildings can be utilized for commercial and public purposes. As mentioned by the director of state assets policy formulation of the ministry of finance, buildings can be leased out to generate income and maintain the utility of the building [30], as was done for the dhanapala building in the ministry of finance complex. Additionally, there are concerns regarding the involvement of investors in funding IKN infrastructure. For this reason, the government needs to improve information transparency about investors, such as publicizing the investors involved and their contributions to build public trust. On the other hand, keywords such as "building", "build", and "design" appear dominant in positive sentiments, indicating public support for the quality and design of the architecture.

Table 6. Discussion on IKN government office based on public response

Keyword	Tweet representation	Interpretation/Suggestion
Investor	Investors are meant to build government buildings. The state budget only covers 20%. Mention investors who participate in putting money in government buildings. The private sector building commercial facilities should not be called IKN Investors.	Inform all investors involved in the IKN infrastructure funding and how funds are allocated between government and commercial projects.
DPR	@Sophistssss @vadalle @elisa_jkt Once IKN is completed, what exactly will the DPR be used for? Is there a plan for the use of government buildings?	Clarify plans for using government buildings that are no longer in use.
Move	If the move to IKN happens, what will become of the government buildings in Jakarta? Turned into dormitories?	
Building	The construction of government buildings in IKN is strengthening the business world's trust.	Continue development with a focus on innovative, energy-efficient, and sustainable development.
Build	The point is to continue development that is not only focused on buildings but on green energy.	
Design	The president's office building was designed by artist Nyoman Nuarta, who also designed the bridge on the IKN toll road.	

3.5. Algorithm comparison results

The key experiment here is the performance of various SVM kernels (linear, RBF, polynomial, and sigmoid), which will be evaluated with different C parameter values (0.5, 1, 2, 3, and 4) on four datasets, including development of IKN, IKN presidential palace, IKN toll road, and IKN government offices. This experiment aims to identify the combination of kernels and C values that provide the highest accuracy on each dataset. By systematically testing each kernel and C value, this experiment aims to determine which configuration best suits the characteristics of each dataset. Accuracy will be used as the primary metric to measure and compare the success of each combination so that the model can produce the most reliable results for each dataset.

The SVM algorithm will first test the cost parameter (C) value on the linear, RBF, polynomial, and sigmoid kernels, with the tested C values being 0.5, 1, 2, 3, and 4. The test results with the highest accuracy will be used for SVM modeling; kernel and parameter values for each dataset can be seen in Table 7. The "Development of IKN" dataset shows the highest accuracy of 0.897 with the RBF kernel at C=4.0. For the "IKN presidential palace" dataset, linear and polynomial kernels achieve perfect accuracies of 1,000 for all tested C values. However, the linear kernel is selected with a C=0.5. The "IKN toll road" dataset reports the highest accuracy of 0.937 using the RBF kernel with a C=2.0. Meanwhile, the "IKN government offices" dataset finds the linear kernel with a C=1.0 to provide the highest accuracy at 0.845. The next phase of this research involves evaluating the results, which aims to assess and compare the quality of each algorithm on each dataset. The evaluation will use a confusion matrix, including accuracy, recall, precision, and F1-score values.

Table 7. Results of SVM parameters

Dataset	Kernel	Accuracy of C values				
		0.5	1.0	2.0	3.0	4.0
Development of IKN	Linear	0.840	0.850	0.857	0.862	0.861
	Polynomial	0.847	0.873	0.874	0.874	0.873
	RBF	0.869	0.891	0.894	0.896	**0.897**
	Sigmoid	0.808	0.796	0.773	0.770	0.764
IKN presidential palace	**Linear**	**1.000**	1.000	1.000	1.000	1.000
	Polynomial	1.000	1.000	1.000	1.000	1.000
	RBF	0.909	1.000	1.000	1.000	1.000
	Sigmoid	0.909	1.000	1.000	1.000	1.000
IKN toll road	Linear	0.890	0.890	0.901	0.895	0.901
	Polynomial	0.812	0.822	0.822	0.822	0.822
	RBF	0.838	**0.937**	0.927	0.916	0.911
	Sigmoid	0.843	0.880	0.859	0.869	0.874
IKN government offices	**Linear**	0.810	**0.845**	0.810	0.776	0.753
	Polynomial	0.718	0.695	0.713	0.718	0.718
	RBF	0.776	0.810	0.805	0.805	0.805
	Sigmoid	0.770	0.799	0.828	0.805	0.764

Based on Table 8, the SVM algorithm has higher accuracy than LR on the various datasets tested. On the IKN development dataset, SVM achieved an accuracy of 89.68%, while LR recorded 83.89%. For the IKN presidential palace dataset, both algorithms attained an accuracy of 100%. On the IKN toll road dataset, the SVM accuracy was 93.72% while LR was 89.53%, and on the IKN government office dataset, SVM achieved 84.48% while LR was 81.03%. The average accuracy for SVM is 91.97%, while LR has 88.61%.

Table 8. Comparison of SVM and LR algorithm

Dataset	Algorithm	Accuracy	Precision	Recall	F1-score
Development of IKN	SVM	89.68%	89.74%	89.68%	89.69%
	LR	83.89%	84.29%	83.89%	83.91%
IKN presidential palace	SVM	100.00%	100.00%	100.00%	100.00%
	LR	100.00%	100.00%	100.00%	100.00%
IKN toll road	SVM	93.72%	93.74%	93.72%	93.69%
	LR	89.53%	90.15%	89.53%	89.31%
IKN government offices	SVM	84.48%	84.77%	84.48%	84.39%
	LR	81.03%	81.59%	81.03%	80.84%
Average	**SVM**	**91.97%**	**92.06%**	**91.97%**	**91.94%**
	LR	88.61%	89.01%	88.61%	88.52%

3.6. Comparison sentiments on previous research

A comparison of the results of this study with previous studies can be seen in Table 9. In analyzing sentiment related to the IKN, various algorithms have been used with varying results. The SVM algorithm consistently shows high accuracy in this study and other studies, with an accuracy above 80%. Naïve Bayes and RF are popular algorithms, but their performance is generally below SVM. Several studies have shown that the accuracy of these two algorithms often does not reach 80%. LR, convolution neural network (CNN), and k-nearest neighbor (KNN) are less popular. However, in this study, LR produced an accuracy above 80%. In terms of sentiment composition, this study shows positive sentiment towards the IKN. This finding aligns with previous studies showing that positive sentiment still dominates perceptions about the IKN. This trend indicates that public views of the IKN are generally more positive than negative.

Table 9. Comparison of sentiments on IKN results

Articles	Algorithm	Accuracy	Positives	Negatives	Neutral
This study	SVM	84.48% to 100%	58%	42%	-
	LR	81.03% to 100%			
Darmawan [31]	SVM	95.24%	85%	15%	-
	RF	86.90%			
Hasugian *et al.* [32]	Naïve Bayes	77,05%	66%	34%	-
Wibowo *et al.* [14]	SVM	89.77%	38.67%	12.94%	48.39%
	Naïve Bayes	79.26%			
	RF	66.88%			
Hasanati *et al.* [33]	SVM	90.65%	34.7%	17.3%	48%
	CNN	84,92%			
Arista *et al.* [34]	Naïve Bayes	94%	80%	20%	-
Syabri *et al.* [35]	Naïve Bayes	72.31%	60.6%	23.9%	15.5%
Huda and Yel [36]	KNN	83.75%	87%	13%	-
Utama *et al.* [37]	CNN	90%	36%	34%	30%

4. CONCLUSION

This study analyzes public sentiment on Twitter regarding the development of the Nusantara Capital city using SVM and LR. This study found that the SVM algorithm consistently outperformed LR across all datasets. It shows that SVM is more effective in sentiment analysis, with higher accuracy in predicting sentiment categories. For each dataset, a different SVM kernel provides the best results. The development of IKN dataset shows the highest accuracy of 89.68% with the RBF kernel and C=4.0. The IKN presidential palace dataset achieved perfect accuracy (100%) with linear and polynomial kernels at all C values, with the linear kernel chosen at C=0.5. The IKN toll road dataset achieved 93.72% accuracy using the RBF kernel and C=2.0, while the IKN government office dataset showed the highest accuracy of 84.48% with the linear kernel and C=1.0.

This study identifies several significant insights into public sentiment regarding the development of IKN. The findings on the most frequently mentioned words help identify the main issues the public discusses, thereby offering valuable input for the government's future communication strategies, evaluating existing policies, and making more informed decisions based on public feedback. Overall, public perception of the IKN tends to be positive, with a positive to-negative sentiment ratio of 54%:46%. Despite some concerns, most public opinion supports the IKN development project. Furthermore, key findings related to sentiment analysis concerns for each dataset are as follows. First, the government needs to clarify the sources and allocation of funds, including the state budget, public-private partnerships, and private investments. Public concerns emphasize the importance of demonstrating how the IKN development will benefit the broader national economy, not just the IKN area. Second is the need for a balanced budget that supports not only the construction of the palace but also other essential public services and infrastructure development. Third, ensuring that resources are allocated fairly between IKN and other regions needing similar infrastructure development. Public sentiment emphasizes the need for the government to balance toll road funding with investments in essential services such as education and healthcare, ensuring that the development benefits the broader population and not just select areas. Fourth, the involvement of private investors in financing the infrastructure. Public responses emphasize the need for greater transparency regarding investor contributions and their roles in the project to build public trust. Additionally, there are concerns about the future use of abandoned government buildings, suggesting that these buildings could be repurposed for commercial or public uses to generate income and maintain utility.

Continuing this sentiment analysis will be very useful for monitoring public perceptions of IKN amidst political changes and helping the new government adjust communication strategies and policies related to this project. Following are several future research possibilities. First, post-Jokowi sentiment

changes. Research could focus on how public sentiment towards IKN changed after Jokowi was no longer in office. Will public support for the project remain strong, decline, or increase under the new administration? This analysis is essential for understanding the dynamics of public opinion influenced by new leadership. Second, the sentiment of governments is compared. Researchers can compare public sentiment towards IKN during and after the Jokowi administration. It can provide insight into how policy changes, government communications, and new approaches to IKN projects influence public opinion. Third, the influence of new government policies. Research can also explore how new policies related to IKN affect public sentiment. If there is a significant change in policy direction, this research can help identify whether the change is welcomed positively or negatively by society.

ACKNOWLEDGMENTS

The authors would like to thank multimedia Nusantara University for its full support, facilities, and conducive academic environment during the research process and preparation of this article.

FUNDING INFORMATION

This research was supported by Multimedia Nusantara University through the Independent Research program. Multimedia Nusantara University did not play a role in the design of the research, data collection, data analysis, interpretation of results, or writing of the manuscript.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Valencia Eurelia Angelie Tania	✓	✓	✓		✓	✓		✓	✓		✓		✓	
Raymond Sunardi Oetama	✓	✓		✓			✓			✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest, either financial or non-financial, that influenced the course of the research or the preparation of this manuscript. All findings and conclusions presented are the result of objective analysis and independent thinking.

DATA AVAILABILITY

Data supporting the findings of this study can be obtained by contacting the corresponding author [RSO].

REFERENCES

[1] I. Perwira, S. D. Harijanti, M. Susanto, and M. Y. Adhihernawan, "Capital city relocation in Indonesia: compromise failure and potential dysfunction," *Cogent Social Sciences*, vol. 10, no. 1, p. 2345930, 2024, doi: 10.1080/23311886.2024.2345930.
[2] The Urban Task Force, "Managing the urban environment," *Towards an Urban Renaissance*, vol. 8, no. 01, pp. 90–104, 2020, doi: 10.4324/9780203497746-13.
[3] P. Rahayu, E. F. Rini, I. Andini, and R. A. Putri, "Development and urbanisation during the COVID-19 pandemic: regional vulnerability in Java, Indonesia," *Town Planning Review*, vol. 94, no. 4, pp. 411–433, 2023, doi: 10.3828/tpr.2022.21.
[4] V. A. Dihni, "Indonesian economy still centered in Java in Q2 2022: [Translation] (in Indonesian): [Ekonomi RI masih terpusat di Jawa pada kuartal II 2022]," *Databoks*, 2022.
[5] S. I. F. Asfianur, S. Suswanta, R. Al-Hamdi, and Z. Qodir, "Discourse on moving the new capital city in East Kalimantan Province a political economy perspective," *Society, Culture and Politics*, vol. 36, no. 3, pp. 379–393, 2023, doi: 10.20473/mkp.v36i32023.379-393.

[6] M. Kamal, "Prospects for the new capital city policy in law and economic perspectives," *Substantive Justice International Journal of Law*, vol. 5, no. 1, p. 86, 2022, doi: 10.56087/substantivejustice.v5i1.205.

[7] F. A. Fahma, "The effect of moving the capital city from Jakarta to East Kalimantan on surrounding land conservation," *International Journal of Technology, Education and Social Humanities*, vol. 1, no. 1, pp. 27–32, 2023.

[8] S. Susilawati *et al.*, "The advantages and disadvantages of moving the capital city of Indonesia (equal development or destruction of land?)" *Journal of Histor*, vol. 6, no. 3, pp. 1031–1036, 2023, doi: 10.24815/jr.v6i3.33762.

[9] N. Kankanamge, T. Yigitcanlar, A. Goonetilleke, and M. Kamruzzaman, "Determining disaster severity through social media analysis: Testing the methodology with south east queensland flood tweets," *International Journal of Disaster Risk Reduction*, vol. 42, p. 101360, 2020, doi: 10.1016/j.ijdrr.2019.101360.

[10] DataReportal, "Twitter users, stats, data, trends, and more," *Global Digital Insights, DataReportal*, 2023.

[11] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A machine learning-sentiment analysis on monkeypox outbreak: an extensive dataset to show the polarity of public opinion from twitter tweets," *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

[12] P. Savci and B. Das, "Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 3, pp. 227–237, 2023.

[13] P. Assiroj, A. Kurnia, and S. Alam, "The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 12, no. 6, pp. 3843–3852, 2023, doi: 10.11591/eei.v12i6.5688.

[14] L. A. Wibowo, N. Y. Pratiwi, M. Suhartana, and E. H. Yossy, "Sentiment analysis of indonesian new capitol (IKN) on twitter using classification algorithm," in *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*, 2023, pp. 1–6, doi: 10.1109/ICCED60214.2023.10425235.

[15] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 4, pp. 1620–1630, 2020, doi: 10.11591/eei.v9i4.2352.

[16] S. Kumar, N. Kaur, Kavita, and A. Joshi, "Tweet sentiment analysis using logistic regression," *IET Conference Proceedings*, vol. 2023, no. 11, pp. 332–336, 2023, doi: 10.1049/icp.2023.1801.

[17] T. H. J. Hidayat, Y. Ruldeviyani, A. R. Aditama, G. R. Madya, A. W. Nugraha, and M. W. Adisaputra, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier," *Procedia Computer Science*, vol. 197, no. 2021, pp. 660–667, 2021, doi: 10.1016/j.procs.2021.12.187.

[18] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and logistic regression in sentiment analysis on marketplace reviews using rating-based labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, 2023, doi: 10.51519/journalisi.v5i3.539.

[19] N. Azzahra, D. Murdiansyah, and K. Lhaksmana, "Toxic comment classification on social media using support vector machine and chi square feature selection," *International Journal on Information and Communication Technology (IJoICT)*, vol. 7, no. 1, pp. 64–76, 2021, doi: 10.21108/ijoict.v7i1.552.

[20] U. I. Larasati, M. A. Muslim, R. Arifudin, and A. Alamsyah, "Improve the accuracy of support vector machine using chi square statistic and term frequency inverse document frequency on movie review sentiment analysis," *Scientific Journal of Informatics*, vol. 6, no. 1, pp. 138–149, 2019, doi: 10.15294/sji.v6i1.14244.

[21] A. P. Natasuwarna, "Support vector machine feature selection in online learning sustainability sentiment analysis," *Techno. Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/tc.v19i4.4044.

[22] Y. Yamasari, A. Qoiriah, N. Rochmawati, K. Yoshimoto, R. A. Ahmad, and O. V. Putra, "Detecting students' behavior on the e-learning system using SVM kernels - based ensemble learning algorithm," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 142–153, 2023, doi: 10.22266/ijies2023.0228.13.

[23] T. Ciu and R. S. Oetama, "Logistic regression prediction model for cardiovascular disease," *IJNMT (International Journal of New Media Technology)*, vol. 7, no. 1, pp. 33–38, 2020, doi: 10.31937/ijnmt.v7i1.1340.

[24] E. Darmaja, V. C. Mawardi, and N. J. Perdana, "Review sentiment analysis of social media applications on google playstore using logistic regression method," *Prosiding Serina*, vol. 1, no. 1, pp. 513–520, 2021, doi: 10.24912/pserina.v1i1.17504.

[25] A. Tyagi and N. Sharma, "Sentiment analysis using logistic regression and effective word score heuristic," *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 2.24, pp. 20–23, 2018, doi: 10.14419/ijet.v7i2.24.11991.

[26] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," in *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*, Elsevier, 2020, pp. 83–106.

[27] F. N. Annisa, R. H. Koestoer, and H. Anggrahita, "Spatial pattern of tourism business district (TBD) In Bogor city," in *IOP Conference Series: Earth and Environmental Science*, 2024, vol. 1291, no. 1, p. 12003, doi: 10.1088/1755-1315/1291/1/012003.

[28] D. Rahmawati and A. Muliawati, "Anies quipped about building the IKN palace when so many people are in urgent need: where is justice?," *detikNews*, 2023.

[29] A. Evandio, "Ridwan Kamil appointed as curator of IKN, anti-western state palace," *Ekonomi*, 2023.

[30] S. Deny, "Moving to IKN, government-owned buildings in Jakarta rented to private sector," *Liputan6.com*, 2023.

[31] A. K. Darmawan, M. W. Al Wajieh, M. B. Setyawan, T. Yandi, and H. Hoiriyah, "Hoax news analysis for the indonesian national capital relocation public policy with the support vector machine and random forest algorithms," *Journal of Information Systems and Informatics*, vol. 5, no. 1, pp. 150–173, 2023, doi: 10.51519/journalisi.v5i1.438.

[32] A. Halim Hasugian, R. Amanda Putri, and M. Alfan Simatupang, "Application of the Naïve Bayes classification algorithm for sentiment analysis about the move of the national capital," *Journal of Science and Social Research*, vol. 4307, no. 2, pp. 635–644, 2024, doi: https://doi.org/10.54314/jssr.v7i2.1815.

[33] N. Hasanati, T. S. Utami, and R. H. Kusumaningtyas, "Sentiment analysis on news headlines of nation's capital relocation using CNN and SVM," in *2023 11th International Conference on Cyber and IT Service Management, CITSM 2023*, 2023, pp. 1–6, doi: 10.1109/CITSM60085.2023.10455228.

[34] D. Arista, Y. Sibaroni, and S. Prasetyowati, "Sentiment analysis on Twitter (X) related to relocating the national capital using the IndoBERT method using extraction features of chi-square," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, pp. 403–411, 2024, doi: 10.30865/mib.v8i1.7198.

[35] I. Syabri, R. Sutriadi, and N. Ramadhany, "Exploring public sentiments using big data on superhub spatial development of nusantara, the new capital city of Indonesia," *Journal of Regional and City Planning*, vol. 35, no. 1, pp. 44–68, 2024, doi: 10.5614/jpwk.2024.35.1.3.

[36] C. Huda and M. B. Yel, "Sentiment analysis about the archipelago capital (IKN) using the K-nearest neighbors (KNN) and naive bayes algorithms," in *Proceedings of Annual Conference on Scientific Writing*, 2024, vol. 1, pp. 146–151.
[37] F. R. Utama, D. S. Maylawati, U. Syaripudin, D. R. Ramdania, E. Nurlatifah, and M. A. Septiadi, "Sentiment analysis regarding the name of 'nusantara' in indonesia's new capital city using convolutional neural network," in *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*, 2023, pp. 1–5, doi: 10.1109/ICCED60214.2023.10425579.

BIOGRAPHIES OF AUTHORS

Valencia Eurelia Angelie Tania is a student in the Information Systems program at Universitas Multimedia Nusantara in Tangerang, Indonesia, specializing in big data analysis. Valencia has a keen interest in data analytics and UI/UX design. Committed to her academic development, Valencia actively participates in projects and coursework that deepen her expertise in these areas. She is highly engaged in her educational pursuits and seeks opportunities to apply her knowledge and skills effectively. She can be contacted at email: valencia.tania@student.umn.ac.id.

Raymond Sunardi Oetama is a lecturer specializing in Mathematics and Computer and Information Science, earned his Bachelor's degree in Mathematics from Institut Teknologi Sepuluh Nopember in 1992. Committed to furthering his education, he obtained his Master's degree from Auckland University of Technology in 2007. As a lecturer, Raymond dedicates himself to educating students, leveraging his extensive knowledge in these academic disciplines. He is fervently dedicated to teaching, aiming to foster his students' intellectual development. Raymond maintains an active presence in the academic community, staying abreast of new trends and developments in his fields of expertise. His scholarly pursuits and his role in educational advancement reinforce his reputation in the academic community. He can be contacted at email: raymond@umn.ac.id.