Sentiment Analysis of Presidential Candidate Debates from YouTube Videos

Ulima Inas Shabrina
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
uishabrina@gmail.com

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Ratih Nur Esti Anggraini
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ratih_nea@gmail.com

Agus Tri Haryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
agustharyono24@gmail.com

Abdullah Faqih Septiyanto
Department of Informatics
Institut Teknologi Sepuluh November
Surabaya, Indonesia

2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) | 979-8-3503-5052-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIMS61812.2024.10512640

***Abstract*— The upcoming Indonesian presidential election holds immense democratic significance. Candidate debates, hosted by prominent journalist Najwa Shihab on her YouTube channel, play a crucial role in articulating visions and addressing national concerns. These debates are pivotal in amplifying public discourse and serve as primary information sources for the electorate. This research presents an extensive evaluation of various machine learning models for sentiment analysis, focusing on their performance metrics in identifying positive sentiments within Presidential Candidate Debates from YouTube videos. Models such as Complement Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) were scrutinized. Notable highlights include Bernoulli Naïve Bayes and LSTM exhibiting exceptional precision rates of 99.85% and 100%, respectively, showcasing their proficiency in accurately identifying positive sentiment instances. However, concerns of potential overfitting due to these high precision scores were raised, prompting the need for validation across diverse datasets to ensure generalizability. The findings underscore the effectiveness of these models in sentiment analysis while emphasizing the importance of further assessment for broader applicability beyond the specific dataset used in this analysis.**

Keywords*— *Sentiment Analysis; Support Vector Machines (SVM); K-Nearest Neighbors (KNN); Convolutional Neural Network (CNN); Long Short-Term Memory (LSTM); Naïve Bayes;

I. INTRODUCTION

The presidential election is one of the most important political events in Indonesia's democratic system. Candidate debates are a crucial element in this election process [1]. These debates allow presidential candidates to present their visions, plans, and views on various important issues affecting the country. These debates also allow the public to better understand the presidential candidates and their political stances [2].

In this digital era, an increasing number of people share their opinions about these debates on platforms such as YouTube. Najwa Shihab, a prominent journalist in Indonesia, often hosts presidential candidate debates on her YouTube channel. These events garner significant attention from the Indonesian public and serve as a primary source of information related to the presidential election. Comments, responses, and discussions beneath the debate videos can offer valuable insights into public sentiments regarding presidential candidates and the issues discussed in these debates.

Numerous studies in leadership assessment have utilized various methods, one of which is the Delphi approach [3]. This evaluation process encompasses multiple stages: conducting a literature review, refining domains within leadership, crafting surveys, categorizing data, and assembling keywords. Leadership, being complex, is influenced by numerous factors affecting a leader's performance. Researchers employ sentiment analysis to scrutinize the collected data. As indicated in [4], sentiment analysis is a widely utilized tool for comprehending emotional content within textual data. Research on sentiment analysis often revolves around user reactions to YouTube videos. The rationale for using YouTube data lies in its status as a ubiquitous video-based social media platform that enables comments from all viewers [5].

Furthermore, a research detailed in [6] employed sentiment analysis concerning U.S. presidential candidate Donald Trump following significant debates, utilizing Twitter to aggregate public opinions. The research aimed to gauge tweet polarity to determine public support for the candidate. By developing a text-based sentiment analysis system to scrutinize debates among Indonesian presidential candidates proves beneficial in comprehending public opinions and aiding decision-making on current affairs [7] [8] [9]. Utilizing sentiment analysis on social media platforms such as YouTube, popular in Indonesia, offers insights into public sentiments towards presidential candidates and their debates [1] [7] [9] [10].

While existing studies have contributed to sentiment analysis in political contexts, this paper stands out through a thorough exploration of sentiment in debates among Indonesian presidential candidates. Leveraging machine learning, the research examines sentiment nuances, acknowledging leadership complexity and diverse influencing factors. Differing from prior research, this research proposes a comparative analysis between traditional machine learning models (Naive Bayes, SVM, KNN) and deep learning models (LSTM, CNN) in sentiment analysis.

By exploring various Naive Bayes models (Multinomial, Bernoulli, Complement), this research aims to offer a nuanced understanding of their performance. The uniqueness lies in the comparative assessment of machine learning paradigms, highlighting methodological differences. Scrutinizing these models against the backdrop of the Indonesian presidential election contributes to nuanced sentiment analysis insights, going beyond conventional approaches. This research compares traditional and deep learning models to discern nuances in Indonesian presidential debate data. The diversity

of models allows a comprehensive evaluation of their strengths and weaknesses, considering dataset characteristics.

II. Related Research

The exploration of threshold limits for presidential and vice-presidential candidates in Indonesia's direct elections has sparked extensive debates concerning its impact on democratic principles and individual rights within the government. A recent investigation [9], utilizing qualitative research methods, aimed to gauge public opinion on the implementation of the presidential threshold subsequent to the 2019 elections. Employing Nvivo 12 Plus, data sourced from Twitter discussions (using #presidentialthreshold) and prominent news outlets such as Kompas, Detik, Tempo, Republika, and TribunNews underwent thorough analysis. The research's findings revealed a predominant focus on discussions revolving around the percentage requirement, notably highlighted in Tempo. The analysis depicted a prevailing negative sentiment, encompassing 63% of the gathered responses, while positive sentiment constituted 37%. Tempo predominantly leaned toward a more positive viewpoint, whereas Republika reflected negative sentiments. Positive perspectives predominantly emphasized parliamentary support, while negative perceptions primarily centered on candidates' prospects within the threshold framework.

A research was carried out by [11], where they amassed numerous posts related to specific keywords, amounting to hundreds of thousands, using Python as a tool for machine learning. These posts underwent several stages of processing until sentiment detection was achieved using both lexicon-based and machine-learning approaches. Their findings comprised a comparative analysis between the primary and secondary candidates, delineating their respective positive and negative sentiments. Public sentiment analysis serves multiple purposes, not only in evaluating candidates' prospects in elections but also in monitoring the consistency and quality of elected leaders.

A research [12] discuss about sentiment analysis focusing on tweets about the 2020 US Presidential election. The goal is to determine whether the sentiment in the tweets is positive or negative. The research compares two classification algorithms, Naïve Bayes and Support Vector Machine (SVM). The evaluation, using 10-fold cross-validation and a mean approach, reveals that the model achieves the highest accuracy of 82% using a linear kernel. Analyzing 10,000 tweets about Donald Trump, the model predicts that 36.88% have a neutral view of Trump, 30.78% have a positive view, and 32.34% have a negative view. Similarly, for 10,000 tweets about Joe Biden, the model predicts that 42.39% have a neutral view of Biden, 29.62% have a positive view, and 27.99% have a negative view.

The next research, [13], conducted research on sentiment analysis in e-sports for the educational curriculum. This research was carried out to measure opinions or distinguish between positive and negative sentiments regarding e-sports education. Two methods were used, namely SVM and Naïve Bayes. In this research, the Synthetic Minority Over-Sampling Technique (SMOTE) was employed for evaluation. After testing, the results showed an accuracy of 50.74% for Naïve Bayes, 70.32% for Naïve Bayes + SMOTE, 70.50% for SVM, and 66.92% for SVM + SMOTE.

This research [5] assesses President Jokowi's performance in Indonesia by analyzing public opinions from social media, particularly focusing on comments from an interview video posted on YouTube. Using machine learning tools like Python and TextBlob, the research categorizes sentiments as positive or negative. The findings highlight positive sentiments towards Jokowi's personality as the most influential aspect, while negative sentiments revolve around translation quality and his English skills in the video. The analysis indicates a slight decrease in Jokowi's popularity compared to his earlier term, with 27% of sentiments emphasizing his personality traits. However, the research acknowledges limitations, suggesting the need for further research to delve into commenter profiles, including demographics and political affiliations, to distinguish rational and irrational supporters. This additional information could offer insights into the backgrounds of respondents, potentially identifying differing opinions from opposition and loyal supporters.

The research conducted by [14] introduced an NLP-based model designed to classify Arabic comments as either positive or negative. The model underwent training using a unique dataset comprising 4212 labeled comments, achieving a notable Kappa score of 0.818. Employing six classifiers—SVM, Naïve Bayes, Logistic Regression, KNN, Decision Tree, and Random Forest—the model demonstrated impressive performance with an accuracy of 94.62% and a significant MCC score of 91.46% using NB. Precision, Recall, and F1-measure for NB were all recorded at 94.64%. However, the Decision Tree showed less effective performance, reaching 84.10% accuracy and an MCC score of 69.64% without TF-IDF. This extensive research provides valuable insights for content creators seeking to improve their content and audience engagement by understanding viewers' sentiments about the videos. Furthermore, it fills a gap in the existing literature by presenting a comprehensive approach to Arabic sentiment analysis, an area that currently lacks extensive exploration in the field.

This research [15] explores the application of deep learning techniques like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) for sentiment analysis in Indonesian finance articles. Its primary objective is to categorize Indonesian news headlines based on positive or negative sentiments, employing a combination of LSTM, LSTM-CNN, and CNN-LSTM methodologies. The dataset comprises headlines sourced from Indonesian articles available on the Detik Finance website. The findings from tests indicate that the LSTM, LSTM-CNN, and CNN-LSTM methods attained accuracies of 62%, 65%, and 74%, respectively.

III. Methodology

As in the Fig. 1, the methodology used by this research initiates by collecting data from diverse sources, followed by a meticulous cleaning process to eradicate duplicates and discrepancies, ensuring the creation of a dependable dataset. This includes translating identifiers into English to standardize data for enhanced readability. Labelling is applied using lexicons, and preprocessing steps involve stopwords removal, stemming and tokenization. Subsequently, the prepared data is utilized for training a classification model. Post-training, the model undergoes an evaluation phase, assessing its accuracy, F1-score, recall, and precision to validate its effectiveness in categorizing and analyzing new data.

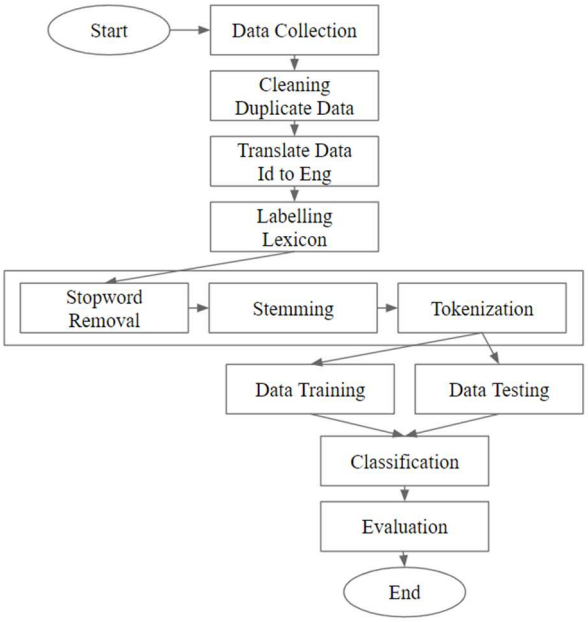

Fig. 1 Process Method

A. Data Crawling

The participants in this research comprise individuals engaging on the 2024 Indonesian Presidential Candidate Debate hosted on the Mata Najwa YouTube channel. The inclusion criteria involve users who posted comments in the specified YouTube video link ('youtube.com/watch?v=C2aZPjVdqyA&t=5s'). There are no specific exclusion criteria. The selection process involved data crawling through the YouTube API, focusing on comments related to presidential candidates Prabowo, Anies Baswedan, and Ganjar Pranowo. The data was crawled at 3 Oct 2023 resulting in a dataset of 33,429 records.

B. Data Cleaning

Identification and elimination of repeated or identical entries within a dataset. This process aims to ensure data accuracy and integrity by removing redundant information, thereby improving the quality and reliability of the dataset for analysis and modeling purposes. In this research, the data underwent cleaning and resulted in a total of 24,138 data.

C. Translate the Data from Id to Eng

Because of the limitation of the dictionary used by this research, data obtained in the Indonesian language will be translated into English using Google Translate. This aids in enhancing the accessibility and comprehensibility of the data for applications that require input in English, thereby improving its usability and interpretative capabilities. The results from the translation can see as in the Table 1.

TABLE I. TRANSLATE FROM ID TO EN

No	Comment	
	Indonesia	**English**
1	Prabowo yess	Prabowo yess
2	Keren lihat dari komentar yang dinotiz hanya calon presiden yang dijadikan penentu arah negara kalau bukan Anies ya Prabowo	Cool see from the comment netized only presidential candidates made to make the direction for the country if not anies prabowo

D. Labelling

One drawback of using machine learning is the time-consuming and subjective nature of manually labeling data to distinguish between negative and positive labels [16]. To address these limitations, the lexical-based method is employed as an alternative. This method heavily relies on the meanings of language embedded within dictionary words to mitigate the shortcomings of machine learning approaches [17]. This method assigns labels or tags to data points by matching them with specific terms or phrases listed in the lexicon. This research use 2 class label, positive and negative. As we can see on Table 2, we have total 7177 data for negative class and 16961 for positive class.

TABLE II. AMOUNT OF DATA FOR EACH CLASS

Class	Amount of Data
Negative	7177
Positive	16961
Total	24138

E. Pre-processing

Pre-processing is the initial data processing stage aimed at preparing data for further analysis. There are three stages that must be carried out for pre-processing. These stages include the following:

- Stopword Removal: Filtering out common words like "and," "the," "is," etc., which carry minimal significance in analysis, to focus on more meaningful content.
- Stemming: Reducing words to their root form to normalize variations (e.g., converting "running" and "runs" to "run") for better consistency in analysis.
- Tokenizing: Breaking down text or sentences into smaller units such as words or phrases (tokens) for analysis or processing.

TABLE III. PRE-PROCESSING

No	Comment	
	Before Pre-Processing	**After Pre-Processing**
1	Prabowo yess	Prabowo yes
2	Cool see from the comment netized only presidential candidates made to make the direction for the country if not anies prabowo	Cool comment netized presidential candidates made make direction country anies prabowo

F. Data Splitting

In this research, the dataset is divided into two sets: the train-set and the test-set, to be inputted into the classification algorithm with an 80% train-set and 20% test-set ratio of the entire dataset.

G. Classification

In this research, a comparative analysis will be conducted between two distinct paradigms of machine learning: deep learning models such as LSTM (Long Short-Term Memory) and CNN (Convolutional Neural Network) , and traditional machine learning models including Naive Bayes, SVM (Support Vector Machines), and KNN (K-Nearest Neighbors).

1) LSTM

LSTM (Long Short-Term Memory) utilizes three gates: the input gate, the forget gate, and the output gate. In LSTM, one gate component is utilized when controlling the information entering the memory, tasked with addressing the issues of

gradient loss and division. Repeated connections add state or memory to the network, enabling it to leverage sequentially observed data. The internal memory implies that the network's output depends on the last context of the input, not just the presented input itself within the network [18].

2) CNN

Convolutional Neural Networks (CNNs) are a component of deep learning architectures that come after Artificial Neural Networks (ANNs). ANNs are inspired by the human brain, where each neuron receives inputs, performs weighted sum and bias additions, and applies an activation function to produce the neuron's output. The output's mathematical representation is typically expressed as equation (1).

$$f(\sum_{1}^{\infty} \omega_i x_i + b) \tag{1}$$

3) Naïve Bayes

The naive Bayes classification method utilizes Bayesian principles and statistical techniques to forecast the probability of belonging to particular classes. It functions under the assumption of class-conditional independence, proposing that the attribute value of one class does not impact the attribute value of another class [19] [20]. This method calculates the overall probability distribution by multiplying the probability distributions of individual data points, as demonstrated in this equation (2) and (3).

$$p(d|c) = \prod_{i=1}^{m} P(w_i|c) \tag{2}$$

$$P(w_i|c) = \frac{\sum_{j=1}^{n} \delta(w_{ji}, w_i)\delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(c_j, c) + n_i} \tag{3}$$

In this context, n represents the count of training instances, n_i signifies the number of values associated with the ith attribute, c_j denotes the class label assigned to the jth training instance, w_{ji} represents the ith attribute value of the jth training instance, and $\delta(\cdot)$ is a binary function. The binary function evaluates to one if its two parameters are identical and zero otherwise.

4) SVM

Support Vector Machine (SVM) is one of the widely developed classification methods at present. The basic concept of this method is to maximize the hyperplane boundary that separates a dataset [21].

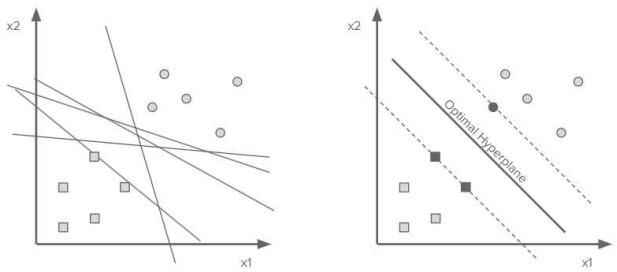

Fig. 2 SVM Illustrasion (A) left and (B) Right

The Fig. 2 illustrates the concept of SVM classification. In Figure (a), there is a set of data with circles representing class -1 and squares representing class +1. Several hyperplanes are possible for this dataset. Figure (b) shows the most optimal hyperplane. Several hyperplanes are possible for this dataset. Calculating the hyperplane involves measuring the margin distance with the closest data from each class. These closest data points are referred to as Support Vector Machines. The essence of this method lies in searching for the best hyperplane from all possibilities [22]. SVM in this study utilized a linear kernel with the function given in the equation (4).

$$K\,(x_i\,, x_j) \;=\; sum\,(x_i\,, x_j) \tag{4}$$

5) KNN

The K-Nearest Neighbor (K-NN) method is one of the top 10 most widely used data mining methods. This method performs classification based on the similarity of one data point to others. K-Nearest Neighbor is used for classifying unlabeled data. The data's characteristics are obtained from the training set and test set [23].

H. Evaluation

The classification report is a visualization used to measure the quality of prediction performance from a classification algorithm by displaying accuracy, recall, precision and F1 Score. Accuracy is used to calculate the ratio of correct predictions, both positive and negative, with respect to the entire available data. Accuration illustrated in the following equation:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{5}$$

True Positives (TP) are correct positive predictions, True Negatives (TN) are correct negative predictions. False Positives (FP) are incorrect positive predictions, and False Negatives (FN) are incorrect negative predictions. These classifications are crucial for evaluating a model's performance and understanding prediction errors.

Recall is the classifier's ability to find all positive cases. Recall illustrated in the following equation:

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

True Positives (TP) are correctly classified positive instances, while False Negatives (FN) are instances incorrectly labeled as negative instead of positive, revealing the model's failure to identify some positive cases. Precision is used to calculate the ratio or accuracy of the predictions made by the model that are correct. As shown in equation below:

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

Where, True Positives (TP) signify instances correctly predicted as the positive class by the model. Conversely, False Positives (FP) indicate instances where the model incorrectly identifies the positive class when the actual class is negative, presenting a type of error in which the model assigns something to a specific class erroneously.

F1-score is the weighted harmonic mean of precision and recall. F1-score has a lower accuracy percentage than accuracy because it embeds precision and recall into its calculation. F1-score illustrated in the following equation:

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{8}$$

The F1-score serves harmonic mean between precision and recall. Precision, evaluating the accuracy of positive predictions among all predicted positives, is represented by True Positives divided by True Positives plus False Positives. Meanwhile, recall measures the model's capability to capture all actual positives, calculated as True Positives divided by True Positives plus False Negatives. The F1-score strikes a balance between precision and recall, offering a single metric useful for assessing model performance, particularly beneficial when dealing with imbalanced class distributions.

IV. Result and Discussion

A. Experiment Setup

This research involved collecting 33,429 YouTube comments through the YouTube API, subsequently cleaning the data by removing duplicates, resulting in 24,138 records. The dataset underwent translation from Indonesian to English using Google Translate for improved accessibility. Data labeling into positive and negative classes utilized a lexical-based method, followed by pre-processing steps such as stopword removal, stemming, and tokenization. The dataset was then split into an 80% train-set and a 20% test-set for classification, employing Naïve Bayes algorithms and Support Vector Machines with various models. This study aims to conduct sentiment analysis on Indonesian presidential debates using both machine learning and deep learning algorithms.

The machine learning models encompass Naïve Bayes Multinomial, Naïve Bayes Bernoulli, and Naïve Bayes Complement, all using default parameters. KNN experiments involved adjustments to K parameters, with optimal results obtained at K=5. SVM in this study utilized a linear kernel. Deep learning models, including CNN and LSTM, underwent parameter adjustments for optimal performance. The CNN model employed parameters like a dropout rate of 0.2, learning rate of 0.01, relu activation, kernel sizes (3,3), (5,5), filters 32 and 64. Meanwhile, the LSTM model utilized a dropout rate of 0.2, learning rate of 0.01, and units of 64 and 128.

B. Experiment Result

This research assess using a variety of machine learning models. The effectiveness of Complement Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) was assessed based on their accuracy, recall, precision, and F1 scores as shown at Table IV below:

TABLE IV. Experiment Result

Model	Accuracy	Recall	Precision	F1
Complement Naïve Bayes	94.24%	99.67%	94.53%	97.03%
Multinomial Naïve Bayes	99.21%	99.54%	99.67%	99.60%
Bernoulli Naïve Bayes	99.40%	99.54%	99.85%	99.70%
SVM	99.59%	99.67%	99.92%	99.80%
KNN	99.50%	99.56%	99.94%	99.75%
CNN	99.54%	99.54%	99.99%	99.77%
LSTM	**99.75%**	**99.75%**	**100%**	**99.87%**

The examination commenced with a diverse range of models, showcasing their prowess in sentiment classification. Complement Naïve Bayes, while achieving an accuracy of 94.24%, demonstrated a commendable recall of 99.67% and an F1 score of 97.03%. However, its precision of 94.53% lagged slightly behind, signaling a proportionately higher rate of false positives. Nevertheless, this model exhibited a notable ability to correctly identify positive sentiments but requires refinement to reduce misclassifications.

Moving forward, Multinomial Naïve Bayes entered the fray with a significant leap in accuracy, recording an impressive 99.21%. This model maintained a well-balanced performance across metrics, showcasing consistency in recall (99.54%), precision (99.67%), and an F1 score of 99.60%. Its robust performance signaled its capacity for precise classification, affirming its suitability for sentiment analysis tasks.

Moreover, Bernoulli Naïve Bayes surged ahead, presenting a remarkable accuracy rate of 99.40%. This model excelled in precision, boasting an exceptional 99.85%, suggesting its ability to minimize false positives significantly. With an F1 score of 99.70% and consistent recall performance, it emerged as a standout model, underscoring its proficiency in accurately categorizing sentiments.

The Support Vector Machines (SVM) model further strengthened the discussion by showcasing a staggering accuracy of 99.59% and an outstanding precision of 99.92%. These metrics highlighted its capability to precisely identify positive sentiments among the predicted instances. With a commendable F1 score of 99.80% and high recall, SVM validated its effectiveness in sentiment classification tasks.

Similarly, K-Nearest Neighbors (KNN) models depicted consistent high performance, recording an accuracy rate of 99.50% and an impressive precision of 99.94%. Their ability to maintain high accuracy and precision, accompanied by strong F1 scores and recall rates, solidified their position as reliable models for sentiment analysis.

Additionally, both CNN and LSTM models excelled with accuracy rates of 99.54% and 99.75%, respectively, displaying impressive precision and F1 scores. The LSTM model's attainment of a perfect precision score of 100% in sentiment analysis within the context of Presidential Candidate Debates from YouTube videos is undeniably remarkable. This achievement underscores the model's exceptional capability in correctly identifying positive sentiment instances without making any false positive predictions. Such precision indicates the LSTM model's proficiency in capturing intricate patterns and nuances within the sentiment expressions present in the debate videos. However, the possibility of overfitting to the dataset should be acknowledged, as achieving perfection in precision might raise concerns about the model's adaptability to new, unseen data. To validate its robustness and generalization, further assessment on diverse datasets or employing cross-validation techniques becomes essential to gain a comprehensive understanding of its performance beyond the confines of the specific dataset used for evaluation.

V. Conclusions and Future Work

In summary, the comprehensive analysis of sentiment in debates among Indonesian presidential candidates, leveraging both traditional machine learning and deep learning models, has yielded highly affirmative results. The models, including Complement Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Network

(CNN), and Long Short-Term Memory (LSTM), consistently demonstrated exceptional performance across multiple evaluation metrics.

The accuracy rates ranging from 94.24% to 99.75% highlight the robustness of these models in correctly identifying sentiments within the YouTube comment dataset. Additionally, the precision scores, a measure of the models' ability to accurately classify positive sentiment instances, further emphasize their effectiveness. Notably, Bernoulli Naïve Bayes and LSTM models exhibited extraordinary precision scores of 99.85% and a perfect 100%, respectively.

These findings highlight the practical applicability of sentiment analysis models, particularly in assessing public sentiment during political events like presidential debates. Unmeasurable factors, such as the translation of Indonesian to English, pose challenges in accuracy due to linguistic nuances and cultural variations. Also, it is crucial to note the possibility of overfitting to the specific dataset, given the exceptionally high precision scores. Consequently, further validation across diverse datasets is recommended to ascertain the generalization and robustness of these models in broader contexts. Overall, this research not only contributes nuanced insights to sentiment analysis methodologies but also lays a strong foundation for employing these approaches in understanding public sentiments on digital platforms during significant political events, offering promising avenues to enhance our comprehension of public sentiments for informed decision-making in political discourse.

REFERENCES

[1] A. A. Firdaus, Mahsun, A. Yudhana and I. R, "Indonesian presidential election sentiment: Dataset of response public before 2024," Data Brief, vol. 52, 2023.

[2] A. S. Wisnubroto, A. S. Saifunas, A. B. Santoso, P. K. Putra and I. Budi, "Opinion-based sentiment analysis related to 2024 Indonesian Presidential Election on YouTube," 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 318-323, 2022.

[3] R. Vaughan, R. G. Kost, D. Brassil and M. Romanick, "3184 Development of a Leadership Assessment Scale in Translational Science," vol. 3, p. 131, 2020.

[4] W. R, "Unsupervised Sentiment Analysis: How to extract sentiment from the data Towards Data Science," 2019. [Online]. Available: https://towardsdatascience.com/unsupervised-sentiment-analysis-a38bf1906483. [Accessed 16 Octobet 2020].

[5] A. Mansur, Z. Allamsyah and P. Amalia, "The Performance of Indonesia's President: A Sentiment Analysis in Social Media," IOP Conf. Series: Materials Science and Engineering, 2021.

[6] A. M and H. M, "Sentiment analysis of twitter data: Emotions revealed regarding Donald Trump during the 2015-16 primary debates," in Int. Conf. Tools with Artif. Intell. ICTAI, 2018.

[7] A. Lestari and D. Karolita, "Summarizing Netizens' Sentiments Towards the 1st Indonesian Presidential Debate using Lexicon Sentiment Analysis," in IOP Conference Series: Materials Science and Engineering, 2019.

[8] W. Budiharto and Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," Journal of Big Data, 2018.

[9] R. Gustiawan, R. D. Ropawandi and Suswanta, "Analyzing public sentiment on implementing the presidential threshold in Indonesia's presidential election system," Journal Civics Media Kajian Kewarganegaraan, 2023.

[10] N. E, "Indonesian Twitter Data Pre-processing for the Emotion Recognition," International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Vols. 58-63, 2019.

[11] O. Oyebode and R. Orji, "Social Media and Sentiment Analysis: The Nigeria Presidential Election 2019," 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0140-0146, 2019.

[12] G. Nugroho, M. D. T and L. K. M., "Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine," in e-Proceeding of Engineering, 2021.

[13] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha and W. Gata, "Sentiment Analysis on E-Sports for Education Curriiculum Using Naive Bayes and Support Vector Machine," Jurnal Ilmu Komputer dan Informasi, vol. 13(2), pp. 109 - 122, 2020.

[14] D. A. Musleh et al., "Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation," Big Data and Cognitive Computing, p. 127, 2023.

[15] D. T. Hermanto, A. Setyanto and E. T. Luthfi, "Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online," Citec Journal, 2021.

[16] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," Artificial Intelligence Review, pp. 1495-1545, 2019.

[17] A. Jurek, M. D. Mulvenna and Y. Bi, "Improved lexicon-based sentiment analysis for social media analytics," Security Information, 2015.

[18] I. Augenstein, K. Bontcheva, A. Vlachos and T. Rocktäschel, "Stance Detection with Bidirectional Conditional Encoding," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, p. 876–885, 2016.

[19] Samsir, Ambiyar, Verawardina and Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," Jurnal Media Informatika Budidarma, vol. 5, pp. 157-163, 2021.

[20] Samsir et al., "Naives Bayes Algorithm for Twitter Sentiment Analysis," Journal of Physics: Conference Series, vol. 1933, 2021.

[21] Prasetyo, Data Mining: Mengolah Data Menggunakan Metlab, Penerbit Andi, 2014.

[22] Neneng, Adi and R. Isnanto, "Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM)," Jurnal Sistem Informasi Bisnis, 2016.

[23] Z. Z, "Introduction to machine learning: K-Neirest Neighbor," Annals of Translational Medicine, 2016.