Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 10, No. 2, May, Pp. 175-186

Sentiment analysis on social media using CNN-RNN hybrid: a case study of Indonesian presidential candidate

Slamet Riyadi[1*], Fayyadh Daffa[1], Cahya Damarjati[1], Megat Syahirul Amin Megat Ali[1]
Department of Information Technology, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia[1]

Article Info
Keywords:
Sentiment Analysis, Deep Learning, Overfitting, Upsampling, Imbalanced Data

Article history:
Received: September 19, 2024
Accepted: February 04, 2024
Published: May 31, 2025

Cite:
S. Riyadi, Fayyadh Daffa, Cahya Damarjati, and Megat Syahirul Amin Megat Ali, "Sentiment Analysis on Social Media Uusing CNN-RNN Hybrid: A Case Study of Indonesian Presidential Candidate", *KINETIK*, vol. 10, no. 2, May 2025. https://doi.org/10.22219/kinetik.v10i2.2125

*Corresponding author.
Slamet Riyadi
E-mail address:
riyadi@umy.ac.id

Abstract
Research on sentiment analysis for Presidential Candidate 01 on social media cannot be ignored because there is no in-depth understanding of public perceptions and opinions circulating online. The CNN model is quite commonly used for sentiment analysis; however, this model still has quite low accuracy so modifications need to be made. This research aims to increase the accuracy of sentiment analysis through the application of a modified Convolutional Neural Network (CNN) method. The research process includes collecting tweet data related to Presidential Candidate 01 using crawling techniques, data preprocessing, sentiment labeling, data balancing, as well as dividing the dataset into training, validation and test data. The CNN model is modified with additional layers to improve the performance. The model is evaluated by measuring its accuracy, precision, recall, and F1 Score. The research results show that the modified CNN-RNN Hybrid model with the Upsampling method achieves an accuracy of 94% and F1 Score of 0.95, while the CNN-RNN Hybrid model has an accuracy of 86% and F1 Score of 0.82, the CNN Model has an accuracy of 90% and F1 Score of 0.88, and the RNN model has an accuracy of 88% and F1 Score of 0.84, which are higher compared to the Naïve Bayes and LSTM methods used in the previous research. Modifying the CNN method can significantly increase the accuracy of sentiment analysis for Presidential Candidate 01, so that it can become a more effective tool for understanding public perceptions and improving political campaign strategies.

1. Introduction

Indonesia is a democracy as indicated by a general election system that elects a president every five years. The entire Indonesian people are looking forward to the presidential election, which is one of the most important political events. In most cases, the presidential candidate, also known as the Capres, is elected by a political party or a combination of several political parties. These decisions are based on how popular they are in the eyes of the public. Twitter is one of the social networks most frequently used by the public to convey their opinions against Capres. The research on Caprres 01's sentimental analysis on social media cannot be ignored because there is no deep understanding of the perceptions and opinions of the public circulating online. It can affect the effectiveness of campaign strategies and help them assess their own reputation and respond to new problems.

Twitter and other kinds of social networks have evolved into virtual public spaces full of conversation and information communication. With the large number of users of the social network like Twitter, a person's status can be converted into sentimental data for processing and analysis by researchers. A sentiment analysis from a Twitter user's status or tweet shows the public's comments on the presidential candidate and the sentiment of the tweet. The purpose of a sentiment analysis is to analyze, identify, and express opinions or sentiments in a text. Sentiment analysis is also a process that serves to identify the opinion or sentiment from the content of a dataset, such as a text about a positive, negative, or neutral topic or event [1].

Sentiment analysis can be done using various approaches, with machine learning and deep learning being the most common approaches. In machine learning, methods such as Naïve Bayes and Support Vector Machines (SVM) are frequently used. Meanwhile, in deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are widely used to process text data [2]. CNNs are valued for their ability to identify local patterns, while RNNs are known for their strength in modeling sequential data. However, despite their effectiveness, RNNs face challenges such as vanishing or exploding gradients when handling long-term dependencies in input data [3].

A review [4], which proves that CNN and RNN models can overcome the shortcomings of short texts in deep learning, also found that the results of combining CNN and RNN outperformed the accuracy results with relatively high accuracy values. The RNN model in this study uses GRU (Gated Recurrent Unit), selected as it can anticipate explosions and vanishing gradients. GRU can control data and function similarly to LSTM, but do so without using additional memory units. The bidirectional variant of this network is a special variant that allows the use of information from subsequent time steps in addition to previous time steps to allow the system to make better predictions about the current state [5]. The aim of this study is to improve the accuracy of sentiment analysis using the CNN-RNN method. This study introduces a novelty by making a CNN-RNN hybrid model with the addition of dropout and batch normalization layers and balanced data.

2. Research Method

Figure 1 is an overview of the stages of the procedure in this study. The first step of this research is to gather the data to become a dataset. The dataset that has been taken were preprocessed in this stage too. The next step is to balance the dataset that was previously labelled. After that, the dataset was divided into three parts: data training, data validation and data test.

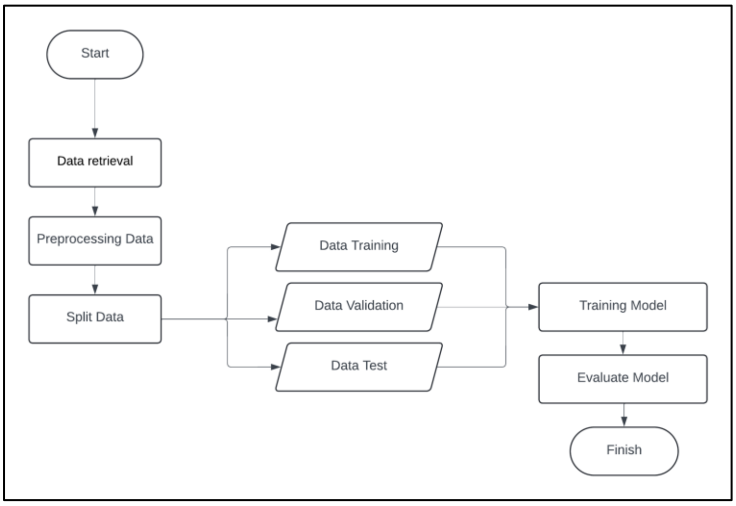

Figure 1. Research Stage

2.1 Data Retrieval

The data in the form of tweets or reviews containing the keyword "Anies" were collected using the Crawling technique. Crawling is the process of retrieving data from a data source that will be used as a dataset [6]. The process is conducted using a library called tweet-harvest developed by Helmi Satria [7]. This study uses primary data from X because X is most suitable for collecting data, especially about public opinion. This is because of the 'tweet' feature on X that makes it easier for individuals to express their opinions and engage in discussions on various issues [8]. Crawling is done by entering search_keyword: 'Anies lang:id until:2024-04-01 since:2024-01-01', which means a tweet containing an Indonesian-language keyword for the period from 01 January 2024 to 01 April 2024. In Table 1, we can see an example of data obtained from crawling in a total of 7,750 data with Anies keywords.

Table 1. Amount of Data

Created	Tweet
Sun Feb 04 14:01:31 +0000 2024	Anies: Bansos itu dibagiinnya bukan di pinggir jalan. Ganjar: https://t.co/t5xtP7Trdv
Wed Feb 07 14:18:07 +0000 2024	followers ku berkurang 2 ribu semenjak posting foto sama Pak Anies WKWKWKKW https://t.co/BJEm8GlDNe
Fri Feb 09 08:46:50 +0000 2024	Desak Anies Surabaya open gate jam 18. Ini belum juga jam 16 tapi kok udah begini :(https://t.co/uxgQWt77xU

2.2 Preprocessing Data

Preprocessing data is done to prepare the dataset that has already been retrieved before continued to the next stage to make it more structured and more accurate for data training [9]. In this study, the data preprocessing has three steps, namely:

- The first step is basic cleaning such as deleting insignificant table columns, changing upper-case to lower-case, removing emoji, repeated letter, excessive spaces, URL and non-HTML, number removal in words, symbols removal and other special characters.

- The second step is deleting the banned words that are unwanted, for example, non-standard abbreviations and so on.
- The third step is changing the slang terms into their standard equivalents using a translated word dictionary. After the preparation process is completed, the dataset will change as described in Table 2.

Table 2. Data after Preprocessing

Created	Tweet	Normalization
Sun Feb 04 14:01:31 +0000 2024	Anies: Bansos itu dibagiinnya bukan di pinggir jalan. Ganjar: https://t.co/t5xtP7Trdv	anies bansos itu dibagiinnya bukan di pinggir jalan ganjar
Wed Feb 07 14:18:07 +0000 2024	followers ku berkurang 2 ribu semenjak posting foto sama Pak Anies WKWKWKKW https://t.co/BJEm8GlDNe	followers saya berkurang ribu semenjak posting foto sama pak anies wkwkwkkw
Fri Feb 09 08:46:50 +0000 2024	Desak Anies Surabaya open gate jam 18. Ini belum juga jam 16 tapi kok udah begini :(https://t.co/uxgQWt77xU	desak anies surabaya open gate jam ini belum juga jam tapi kok sudah begini

2.3 Labeling Data

Data labelling is a process to decide sentiment towards a sentence or statement to decide whether the data is positive, neutral or negative [10]. The process of data labelling uses the TextBlob method, which is a phyton library used for Natural Language Processing (NLP). Textblob works by inserting a sentence into the textblob, which then gives two outputs, namely polarity and subjectivity. Polarity is the output that lies between [-1,1], where -1 refers to negative feelings and +1 refers to positive feelings. Subjectivity is an output which lies at [0,1] and refers to personal opinions and judgments. Textblob supports several languages, but for some languages it can also be a challenge in the process of data labelling [11]. TextBlob outperforms VADER and BERT in certain contexts, particularly in effectively handling neutral sentiment. TextBlob lexicon-based approach enables deeper and more accurate sentiment understanding than other models [12]. Since the dataset use the Indonesian language, it needs to be prepared in advance by translating it into English. The labelling results can be seen on Table 3.

Table 3. Data after Labeling

Tweet	Subjektivitas	Polaritas	Sentimen
anies bansos itu dibagiinnya bukan di pinggir jalan ganjar	0.066666666666	0.0333333333333	positif
followers saya berkurang ribu semenjak posting foto sama pak anies wkwkwkkw	0.55	-0.1	negatif
desak anies surabaya open gate jam ini belum juga jam tapi kok sudah begini	0.5	0	netral

2.4 Balancing Data

After the data labelling phase, out of 7,750 total data, 3,594 were classified as positive data, 2,569 as neutral data and 1,587 as negative data. In Figure 2, there is an imbalance data class found, requiring a data balancing. Balancing data is a technique used to match data in any class [13]. The data would be vulnerable to overfitting if they are not balanced. In this study, the method of Upsampling is used to balance the data because this method can preserve information from the minority class and improve the model's ability to predict the class without sacrificing the majority class [14]. This method works by multiplying the number of samples in the minority class (negative and neutral sentiment) to match the majority class (positive sentiment). As a result, models can effectively learn from all classes without sliding to the majority class. This approach enhances the model's ability to capture minority class intricacies, thereby improving overall classification performance without bias towards the majority class [15]. Figure 3 shows that the data becomes equal on each data class.

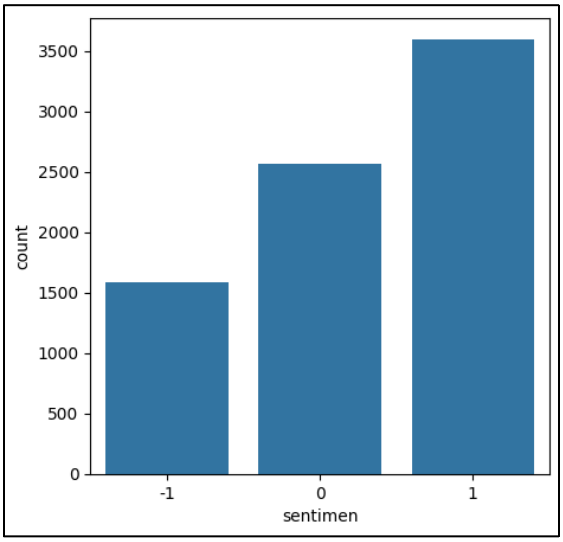

Figure 2. Data before Balancing

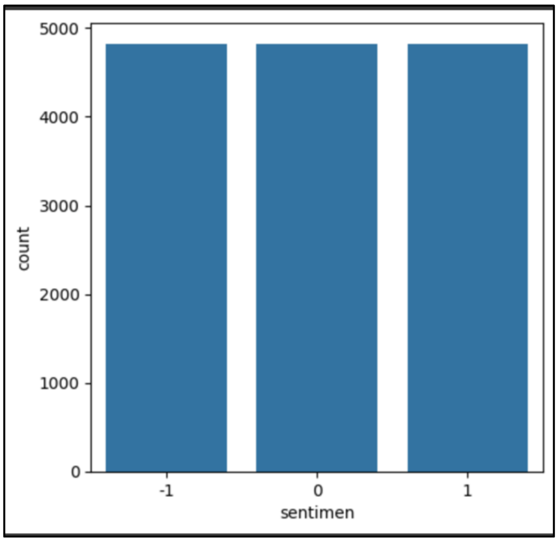

Figure 3. Data after Balancing

2.5 Data Spliting

Data splits are the process of dividing datasets into three parts: 80% training, 10% validation and 10% testing data. Training data is used to train the classification model, while test data is used in the model evaluation process. This process uses a Python library, namely sci-kit-learn. Data splits are done using a cross-validation rule that repeats sampling of test data as many folds and requires a different sample each time for testing [16]. Shared data is useful to ensure model performance on evaluation data first before applying it to the testing data. If the performance of the model at the time of training and evaluation is not quite different, then it can be said that the model has worked well.

2.6 Modeling CNN-RNN Hybrid

Modeling is the process of creating, training, and evaluating a CNN-RNN Hybrid model. In this study, the basic CNN model is used and then modified so that it has several RNN core layers that form the model to optimize predictive performance and accuracy on the CNN model validation set [17]. The CNN model architecture formed after the modification, as seen in Figure 4. This study compared the CNN model with RNN model, CNN - RNN hybrid without upsampling data, and CNN - RNN hybrid with upsampling data.

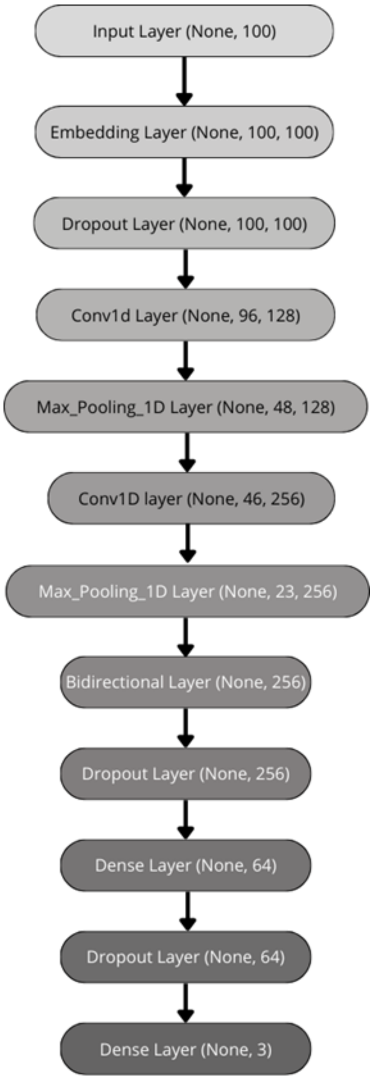

Figure 4. Modified CNN Architecture Model

The CNN architecture used is a text classification model that uses several layers to extract features from a word series in a review. This model has six convolutional layers with different filters and different filter sizes to capture various input text features. There are the Embedding layer, the Conv1d layer and the global_max_pooling_1d, the Dropout layer as well as the Dense and Output layer.

a. Embedding Layer
   Layer Embedding serves to convert words in text into fixed-dimensional vectors that can be processed by neural networks. These vectors contain semantic information of those words.
b. Conv1d Layer
   The Conv1d layer serves to apply a convolution filter to detect spatial features in text. This filter serves as a pattern detector that searches for n-grams (e.g., bigrams, trigrams) in the text.
c. max_pooling Layer
   Max pooling retains the maximum value of the input but ignores its position information, so it has limitations in maintaining positional invariance in the feature map [18].
d. Bidirectional Layer
   Bidirectional layers ensure that the model considers the entire sentence, not just the beginning or end, to understand the full meaning [19].
e. Dense Layer
   A layer dense is a fully connected layer that connects all the neurons of the previous layer with all neurons in this layer. ReLU activation is used to introduce non-linearity, helping models to capture complex relationships in data.

f. Dropout Layer
Layer Dropout works to reduce overfitting by randomly dropping out a number of neuron units during training. A 50% dropout rate means half of the neuron will be disabled on each training update.

g. Output Layer
The output layer acts as the final layer that generates a probability prediction for each class (positive, negative, neutral) using softmax activation.

2.7 Model Evaluation

After the data classification is done, it is important to evaluate the performance of the model. In this evaluation, the metrics used are accuracy and F1 Score. To calculate both metrics, the initial step is to determine the values of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The evaluation process is also carried out using the Confusion Matrix, which is an evaluation tool to tell the number of correct and wrong predictions [20]. By using the confusion matrix, we can see the values of accuracy, precision, recall, and F1 score using the formula below:

a. Accuracy
Accuracy is a metric that measures the number of true predictions corresponding to the original label of the total prediction made by the model [21]. The accuracy is measured by using Equation 1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

b. Precision
Precision is the ratio of the number of correctly predicted sentiments to the total number of classified sentiments [22]. The precision is measured by using Equation 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

c. Recall
Recall is the ratio of the number of true positive sentiments correctly classified to the total number of positive samples [23]. The recall is measured by using Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

d. F- Measure
F1-Score is an evaluation metric that calculates the harmonic averages of recall and [23]. The F1-Score is measured by Equation 4.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

3. Results and Discussion

This section discusses the results of testing the CNN model before and after the modification to find out the differences in the accuracy of sentiment analysis against Capres 01 of 2024. The process of testing the CNN model begins with training and testing the data using the model. Based on the scenarios that have been performed, the results show the CNN model testing before and after the modification with the data that has been balanced. CNN models were modified before using Layer embedding, convolution, pooling and some Layer dense, which then combined with 'Adam' for optimizer,'sparse_categorical_crossentropy' for loss function, and 'accuracy' for metric for testing.

The CNN-RNN Hybrid model also uses embedding, convolution, pooling layer and some dense layer. Bidirectional layer has been added, along with dropout layer, to reduce overfitting. The output layer with the softmax function is also used to generate class predictions for three categories of sentiments [17].

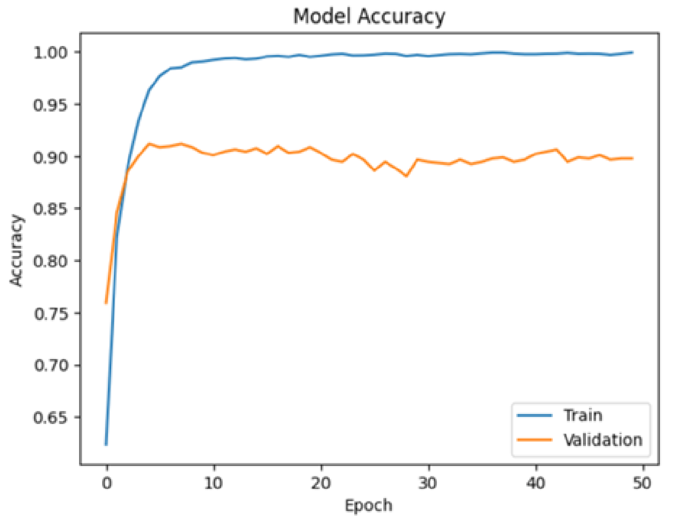

Figure 5. CNN

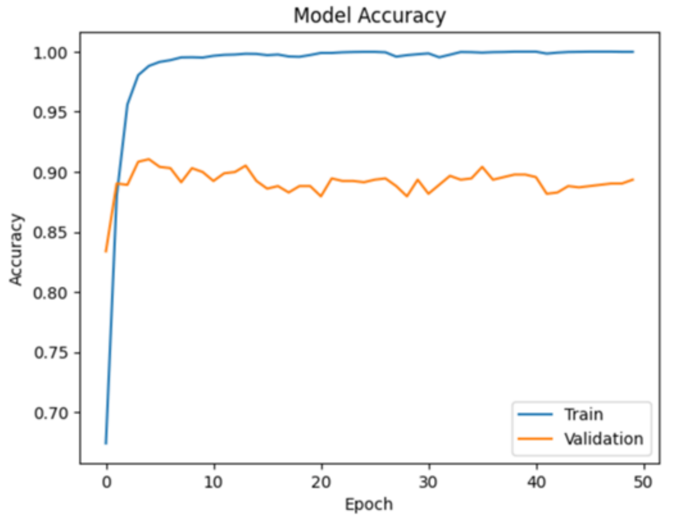

Figure 6. RNN

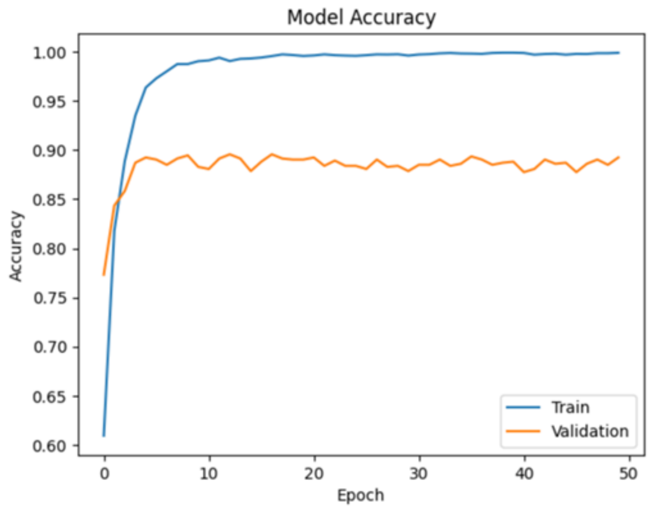

Figure 7. CNN-RNN Hybrid

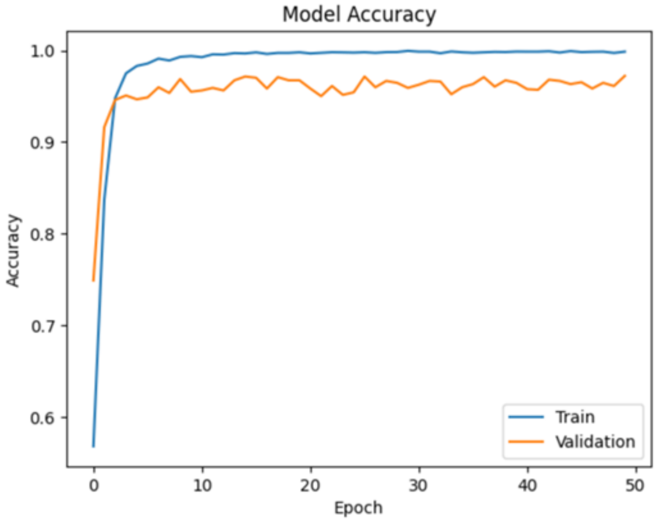

Figure 8. CNN-RNN Hybrid + Upsampling

It can be seen from the four graphs above that the curve results show the differences in CNN model validation results before and after the modification. In Figure 5, CNN model curves show overfitting. This indicates that the model is over studying the training data patterns that make the performance on new data bad [24]. Figure 6 shows the training and validation curves of the RNN model, which show fluctuations in validation accuracy indicating possible imbalances in the model's generalization ability to unseen data. Then, Figure 7 shows the improvement in stability of the hybrid CNN-RNN model compared to the baseline CNN and RNN models. Although not perfect, the curves show a trend towards more stable performance improvements in training and validation. Meanwhile in Figure 8, the modified model underwent an improvement in accuracy. This is due to good weight initialization and adequate data quality, so the model starts learning well at the start of the training process and continues to show good and stable performance in the next epoch. This increase is due to the addition of layers to the components, mainly in the use of Layer bidirectional using RNN model, which helps to stabilize the model during testing and minimize overfitting.

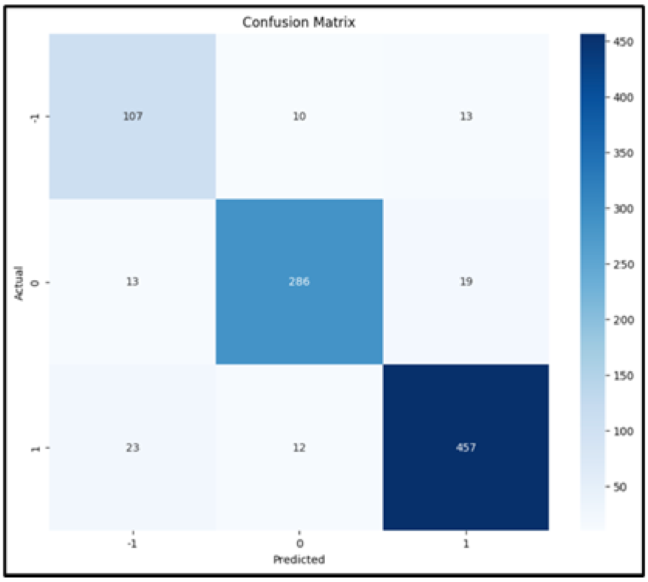

Figure 9. Confusion Matrix CNN

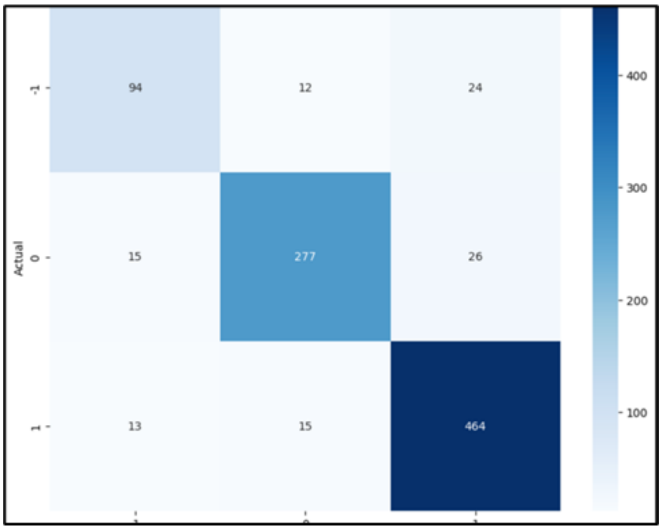

Figure 10. Confusion Matrix RNN

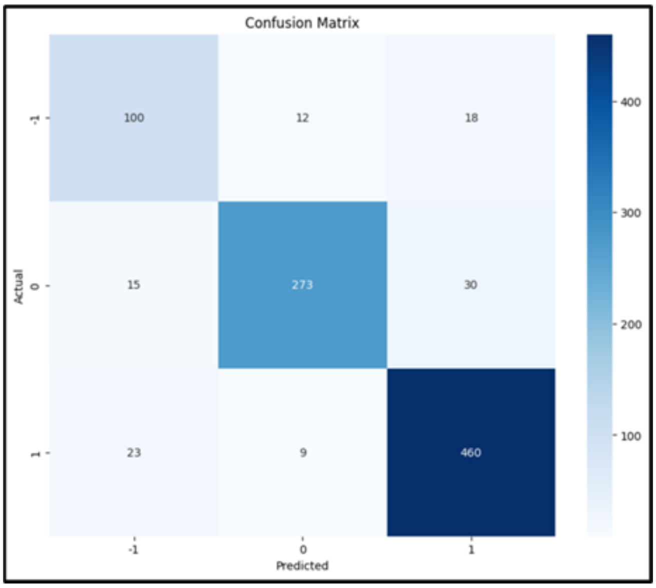

Figure 11. Confusion Matrix CNN-RNN Hybrid

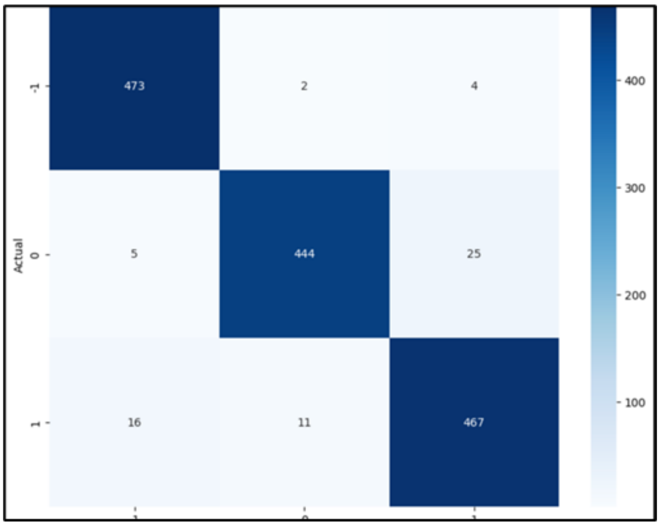

Figure12. Confusion Matrix CNN-RNN Hybrid + Upsampling

Table 4. Accuracy Results for Each Model

Metode	*Labels*	*Precision*	*Recall*	F1 Score	*Accuracy*
CNN	-1	0.75	0.82	0.78	0.82
	0	0.93	0.90	0.91	0.90
	1	0.93	0.93	0.93	0.93
RNN	-1	0.76	0.68	0.72	0.68
	0	0.90	0.87	0.88	0.87
	1	0.90	0.95	0.92	0.95
CNN-RNN Hybrid	-1	0.71	0.68	0.69	0.68
	0	0.91	0.85	0.87	0.85
	1	0.88	0.93	0.90	0.93
CNN-RNN Hybrid + Upsampling	-1	0.92	0.99	0.95	0.99
	0	0.97	0.94	0.95	0.94
	1	0.95	0.90	0.93	0.90

To measure the classification performance in more detail, the Confusion Matrix is used as shown in Figures 9 to 12, which illustrates the distribution of predictions for each sentiment class (positive, neutral, negative). Figure 9 (CNN) shows that the model is quite strong in predicting the positive class, but it is still weak in identifying the negative class. Figure 10 (RNN) indicates lower performance on negative sentiment, with many misclassifications in that class. Figure 11 shows the distribution of results from the CNN-RNN Hybrid model, where the positive class prediction performance increases, but the prediction for the negative class is still not optimal. Significant improvements are seen in Figure 12, where the CNN-RNN Hybrid with Upsampling successfully provides a more even and accurate prediction distribution across the three sentiment classes. The upsampling technique has proven effective in helping the model understand the characteristics of minority data, especially negative sentiment.

Table 4 shows the results of the performance evaluation of the CNN, RNN and Hybrid model that have been carried out. From the table, it can be observed that the CNN model obtains an accuracy of 0.90 and F1 score of 0.88, which is still considered low. Next, the result on the RNN model obtains an accuracy of 0.88 and F1-score of 0.84. Then, the CNN-RNN Hybrid Model obtains an accuracy of 0.86 and F1Score of 0.82. Further testing on the CNN-RNN Hybrid + Upsampling model obtains an accuracy of 0.94 and Score F1 of 0.95. These results show that the performance of the model has improved.

4. Conclusion

The study aims to analyze the public opinion of Capres 01 in the 2024 elections and to evaluate the effectiveness of the CNN-RNN Hybrid model for the sentiment analysis. The study focuses on the sentiment analysis associated with Capres 1 in the election of 2024 using the CNN method. The comparison shows that the CNN-RNN Hybrid with the Upsampling method obtains high accuracy of 94% with the F1 Score of 0.95, while the CNN-RNN Hybrid model has an accuracy of 86% and F1 score of 0.82, the RNN model has an accuracy of 88% and F1 scores of 0.84, and the CNN model has an accuracy of 90% and F1 score of 0.88.

This can happen because some additional RNN layers are given to the basic CNN, mainly the use of Bidirectional layers ensure that the model considers the entire sentence, not just the beginning or the end, to understand the full meaning on the CNN model [25]. The conclusion of this study is that the modification on the CNN architecture has a positive impact on an improved model performance in Capres 01's sentiment analysis of the 2024 election.

References

[1] F. Aftab *et al.*, “A Comprehensive Survey on Sentiment Analysis Techniques,” *International Journal of Technology*, vol. 14, no. 6, pp. 1288–1298, 2023. https://doi.org/10.14716/ijtech.v14i6.6632

[2] M. R. Fais Sya’ bani, U. Enri, and T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 265, Apr. 2022. https://doi.org/10.30865/jurikom.v9i2.3989

[3] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, “ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis,” *Future Generation Computer Systems*, vol. 115, pp. 279–294, Feb. 2021. https://doi.org/10.1016/j.future.2020.08.005

[4] C. N. Dang, M. N. Moreno-García, and F. De La Prieta, “Hybrid Deep Learning Models for Sentiment Analysis,” *Complexity*, vol. 2021, 2021. https://doi.org/10.1155/2021/9986920

[5] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, “Sentiment Analysis Using Gated Recurrent Neural Networks,” Mar. 01, 2020, *Springer*. https://doi.org/10.1007/s42979-020-0076-y

[6] M. Z. Rahman, Y. A. Sari, and N. Yudistira, “Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM),” 2021.

[7] Ridho Darman, “Analisis Sentimen Respons Twitter Terhadap Persyaratan Badan Penyelenggara Jaminan Sosial Kantor Pertanahan” Kantor Pertanahan Kabupaten Agam, K. Agraria dan Tata Ruang, B. Lubuk Basung, K. Agam, and P. Sumatera Barat, “Jurnal Widya Bhumi,” 2023. https://doi.org/10.31292/wb.v3i2.61

[8] M. Li and Y. Shi, “Sentiment analysis and prediction model based on Chinese government affairs microblogs,” *Heliyon*, vol. 9, no. 8, Aug. 2023. https://doi.org/10.1016/j.heliyon.2023.e19091

[9] M. Işik and H. Dağ, "The impact of text preprocessing on the prediction of review ratings," 2020, *Turkiye Klinikleri*. https://doi.org/10.3906/elk-1907-46

[10] J. Khatib Sulaiman, D. Setiyawati, N. Cahyono, and U. Amikom Yogyakarta, "Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 1, pp. 2023–262, 2023. https://doi.org/10.33022/ijcs.v12i1.3154

[11] S. Dewi and D. B. Arianto, "Twitter Sentiment Analysis Towards Qatar as Host of The 2022 World Cup Using Textblob". https://doi.org/10.55324/josr.v2i2.615

[12] H. Dia, "Evaluating the Accuracy of Sentiment Analysis Models when Applied to Social Media Texts; Evaluating the Accuracy of Sentiment Analysis Models when Applied to Social Media Texts; Utvärdering av noggrannheten hos sentimentanalysmodeller när de tillämpas på texter från sociala medier."

[13] C. Magnolia, A. Nurhopipah, D. Bagus, and A. Kusuma, "Edu Komputika Journal Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter," 2022. https://doi.org/10.15294/edukomputika.v9i2.61854

[14] J. Prasetya, "Leibniz : Jurnal Matematika Penerapan Klasifikasi Naive Bayes Dengan Algoritma Random Oversampling dan Random Undersampling Pada Data Tidak Seimbang Cervical Cancer Risk Factors," vol. 2, no. 2. https://doi.org/10.59632/leibniz.v2i2.173

[15] A. Q. Md, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease," *Biomedicines*, vol. 11, no. 2, Feb. 2023. https://doi.org/10.3390/biomedicines11020581

[16] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, pp. 746–757, 2023. https://doi.org/10.26555/jiteki.v9i3.26490

[17] M. M. A. Monshi, J. Poon, V. Chung, and F. M. Monshi, "CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR," *Comput Biol Med*, vol. 133, Jun. 2021. https://doi.org/10.1016/j.compbiomed.2021.104375

[18] L. Zhao and Z. Zhang, "A improved pooling method for convolutional neural networks," *Sci Rep*, vol. 14, no. 1, Dec. 2024. https://doi.org/10.1038/s41598-024-51258-6

[19] H. Asrawi, E. Utami, and A. Yaqin, "LSTM and Bidirectional GRU Comparison for Text Classification," *sinkron*, vol. 8, no. 4, pp. 2264–2274, Oct. 2023. https://doi.org/10.33395/sinkron.v8i4.12899

[20] M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, p. 209, Jul. 2019. https://doi.org/10.22146/ijccs.41236

[21] F. A. Irawan and D. A. Rochmah, "Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19," *JURNAL INFORMATIKA*, vol. 9, no. 2, 2022. https://doi.org/10.31294/inf.v9i2.13257

[22] R. Yacouby Amazon Alexa and D. Axman Amazon Alexa, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models." https://doi.org/10.18653/v1/2020.eval4nlp-1.9

[23] D. Hidayatul Qudsi, J. Hakim Lubis, K. Umam Syaliman, and N. Fadilah Najwa, "Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Convolutional Neural Network," 2021. https://doi.org/10.25126/jtiik.202184842

[24] S. Pavlitskaya, J. Oswald, and J. M. Zöllner, "Measuring Overfitting in Convolutional Neural Networks using Adversarial Perturbations and Label Noise," Sep. 2022. https://doi.org/10.48550/arXiv.2209.13382

[25] M. Ainur Rohman and T. Chamidy, "Bidirectional GRU dengan Attention Mechanism pada Analisis Sentimen PLN Mobile Bidirectional GRU with Attention Mechanism on Sentiment Analysis of PLN Mobile." https://doi.org/10.33633/tc.v22i2.7876