Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

Sentiment Analysis About Legislative Elections using Deep Learning with LSTM and CNN Models

Nadya Arda Angraini, Kemas Muslim Lhaksmana*

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: [1]nadyrdangrainiii@student.telkomuniversity.ac.id, [2,*]kemasmuslim@telkomuniversity.ac.id
Correspondence Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **04/06/2024**; Accepted: **29/06/2024**; Published**: 29/06/2024**

Abstract−The election of legislative members is a significant moment from the perspective of democracy, influencing the policies and direction of a country. In the digital era, sentiment analysis regarding the election of legislative members through social media has become increasingly important for analyzing public opinions and providing insights into how people respond to and feel about candidates, parties, or specific issues. The authors of this study employ deep learning methods, specifically Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models, for sentiment analysis related to legislative member elections. These models were developed and trained using preprocessed datasets. The aim of this research is to identify the highest accuracy values of the LSTM and CNN models and to analyze and classify public sentiment regarding the 2024 DPR member election.The results of this study indicate that deep learning methods can provide valuable insights into public sentiment during the 2024 legislative elections. Using a CNN model with a data ratio of 80:20, the proposed model can categorize and identify sentiments with the highest testing accuracy. It is clear that the data ratio, which provides an optimal balance between training and testing data, has a significant impact on model performance. As a result, the CNN model achieves the best results, with an accuracy of 93.27%, an F1 score of 93.19%, precision of 93.52%, and recall of 92.73. This research makes an important contribution by applying the CNN model, which succeeded in achieving the best results in categorizing sentiment, demonstrating the highest test accuracy in analyzing public sentiment towards the 2024 DPR member elections.

Keywords: Sentiment Analysis; Legislative Elections; DPR; LSTM; CNN

1. INTRODUCTION

Indonesia is a democratic country, one of which involves the election of legislative members, namely the DPR. The DPR, as a state institution in the political and legal spheres, plays a significant role in drafting legislation that adheres to the constitutional foundations of the Republic of Indonesia's 1945 Constitution [1]. The election of legislative members (DPR) is the main pillar in a democratic system that allows citizens to cast their votes in determining the representatives who will make legislative decisions. Twitter is one of the most popular social media platforms available today. Quoted from dataindonesia.id, Indonesia is ranked fifth with the largest number of Twitter users in the world as of February 2023, reaching 24 million users [2]. Opinions or viewpoints expressed via Twitter do not always receive favorable responses, but they can also elicit rejection. This makes Twitter a valuable online resource for examining opinions and preferences to conduct sentiment analysis.

Sentiment analysis is an automatic process that uses special algorithms to extract and process data, with the aim of obtaining information about the sentiments implied in an opinion text. Sentiment analysis functions as an approach to explore information about individual perceptions of an issue or classify text sentiment as positive, negative, or neutral [3]. In the context of DPR member elections, sentiment analysis is useful for analyzing public opinions and providing insight into how people respond and feel regarding certain candidates, parties, or issues.

As methods based on deep learning can capture intricate patterns and convey deeper meanings from text, they are becoming increasingly popular in sentiment analysis applications. An artificial intelligence technique called deep learning builds multilevel networks to learn features [4]. Deep learning methods adapt multilayer methodologies to neural network structures. Extracting the model naturally will result in more efficient accuracy and effectiveness [5].

In recent years, sentiment analysis (SA), a natural language processing (NLP) activity, has grown in popularity and significance for data analysis and information extraction [6]. Also known as opinion mining, it involves identifying, processing, extracting, and understanding sentiment in text or data. The main goal is to identify whether the text is positive, negative, or neutral. Sentiment analysis can be conducted at three levels: aspect, sentence, and document. At the document level, each document is assumed to have a dominant polarity. The sentence level improves the overall document polarity predictions. The aspect level analyzes the detailed sentiment toward entities or aspects within the document [7]. Sentiment analysis provides insights into the public's views on a topic, and a robust method ensures the accuracy of sentiment according to actual conditions.

This research aims to apply a deep learning-based sentiment analysis method to the election of DPR members by analyzing public opinion on social media, which is important because understanding public opinion and sentiment allows us to gain deeper insight into voter preferences, perceptions of candidates, political parties, and crucial issues that influence the process of selecting members of the DPR. Thus, by applying sentiment analysis in this context, we can develop a more comprehensive understanding of political dynamics and contribute to a more effective decision-making process. By utilizing and evaluating sentiment based on deep learning to understand public opinion regarding the election of DPR members via social media, and by evaluating sentiment towards public authorities and regulations

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

set by the DPR, it is hoped that we can more effectively and accurately assess the news and input submitted by the public regarding decisions that will be issued next.

This research aims to apply a deep learning-based sentiment analysis method to the election of DPR members by analyzing public opinion on social media. This is important because understanding public opinion and sentiment allows us to gain deeper insight into voter preferences, perceptions of candidates, and political parties. By applying sentiment analysis in this context, we can develop a more comprehensive understanding of political dynamics and contribute to a more effective decision-making process. Furthermore, by utilizing and evaluating sentiment based on deep learning to understand public opinion regarding the election of DPR members via social media, and by evaluating sentiment towards public authorities and regulations set by the DPR, it is hoped that we can more effectively and accurately assess the news and input submitted by the public regarding decisions that will be issued next.

Several studies have been conducted on sentiment analysis using deep learning applications. In 2022, Dianati Duei Putri and colleagues [1], conducted research using the Naïve Bayes model and achieved an accuracy of 0.8% or 80%, while the total test data was 20%. In 2023, Rajesh Kumar Das and his colleagues [8], conducted sentiment analysis research in a multilingual context by applying the Support Vector Machine (SVM) model and four deep learning models, namely Long Short Term Memory (LSTM), BiDirectional LSTM (Bi-LSTM), Convolutional 1D (Conv1D), and hybrid Conv1D-LSTM. The Support Vector Machine (SVM) model outperformed other models according to the study findings, obtaining 86.53% accuracy for the Bangla language sentiment analysis and 82.56% accuracy for the English text sentiment analysis. In 2022, Pavitha N. and colleagues [9], this study implemented machine learning to perform sentiment analysis using the Naïve Bayes (NB) Classifier and Support Vector Machine (SVM) Classifier methods. As NB and SVM were compared using metrics such as accuracy, precision, recall, and F1 score, the results showed that NB had an accuracy score of 97.33%, whereas SVM had an accuracy score of 98.63%. In the same year, Nurfatima Selle and colleagues [10], compared deep learning methods, namely Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). Based on the results of this study, the LSTM model with 70% training data produced an average RMSE of 46.72. The ideal scenario for nighttime data features produced an average RMSE of 51.05 with a sequence length of 30, hidden size of 8.1, one LSTM layer, and 80% training data. Then, Lilis Kurniasari and Arif Setyanto [11], compared the CNN-based + Word2Vec feature expansion, Naïve Bayes, RNN Conv, and RNN + Word2Vec feature expansion methods. The dataset used comprised reviews from the Traveloka website. According to the research findings, the RNN + Word2Vec method achieved the highest accuracy of 91.98%. The CNN + Word2Vec method ranked second, with an accuracy of 89.23%. RNN Conv ranked third, with an accuracy of 88.77%. Meanwhile, the Naïve Bayes method ranked fourth, with a significantly lower accuracy of 44.1%.

Based on these studies, there are still variations in the accuracy of determining sentiment analysis. This is due to the differences in models and the large number of datasets used. As a result, the author used methods of deep learning with LSTM (Long Short-Term Memory) and Convolutional Neural Network (CNN) models to analyze sentiment in the 2024 DPR member election. The Long Short-Term Memory (LSTM) component of the Recurrent Neural Network (RNN) architecture uses a recurrent neural network method. Because LSTM can retain knowledge over a long period of time, it can outperform standard RNNs [12]. CNN is a type of artificial neural network architecture that is intended to process spatial data and perform visual tasks, such as analyzing photos and videos. The CNN consists of several layers that gradually process data until the output layer provides the final results of the model [13]. This study chose the LSTM algorithm combined with CNN. The objective of this research is to investigate and compare the CNN and LSTM models to determine which model produces the highest accuracy value for categorizing sentiment on the 2024 DPR member election issue based on public perceptions on social media. The integration of CNN and LSTM models is anticipated to improve the performance of sentiment analysis. [14]. The scope of this study is limited to a dataset of 10,068 Indonesian-language tweets.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this research, a system was built that includes several stages aimed at analyzing sentiment related to the 2024 DPR member election. The dataset utilized in this study originates from Twitter and was analyzed employing the LSTM and CNN models/algorithms to ascertain the highest accuracy value of the two models. The following is the system flowchart design shown in Figure 1.

Based on the system flowchart design in Figure 1, the first step in planning this sentiment analysis is data collection through Twitter crawling. After the data crawling process is completed, the obtained dataset will be prepared for further implementation. The next stage involves performing data preprocessing to convert the raw data into a format that is ready for use. The preprocessing process encompasses several steps, including data cleaning, case folding, tokenization, stopwords removal, stemming, and normalization. The following data processing, feature extraction is conducted to reduce the complexity of the large dataset. One of the feature extraction methods used is Word2Vec with skip-gram and SMOTE models. Subsequently, the data will be partitioned into two parts, training data and test data. In cases where an imbalance is detected between the positive and negative classes, oversampling is employed to address this issue. Upon achieving balanced data classes, training of the LSTM and CNN models is

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

conducted to generate sentiment analysis predictions. Following this, the accuracy, precision, recall, and F1 values will be evaluated for both models.

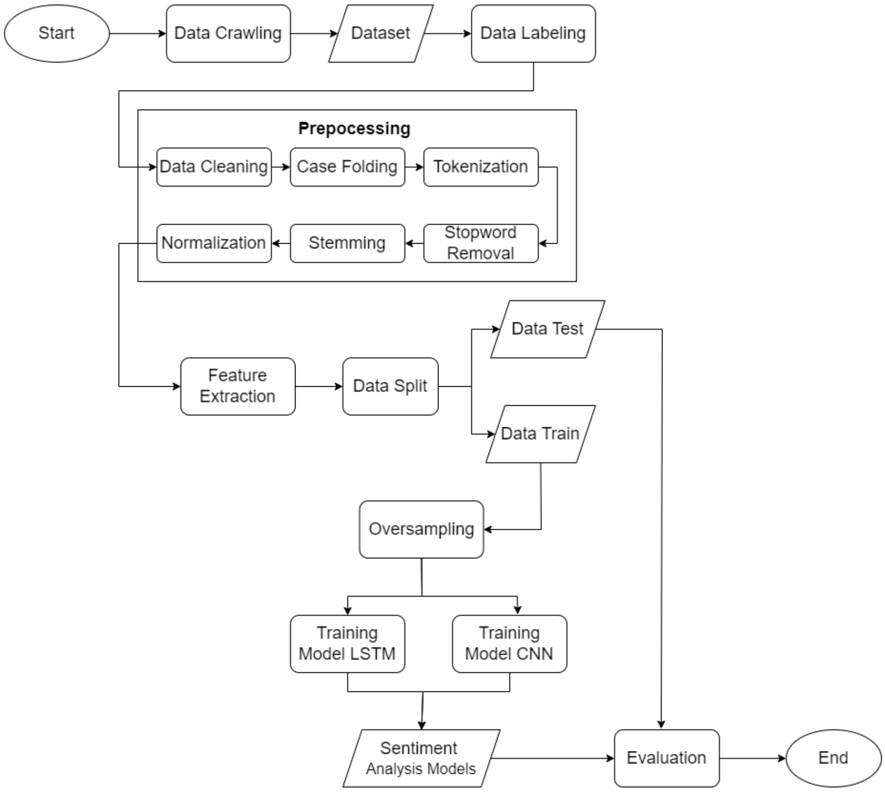

Figure 1. System Flowchart Design

2.2 Data Crawling

Data crawling involves collecting information or data from social media platforms using web crawlers or APIs (Application Programing Interface) [15]. The data used in this research were obtained by crawling Twitter, a social media network. The dataset contains public opinions or views regarding comments on the issue of the 2024 DPR member election. The dataset used in this study consists of 10.068 data.

2.3 Data Labeling

Data labeling is an integral part of the development of machine learning and involves the process of marking or labeling each data point in a dataset according to a certain class or category. This data labeling process involves categorizing each text data (tweet or social media post) based on the sentiment it contains, such as positive, negative, or neutral, regarding the issue of the 2024 DPR member election on social media. The data are grouped into three labels/classes: positive, negative, and neutral. The results of the data labeling process are shown in Table 1.

Table 1. Data Labeling

Label	Total	Percentage (%)
Positive	857	8,51
Negative	470	4,66
Neutral	8741	86,8
Total Data	**10068**	**100**

Based on Table 1, the labeling results indicate that 857 data points were labeled as positive, comprising 8.51% of the total data. A total of 470 data points were labeled negative, making up 4.66% of the total data. Meanwhile, 8741 data points were labeled neutral, which constituted the majority at 86.8% of the total data. Therefore, it can be concluded that most data falls into the neutral category, with a very high percentage of 86.8%. Data showing positive and negative sentiments accounted for only 8.51% and 4.66%, respectively.

2.4 Preprocessing

Data preprocessing is the process of converting raw data into a format that is easy to understand and use. Several steps are required during data preprocessing, as follows:

a. Data Cleaning

ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

Data cleaning is the stage of processing raw data by selecting and removing unnecessary data. This includes removing special characters, punctuation, symbols, tags, URL, emails, numbers, dates, stopwords, emojis, and spaces to prevent noise and make the text easier to understand. An example of the results of data cleaning is shown in Table 2.

Table 2. Data Cleaning

Before	After
#temanpemilih Berikut ini adalah ringkasan dan penentuan hasil perolehan suara Pemilu 2024 untuk Pemilihan DPR RI di tingkat Provinsi Jawa Tengah (#temanpemilih Here is the summary and determination of the 2024 Election results for the DPR RI election at the Central Java Provincial level)	Berikut ini adalah ringkasan dan penentuan hasil perolehan suara Pemilu 2024 untuk Pemilihan DPR RI di tingkat Provinsi Jawa Tengah (Here is the summary and determination of the 2024 Election results for the DPR RI election at the Central Java Provincial level)

Based on Table 2, the data were cleaned by removing punctuation marks, symbols, and special characters, which aimed to prevent noise and make the text easier to understand.

b. Case Folding

Case folding is the process of changing all uppercase letters in the text to lowercase. This includes removing accents and treating characters equally. Examples of the case processing results are listed in Table 3.

Table 3. Case Folding

Before	After
Berikut ini adalah ringkasan dan penentuan hasil perolehan suara Pemilu 2024 untuk Pemilihan DPR RI di tingkat Provinsi Jawa Tengah (Here is the summary and determination of the 2024 Election results for the DPR RI election at the Central Java Provincial level)	berikut ini adalah ringkasan dan penentuan hasil perolehan suara pemilu 2024 untuk pemilihan dpr ri di tingkat provinsi jawa tengah (here is the summary and determination of the 2024 election results for the dpr ri election at the central java provincial level)

Based on Table 3, the data before the case folding process still used capital letters. Therefore, at this stage, the capital letters in the text were changed to lowercase.

c. Tokenization

Tokenization involves breaking text into smaller units, with each word in a particular column separated [16]. This step highlights each word in the sentence text. Each word in the sentence is separated by a space. After tokenization, the sentence becomes a collection of arrays, where each cell contains a word from the sentence. An example of the results of tokenization is shown in Table 4.

Table 4. Tokenization

Before	After
berikut ini adalah ringkasan dan penentuan hasil perolehan suara pemilu 2024 untuk pemilihan dpr ri di tingkat provinsi jawa tengah (here is the summary and determination of the 2024 election results for the dpr ri election at the central java provincial level)	[‘berikut’, ‘ini’, ‘adalah’, ‘ringkasan’, ‘dan’, ‘penentuan’, ‘hasil’, ‘perolehan’, ‘suara’, ‘pemilu’, ‘2024’, ‘untuk’, ‘pemilihan’, ‘dpr’, ‘ri’, ‘tingkat’, ‘provinsi’, ‘jawa’, ‘tengah’] ([‘here’, ‘is’, ‘the’, ‘summary’, ‘and’, ‘determination’, ‘of’, ‘the’, ‘2024’, ‘election’, ‘results’, ‘for’, ‘the’, ‘dpr’, ‘ri’, ‘election’, ‘at’, ‘the’, ‘central’, ‘java’, ‘provincial’, ‘level’])

Based on Table 4, the tokenization stage is conducted to divide the text into smaller units with the aim of separating each word from a sentence [16].

d. Stopword Removal

Stopword removal eliminates common words that frequently appear and have no significant meaning. These common words include “and,” “or,” “the,” “of,” “on,” and others. Table 5 shows an example of the results of stopword removal.

Table 5. Stopword Removal

Before	After
[‘berikut’, ‘ini’, ‘adalah’, ‘ringkasan’, ‘dan’, ‘penentuan’, ‘hasil’, ‘perolehan’, ‘suara’, ‘pemilu’, ‘2024’, ‘untuk’, ‘pemilihan’, ‘dpr’, ‘ri’, ‘tingkat’, ‘provinsi’, ‘jawa’, ‘tengah’] ([‘here’, ‘is’, ‘the’, ‘summary’, ‘and’, ‘determination’, ‘of’, ‘the’, ‘2024’, ‘election’, ‘results’, ‘for’, ‘the’,	[‘berikut’, ‘ringkasan’, ‘penentuan’, ‘hasil’, ‘perolehan’, ‘suara’, ‘pemilu’, ‘2024,’ ‘pemilihan’, ‘dpr’, ‘ri’, ‘tingkat’, ‘provinsi’, ‘jawa’, ‘tengah’] ([‘here’, ‘summary’, ‘determination’, ‘2024’, ‘election’, ‘results’, ‘dpr', ‘ri’, ‘election’, ‘central’, ‘java’, ‘provincial’, ‘level’])

Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

‘dpr’, ‘ri’, ‘election’, ‘at’, ‘the’, ‘central’, ‘java’, ‘provincial’, ‘level’])

Based on Table 5, the stopword removal stage was conducted to eliminate common words that lack significant meaning, such as "ini," "adalah," "dan," "untuk," and others, from the data.

e. Stemming

Stemming is the process of converting words into their basic forms. The main goal of stemming is to remove affixes (prefixes and suffixes) so that words derived from the same root can be transformed into a uniform form [16]. Examples of the results obtained from stemming are shown in Table 6.

Table 6. Stemming

Before	After
[‘berikut’, ‘ringkasan’, ‘penentuan’, ‘hasil’, ‘perolehan’, ‘suara’, ‘pemilu’, ‘2024,’ ‘pemilihan’, ‘dpr’, ‘ri’, ‘tingkat’, ‘provinsi’, ‘jawa’, ‘tengah’]	[‘ikut’, ‘ringkas’, ‘tentu’, ‘hasil’, ‘oleh’, ‘suara’, ‘pemilu’, ‘2024’, ‘pilih’, ‘dpr’, ‘ri’, ‘tingkat’, ‘provinsi’, ‘jawa’, ‘tengah’]
([‘here’, ‘summary’, ‘determination’, ‘2024’, ‘election’, ‘results’, ‘dpr', ‘ri’, ‘election’, ‘central’, ‘java’, ‘provincial’, ‘level’])	([‘here’, ‘summar’, ‘determin’, ‘2024’, ‘elect’, ‘result’, ‘dpr’, ‘ri’, ‘elect’, ‘central’, ‘java’, ‘provinc’, ‘level’])

Based on Table 6, the stemming stage is conducted to transform words into their basic forms, such as "ringkasan" becoming "ringkas," "perolehan" becoming "oleh," and etc [16].

f. Normalization

Normalization is the process of simplifying and standardizing data to enable more efficient processing by an algorithm or model. This involves converting nonstandard words into standard words using a normalization dictionary.

2.5 Feature Extraction

Feature extraction is the process of identifying and extracting relevant features from text to calculate feature values and simplify data complexity. In this stage, Word2Vec is used with skip-gram and CBOW models to produce word representations as deep numerical vectors in a dimensional space [12]. Skip-gram in Word2Vec aims at predicting context words based on a specific target word, while CBOW (Continuous Bag of Words) predicts the target word (middle word) based on the context (surrounding words) [17].

2.6 Splitting Data

Splitting Data is the process of dividing a dataset into several parts for training and testing a model. The goal is to ensure that the model can be tested correctly and its performance can be measured after training. The data are divided into two parts: test data and training data. Test data are used to evaluate the model’s performance and ensure accurate predictions. On the other hand, training data are used to train the model and refine it, allowing for accurate performance estimation on new, unseen data.

2.7 Oversampling

Oversampling is a data processing technique that increases the amount of data (samples) required to balance minority and majority classes. This helps improve model performance but can cause overfitting problems. The most widely used oversampling method is SMOTE (Synthetic Minority Over-sampling Technique) [18]. Sample synthesis is based on existing minority samples in the dataset, which can effectively balance minority classes with SMOTE. This is done without needing to replicate existing sample data but by synthetically creating additional instances. This process produces new samples in minority classes to achieve a better balance.

2.8 LSTM

Long short-term memory is a type of RNN architecture designed to handle sequences of different lengths. The LSTM architecture involves layers such as the embedding layer, LSTM layer, dense layer, and fully connected layer. LSTM has a structure that includes memory cells and cell gates, which consist of three main parts: the forget gate, input gate, and output gate [19]. There are several stages in the cell gates. The first layer, the "forget gate," determines which information will be deleted from the cell state. The second layer, the "input gate," determines what new information will be stored in the cell state. Third, the "output gate" filters the output to produce relevant data [20]. LSTM uses gates to control the input of information to memory, thus overcoming the problems of gradient loss and explosion. Recurrent connections add state or memory to the network, enabling the use of sequential observations [21]. The LSTM architectural design is shown in Figure 2.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

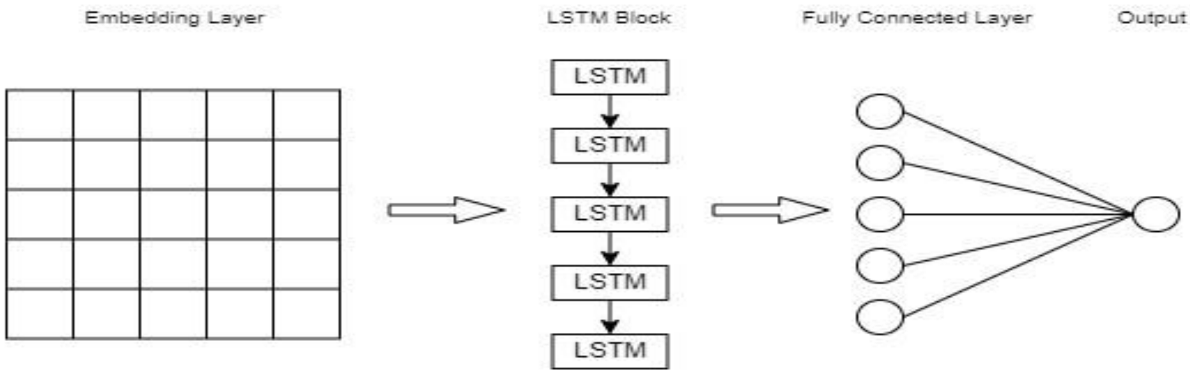

Figure 2. LSTM architecture illustration [13]

Based on Figure 2, it show the design of the LSTM Architecture in the stages of the sentiment analysis process. First, the input is transformed into an embedded vector by the embedding layer. The results of the vector embedding are then processed through LSTM blocks and subsequently forwarded to the fully connected layer for sentiment classification. At the embedding layer stage, the vectors produced undergo further processing through the LSTM block, which functions to comprehend the context and intricate patterns in the text. These LSTM blocks enable the model to grasp the relationships between words by considering their order and context. Next, the vector generated from the LSTM block is passed to the fully connected layer to perform sentiment classification. This layer is responsible for combining information from all vectors and producing a final prediction regarding the sentiment of the analyzed text. Thus, we can observe how the sentiment analysis process proceeds through a series of structured steps, from input transformation to final sentiment prediction.

2.9 CNN

CNN is a type of artificial neural network architecture in deep learning specifically designed for visual form or image processing. CNN processes input images by converting them into an array of pixel values and then classifying them into certain categories [22]. CNN performance depends on the number of hidden layers between the input and output layers [16]. Several layers of the CNN model include the input layer, convolutional layer, Rectified Linear Unit (ReLU) layer, pooling layer, fully connected layer, and output layer. The input layer is the first layer of the CNN that receives image data as input. The convolutional layer is responsible for extracting features from the input image using filters or convolution kernels. The Rectified Linear Unit (ReLU) layer functions to change every negative value to zero. This helps to recognize relevant patterns and prevent gradient problems. The pooling layer reduces the spatial dimensions of the representation produced by the convolutional layer. Max pooling is a common technique for obtaining the maximum value from a group of values, thereby reducing image resolution. Finally, the fully connected layer connects each neuron in that layer to the neuron in the next layer. The goal is to compile a complex representation of the input for classification purposes, which is then passed to the output layer that provides the model's output [13].

The advantages of CNN are its ability to automatically identify key features in images without human involvement and its higher efficiency in memory usage and complexity. However, CNN requires a large amount of training data, its training process takes a long time, and it has the potential to experience overfitting, which can reduce the algorithm's generalization ability [22]. The CNN architectural design is shown in Figure 3.

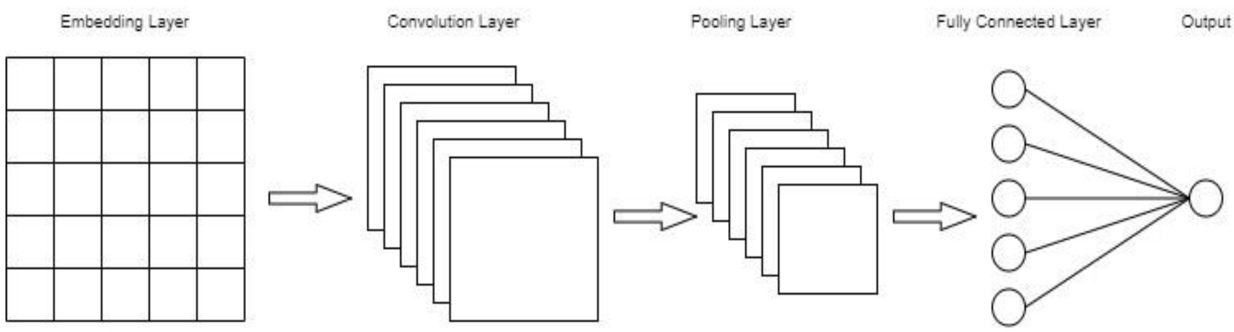

Figure 3. CNN architecture illustration [13]

Based on Figure 3, it show the design of the CNN Architecture in the sentiment analysis process is explained in detail. The first stage of this process is to convert the input into an embedded vector using an embedding layer. The embedding layer converts text data into a numerical representation that can be further processed by the model. After the input is converted into an embedded vector, the data will pass through the Convolution Layer. This convolution layer extracts important features from the input by applying filters or kernels to the data. The features extracted at this stage are crucial because they help in understanding patterns in the text data.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

After the feature extraction process, the data are forwarded to the Pooling Layer. The Pooling Layer functions to reduce the dimensions of the features extracted by the Convolution Layer. By reducing feature dimensions, the computing process becomes faster and more efficient as the number of parameters that must be processed is reduced. Pooling also helps reduce overfitting by abstracting the most important information from the extracted features.

The next step is to flatten the data, where the results from the Pooling Layer are converted into a single column. This process, known as flattening, transforms the data into a form suitable for the subsequent classification stage. Once the data are flattened, they are passed to the Fully Connected Layer. The Fully Connected Layer functions to perform sentiment classification based on the features that have been previously extracted and processed. At this stage, the model determines whether the sentiment from the input data is positive, negative, or neutral. Thus, this sentiment analysis process provides accurate classification results based on patterns and features extracted from text data [22].

3. RESULT AND DISCUSSION

This section discusses the experimental results of the models created. The research constructs test scenarios for each existing model to analyze the performance and effectiveness of each model. The method employed in this study uses LSTM and CNN modeling. Further explanations will be presented in the subsequent sections on the test results and their analysis.

3.1 Test Result

In this subsection, following the predetermined modeling stages, the subsequent stage is evaluation, which aims to test the performance of the LSTM and CNN algorithm models with the test data ratios used, namely (10:90, 20:80, and 30:70). Test results are reviewed using an evaluation matrix, including accuracy, F1-score, precision, and recall. The evaluation's conclusions are based on the weighted average section calculation and include the F1 score, recall, accuracy, and precision. Subsequently, the test results of each model are divided into two categories, namely LSTM and CNN, to obtain a deeper understanding of their performance levels in various test scenarios and different test data. The results of the model analysis above can be seen in Table 7.

Table 7. Model Testing Results

Model	Proportion Data	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
LSTM	**90:10**	**90.43 %**	90.35 %	90.74 %	90.43 %
	80:20	88.90 %	88.81 %	89.12 %	88.90 %
	70:30	88.75 %	88.62 %	89.04 %	88.75 %
CNN	90:10	93.02 %	92.98 %	93.06 %	93.02 %
	80:20	**93.27 %**	93.19 %	93.52 %	93.27 %
	70:30	92.73 %	92.67 %	92.82 %	92.73 %

Based on Table 7, shows a performance comparison between the LSTM and CNN models in classifying sentiment on a dataset that has been oversampled using three different data ratios. From the test results above, it can be concluded that the CNN model that provides the best accuracy value is the one using Word2Vec at a data split ratio of 80:20, with an accuracy value of 93.02%, an F1-score value of 92.98%, a precision value of 93.06%, and a recall value of 93.02%. Meanwhile, the LSTM model achieved the highest accuracy at a 90:10 ratio, with an accuracy score of 90.43%, an F1-score of 90.35%, a precision of 90.74%, and a recall of 90.43%. These results show better performance than the LSTM model. This confirms that the use of CNN with Word2Vec has advantages in sentiment classification in oversampled datasets, especially at a data split ratio of 80:20. Here is an illustration of the best scores from the CNN and LSTM models represented in graph form in Figure 4 and Figure 5.

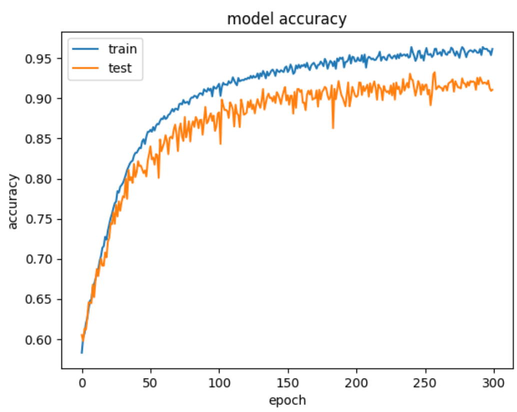

Figure 4. Best CNN Accuracy Graph

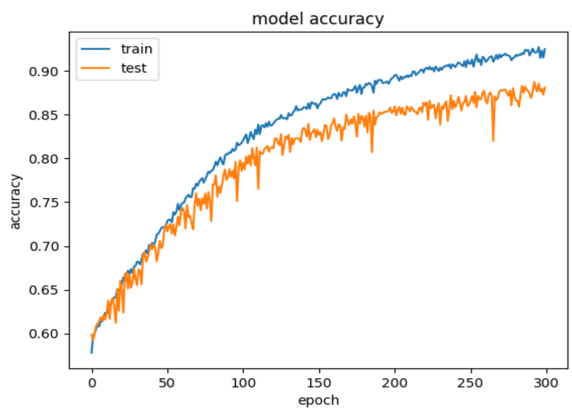

Figure 5. Best LSTM Accuracy Graph

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

Based on Figures 4 and 5, each model undergoes 300 epochs of learning, with each learning experience showing a significant increase. In Figure 4, the CNN model's learning process demonstrated considerable improvement, ultimately achieving the highest accuracy value in this research. Conversely, Figure 5 illustrates that the LSTM model also showed a substantial increase during the epoch learning process but resulted in the lowest accuracy value.

The findings of this study suggest that the CNN model's architecture is more appropriate for the given task than the LSTM model. The CNN's proficiency in capturing spatial hierarchies in the data likely contributed to its superior performance. Conversely, the LSTM model, which excels in processing sequential data, might not have been the best match for this particular dataset, resulting in lower accuracy. In addition, the learning rate and optimization strategies used during training could have significantly influenced the observed performance differences. The CNN model may have benefited from a learning rate that effectively navigated the loss landscape, allowing it to converge to a more optimal solution. In contrast, the LSTM model's training process might have encountered challenges such as vanishing gradients, which can hinder learning in deep networks.

Furthermore, the nature of the dataset itself may have impacted the results. If the data has features that are more conducive to spatial representation, the CNN model would naturally have an advantage. However, if the data required more temporal or sequential understanding, the LSTM model might have been expected to perform better, despite its lower accuracy in this experiment. Another important aspect to consider is the potential for overfitting in the models. Overfitting occurs when a model learns the training data too well, capturing noise and irrelevant details, which can negatively affect its performance on unseen data. Regularization techniques such as dropout, weight decay, and data augmentation can help mitigate this risk and improve generalization performance. It is possible that the CNN model's architecture and training regimen incorporated more effective regularization strategies, contributing to its higher accuracy.

Moreover, the evaluation metrics used in this study are crucial for interpreting the results. While accuracy is a commonly used metric, it might not always provide a complete picture of a model's performance, especially in cases of imbalanced datasets. Other metrics such as precision, recall, F1-score, and AUC-ROC can offer more nuanced insights into the models' strengths and weaknesses. For example, if the LSTM model exhibited higher recall but lower precision, it might still be considered effective in scenarios where identifying as many true positives as possible is critical. The computational resources and time required for training and evaluating these models also deserve attention. CNN models can be computationally intensive because of their complex architecture and the need to process high-dimensional data. However, they often benefit from parallelization on GPUs, which can significantly speed up training. LSTM models, while potentially less demanding in terms of memory, may require longer training times because of their sequential nature. Balancing these factors is essential for practical implementation, especially in resource-constrained environments. Finally, the potential for future work in this area is substantial. Exploring hybrid models that combine the strengths of CNNs and LSTMs could lead to better performance by leveraging both spatial and temporal features. In addition, experimenting with different hyperparameters, architectures, and preprocessing techniques could yield further improvements. Understanding the underlying reasons for the performance differences observed in this study can inform the development of more robust and effective models for similar tasks.

In conclusion, the research findings highlight the importance of selecting an appropriate model architecture based on the characteristics of the dataset and the specific task requirements. The CNN model's superior performance underscores its suitability for tasks involving spatial data, while the LSTM model's lower accuracy suggests a need for further optimization or alternative approaches for tasks requiring temporal understanding. Future research should continue to explore these avenues to enhance the accuracy and efficiency of machine learning models across various applications.

3.2 Analysis of the Test Results

Based on the results of research utilizing CNN and LSTM models with three different data ratios (90:10, 80:20, and 70:30), the data ratio significantly influences the performance of the two models. Further observations indicate that oversampling the dataset specifically enhances the performance of these models. This highlights that data manipulation indeed has a tangible impact on the final modeling results. In particular, the CNN model emerges as the superior choice over the LSTM, exhibiting more stable and superior overall performance. At a ratio of 80:20, the CNN model with the Word2Vec extraction feature achieves the highest performance in terms of accuracy, F1-score, precision, and recall, underscoring its superiority in managing oversampled data. Recommendations derived from these test results offer clear guidance on the composition of training and test data, concluding that an 80% training data and 20% test data ratio represents the optimal choice to attain satisfactory results. By employing the CNN model, this research achieves the best outcomes with a precision value of 93.27%, F1-score of 93.19%, precision of 93.52, and recall of 92.73%, thus confirming the superiority of this model in the context of sentiment classification regarding the issue of the 2023 DPR member election.

The Confusion Matrix is a table consisting of rows of test data predicted correctly and incorrectly by a classification model. This table serves to evaluate the performance of a classification model, aiding in discerning its proficiency in predicting various classes. This methodology is suitable for classification scenarios involving multiple classes or more than two classes, where the precision of predictions for each class is of paramount significance in determining the overall efficacy of the model. Through use of the Confusion Matrix, researchers gain insight into the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 1, June 2024 Page: 289−299
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

nuanced error patterns exhibited by the model, including any tendencies toward misclassification of specific classes. Consequently, this analysis facilitates the identification of areas necessitating model refinement. The results of performance evaluation using the confusion matrix for the LSTM model with the best performance are shown in Table 8.

Table 8. LSTM Confusion Matrix Model

		Prediction		
		0	**1**	**2**
	0	**833**	20	22
Actual	**1**	82	**715**	77
	2	33	17	**824**

Based on Table 8, in this case, for negative labels (0), the model correctly classifies 833 as negative (0), while 20 are predicted as neutral (1) and 22 are predicted as positive (2). For the neutral label (1), the model correctly classified 715, with 82 as negative (1) and 77 as positive (2). Finally, for positive labels (2), the model correctly classified 824, with 33 as negative (0) and 17 as neutral (1). In terms of the performance of this LSTM model, it can be observed that for each label, the model tends to perform a significant number of correct classifications, with a relatively small number of errors. This demonstrates the model's ability to recognize complex patterns and make predictions with a high level of accuracy.

The results of performance evaluation using the confusion matrix for the CNN model with the best performance are shown in Table 9.

Table 9. CNN Confusion Matrix Model

		Prediction		
		0	**1**	**2**
	0	**1714**	17	18
Actual	**1**	132	**1487**	129
	2	34	123	**1691**

Based on Table 9, in this case, for negative labels (0), the model correctly classifies 1714 as negative (0), while 17 are predicted as neutral (1) and 18 are predicted as positive (2). For the neutral label (1), the model correctly classified 1487, with 132 as negative (1) and 129 as positive (2). Finally, for the positive label (2), the model correctly classified 1691, with 34 as negative (0) and 123 as neutral (1). Analysis of the confusion matrix reveals that this CNN model maintains consistency in classification for each label, with the number of errors remaining relatively small. This affirms the reliability and stability of the model in recognizing patterns in the data and making accurate predictions. With such consistent results, this model can be deemed a reliable tool for intricate classification tasks, particularly those associated with the data used in this study.

4. CONCLUSION

In this research, it can be concluded that the results of the analysis regarding the issue of legislative member elections, especially the 2024 DPR member election, on Twitter social media exhibit a tendency toward sentiment, the majority of which is negatively oriented. The research was conducted using CNN and LSTM models with three different data ratios: 90:10, 80:20, and 70:30. It was observed that the data ratio significantly influenced the performance of the two models. Furthermore, it was noted that the performance of these models improved with the implementation of the oversampling process. Specifically, the CNN model demonstrated superior overall performance compared with the LSTM model. At a ratio of 80:20, the CNN model with Word2Vec extraction features achieved the highest performance in terms of accuracy, F1-score, precision, and recall. The results of this assessment indicate that the CNN model outperforms the other models, boasting an accuracy value of 93.27%, an F1-score of 93.19%, a precision of 93.52%, and a recall value of 92.73%. In conclusion, the CNN model exhibits the highest accuracy and performance compared with the LSTM model. As recommendations for future research, it is suggested to explore alternative feature extraction methods, integrate multiple models, and assess model performance on larger datasets across diverse contexts. This endeavor identifies optimal combinations and enhance the overall performance of the model.

REFERENCES

[1] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," Jurnal Informatika dan Teknik Elektro Terapan, vol. 10, no. 1, 2022.
[2] Sarnita Sadya, "Indonesia Masuk Negara Paling Banyak Main Twitter pada Awal 2023," Dataindonesia.id.
[3] M. I. Aditama, R. I. Pratama, K. H. Wiwaha, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," Journal Information Engineering and Educational Technology) ISSN, vol. 2549, p. 869X, 2020.
[4] C. Huang, J. Wang, S. Wang, and Y. Zhang, "A review of deep learning in dentistry," Neurocomputing, p. 126629, 2023.

ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i1.5283

[5] N. C. Dang, M. N. Moreno-Garc\'\ia, and F. la Prieta, "Sentiment analysis based on deep learning: A comparative study," Electronics (Basel), vol. 9, no. 3, p. 483, 2020.
[6] N. Habbat, H. Nouri, H. Anoun, and L. Hassouni, "Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning," Eng Appl Artif Intell, vol. 126, p. 106999, 2023.
[7] B. Liu, Sentiment analysis and opinion mining. Springer Nature, 2022.
[8] R. K. Das, M. Islam, M. M. Hasan, S. Razia, M. Hassan, and S. A. Khushbu, "Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models," Heliyon, vol. 9, no. 9, 2023.
[9] N. Pavitha et al., "Movie recommendation and sentiment analysis using machine learning," Global Transitions Proceedings, vol. 3, no. 1, pp. 279–284, 2022.
[10] N. Selle, N. Yudistira, and C. Dewi, "Perbandingan Prediksi Penggunaan Listrik dengan Menggunakan Metode Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN)," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 9, no. 1, pp. 155–162, 2022.
[11] L. Kurniasari and A. Setyanto, "Sentiment analysis using recurrent neural network," in Journal of Physics: Conference Series, 2020, p. 12018.
[12] M. I. Prabha and G. U. Srikanth, "Survey of sentiment analysis using deep learning techniques," in 2019 1st international conference on innovations in information and communication technology (ICIICT), 2019, pp. 1–9.
[13] K. A. Manurung and K. M. Laksana, "Sentiment Analysis of Tourist Attraction Review from TripAdvisor Using CNN and LSTM," International Journal on Information and Communication Technology (IJoICT), vol. 9, no. 1, pp. 73–85, 2023.
[14] D. Jain, A. Kumar, and G. Garg, "Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN," Appl Soft Comput, vol. 91, p. 106198, 2020.
[15] A. Faisal, Y. Alkhalifi, A. Rifai, and W. Gata, "Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization," JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 5, no. 2, p. 61, 2020, doi: 10.31328/jointecs. v5i2. 1362, 2020.
[16] S. Ullah, M. R. Talib, T. A. Rana, M. K. Hanif, and M. Awais, "Deep learning and machine learning-based model for conversational sentiment classification," Comput. Mater. Contin, vol. 72, pp. 2323–2339, 2022.
[17] M. Deja, I. Huvila, G. Widén, and F. Ahmad, "Seeking innovation: The research protocol for SMEs' networking," Heliyon, vol. 9, no. 4, 2023.
[18] K. Shin, J. Han, and S. Kang, "MI-MOTE: Multiple imputation-based minority oversampling technique for imbalanced and incomplete data classification," Inf Sci (N Y), vol. 575, pp. 80–89, 2021.
[19] B. A. H. Kholifatullah and A. Prihanto, "Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech," Journal of Informatics and Computer Science (JINACS), pp. 292–297, 2023.
[20] S. Utami, K. M. Lhaksmana, and Y. Sibaroni, "Deep Learning and Imbalance Handling on Movie Review Sentiment Analysis," Sinkron: jurnal dan penelitian teknik informatika, vol. 8, no. 3, pp. 1894–1907, 2023.
[21] D. T. Hermanto, A. Setyanto, and E. T. Luthfi, "Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online," Creative Information Technology Journal, vol. 8, no. 1, pp. 64–77, 2021.
[22] M. F. Herlambang, "TA: PENGENALAN KARAKTER HURUF BRAILLE DENGAN METODE CONVOLUTIONAL NEURAL NETWORK," Institut Teknologi Nasional Bandung, 2020.