

Sentiment Analysis on Acute Kidney Syrup Videos Using CNN and LSTM Algorithms

Guido Tamara ¹, Kemas Muslim L ^{2*}

^{1,2}*School Of Computing, Telkom University*

Jl. Telekomunikasi No.1 Terusan Buah Batu, Bandung, Jawa Barat, Indonesia, 40257

* kemasmuslim@telkomuniversity.ac.id

Abstract

The issue of acute kidney failure, particularly caused by the consumption of cough syrup, was circulating around October 2022 and has become a serious public health concern. This issue has drawn extensive attention and sparked various reactions on social media. In this digital era, public opinion expressed in comments on social media platforms like YouTube significantly impacts societal perceptions. Therefore, in the context of the aforementioned issue, sentiment analysis on YouTube video comments can provide valuable insights into societal perceptions and people's reactions. Therefore, this study focuses on the sentiment analysis of public opinions expressed in YouTube comments related to this matter. The methods employed for this analysis include Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) with Word2Vec feature extraction. The findings of this study indicate that both these methods produce good performance results with an oversampling dataset. In the performance comparison, CNN yielded the highest accuracy, at 0.92, while LSTM was at 0.90.

Keywords: Acute Kidney Injury, Convolutional Neural Network, Long Short-Term Memory, Sentiment Analysis, Youtube

I. INTRODUCTION

Acute kidney injury is a condition in which the kidneys suddenly stop functioning. This can occur due to disrupted blood flow, kidney problems, or urinary tract blockages. Acute kidney injury requires immediate treatment to prevent permanent kidney damage [1]. As of October 18, 2022, acute kidney injury has affected many children aged 6 months to 18 years, with 189 reported cases [2]. This resulted from the cough syrup being contaminated with risky substances such as ethylene glycol (EG), diethylene glycol (DEG), and ethylene glycol butyl ether (EGBE) [3].

As a result of acute kidney failure caused by contaminated cough syrup, the Ministry of Health has issued instructions to investigate cases of acute kidney injury in children. In addition, healthcare professionals and pharmacies have been advised not to prescribe or sell cough syrup medications until an official announcement from the government is made [4]. This issue has been widely discussed on social media platforms, including YouTube, where videos discussing the cough syrup issue have been posted. YouTube users utilize the comment section of these videos to express their opinions on the matter in text form. These opinionated comments can serve as a valuable data source for understanding public sentiment on a particular topic [5]. Therefore, sentiment analysis can be conducted to gauge public response to the issue of cough syrup causing acute kidney injury in children. However, it should be noted that data extracted from YouTube comments about this issue may face challenges due to data imbalance. In sentiment analysis, the distribution of comments among positive, neutral, and negative opinions might not be proportionally balanced. This can lead to problems in developing an accurate

and reliable model for classifying the sentiment of comments. This imbalance can lead to a model that tends to produce results leaning towards the majority sentiment, disregarding minority sentiments that also hold important contributions in the analysis [6].

Sentiment analysis is the computation of sentiment, opinion, and emotion towards an object expressed in text form. This analysis extracts attributes and components within the text and determines whether the related comments are categorized as negative or positive [7]. Sentiment analysis can be performed using various classification methods. One such method involves deep learning, specifically the Convolutional Neural Network (CNN) algorithm and Word2Vec word embeddings for word representations. Word2Vec represents dense vectors that can effectively represent relationships between words in suggestion data. This research achieved a high accuracy rate of approximately 98% [8]. Another study entitled "Text-Based Sentiment Analysis Using LSTM" where LSTM managed to get an accuracy of 85% [9]. Additional sentiment analysis research has explored deep learning for identifying fake news in the Indonesian language. The techniques employed included Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), with Word2Vec for feature extraction. The dataset comprised 1786 news items, of which 802 were factual and 984 were identified as fake news. The findings suggested that the CNN method yielded a test accuracy, precision, and recall rate of 0.88. On the other hand, the LSTM model displayed a test accuracy and precision rate of 0.84 and a recall rate of 0.83. Consequently, the CNN method appeared more effective than the LSTM approach, yet both methods could distinguish between factual and fake news in Indonesian [10].

Based on the description above, this research conducts a sentiment analysis of comments on YouTube videos related to the issue of acute kidney disease syrup medication, using a comparison of Convolutional Neural Network & Long Short-Term Memory algorithms with Word2Vec feature extraction. It's important to note that the data might need to be balanced across sentiments. This can impact the model's learning process, causing it to focus more on the majority sentiment. The dataset is taken from comments on YouTube videos about the syrup medication, on videos owned by Dr. Richard Lee, MARS. The results of this research are expected to provide an understanding of public opinion and the performance of modelling in sentiment analysis related to the issue of acute kidney disease syrup medication in Indonesia on YouTube videos.

II. LITERATURE REVIEW

The Study by Sartini [11] discussed sentiment analysis of Indonesian language tweets using the Convolutional Neural Network (CNN) algorithm. The dataset used in the study consisted of 10,806 Indonesian language tweets. This research aimed to compare the accuracy levels achieved by machine learning and deep learning methods. This study used 12 variations of CNN models with different parameters. This research showed that the 12th variation of the CNN model achieved the highest accuracy rate of 81.14%, while the 1st variation had an accuracy rate of 70.5%. In comparison, the Support Vector Machine (SVM) had an accuracy rate of 61.4%, followed by the K-Nearest Neighbors (KNN) algorithm with an accuracy rate of 52.3% and the Gradient Descent (GD) algorithm with an accuracy rate of 62.1%. Thus, it can be concluded that the CNN algorithm demonstrated better accuracy performance than machine learning methods.

Besides that, the study by Angga Kurniawan dan Metty Mustikasari [10] talked about applying deep learning to detect fake news in the Indonesian language. The deep learning methods used were CNN and LSTM, with Word2Vec employed for feature extraction. The dataset consisted of 1786 news pieces, among which 802 were factual, and 984 were classified as fake news. The research findings revealed that the CNN method boasted a test accuracy, precision, and recall rate of 0.88. Meanwhile, the LSTM model demonstrated a test accuracy and precision rate of 0.84, with a recall rate of 0.83. Consequently, the CNN method was more effective than the LSTM method, but both methods could still effectively distinguish between factual and fake news in Indonesian.

The use of CNN was also done by Listyarini dan Anggoro [12] focused on Sentiment Analysis of Regional Elections (Pilkada) during the Covid-19 pandemic using Convolutional Neural Networks (CNN). This study examined public perceptions of the regional elections held in 2020 during the Covid-19 pandemic. The research used a dataset of 500 tweets obtained through crawling data from the Twitter API. The convolutional neural network method divided data classification into positive and negative categories. Additionally, the study employed a 4-layer convolutional model and varied the number of epochs to observe their influence on

accuracy. Four variations of epochs were used in the study, precisely 50, 75, and 100. This research showed that the highest accuracy achieved using the CNN method was 90% with 100 epochs. Furthermore, it was concluded that the accuracy level would improve as the number of epoch variations increased.

The study by Widi Widayat [13] focused on sentiment analysis of movie reviews using the deep learning method LSTM and Word2Vec feature extraction. The dataset comprised 25,000 movie reviews, with the average review having 233 words. The CBOW and Skip-Gram techniques in Word2Vec were utilized to generate vector representations (word vectors) of each term in the corpus. Various dimensions of word vectors, including 50, 60, 100, 150, 200, and 500, were utilized to evaluate their influence on the ensuing accuracy. The topmost accuracy was attained with a word vector dimension of 100, hitting 88.17%, whereas the least accuracy was recorded at 85.86%, using a word vector dimension of 500.

Then, the study by Ihsan.,M et.al [14] discussed implementing the LSTM method for classifying COVID-19 vaccine sentiment on Twitter. In this study, Word2Vec was applied as input, testing the use of a trained Indonesian language model from the Wikipedia corpus. After balancing, the data consisted of 2,563 training data, 778 validation data, and 400 test data, with 1,802 neutral data, 1,066 negative data, and 566 positive data. The best result from various parameter optimization processes yielded an F1 Score of 54% with an accuracy of 66%. The study produced a model capable of classifying sentiment with new sentences.

Last, the study by Ahmad et.al [15] discusses where the LSTM method is used to analyze the sentiments of the Indonesian public regarding face-to-face learning through comments on Youtube. The analysis involves data collection stages, preprocessing, manual labelling, imbalanced classes, Splitting, Word Embedding with Word2Vec, model creation, and evaluation. With the highest accuracy of 78%, this study indicates that public sentiment tends to be neutral and positive, indicating their enthusiasm for face-to-face learning. These results are expected to serve as a reference for the government in designing strategies to increase public enthusiasm.

III. RESEARCH METHOD

A. System Design

In this research, a system is built with feature extraction intended to analyze sentiment related to the issue of acute kidney disease syrup medication on YouTube videos. The analysis is based on negative, neutral, and positive comments. The following is Figure 1 regarding the system flow.

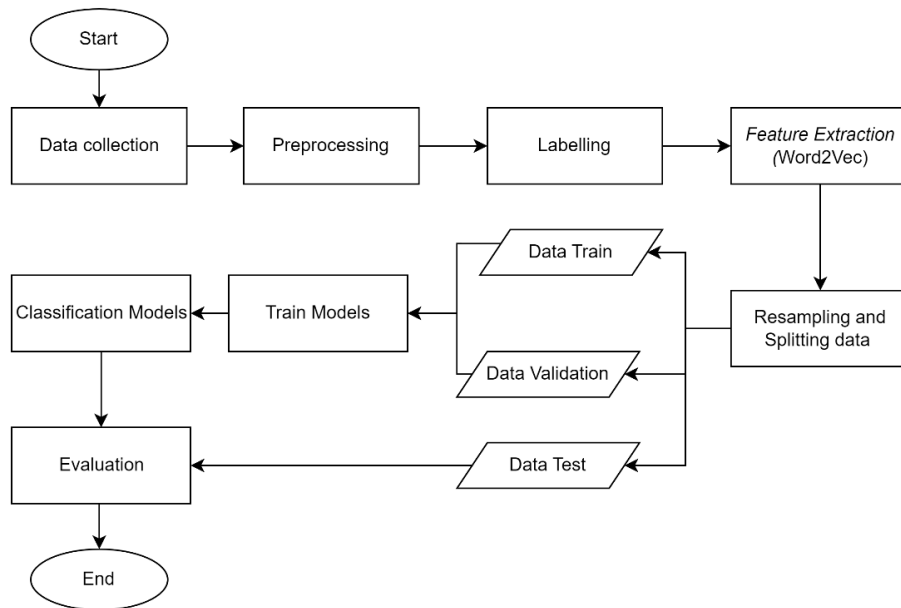


Fig. 1. Flowchart of Sentiment Analysis System

B. Data Collection

Data is collected from videos on Dr. Richard Lee’s, MARS account by scraping YouTube comments related to the syrup medication issue. After removing duplicates, 4,776 comments were obtained from three videos on Dr. Richard Lee’s, MARS YouTube channel, as shown in Table I.

TABLE I
NUMBER OF COMMENTS ON DR. RICHARD LEE’S MARS YOUTUBE CHANNEL ABOUT THE SYRUP MEDICATION ISSUE

| Video Title | Number of Comments |
|---|--------------------|
| MERESAHKAN!! OBAT SIRUP BUAT GAGAL GINJAL!?! PADAHAL SUDAH BPOM?? <i>(UNSETTLING!! A SYRUP MEDICINE FOR KIDNEY FAILURE!?! EVEN THOUGH IT'S ALREADY BPOM IMPROVED???)</i> https://youtu.be/IAQrhMowzZY | 2884 |
| UPDATE TERBARU!! INI 102 OBAT SIRUP YG DIDUGA MENJADI PENYEBAB GAGAL GINJAL ANAK!?! <i>(LATEST UPDATE!! THESE ARE THE 102 SYRUP MEDICINES SUSPECTED TO CAUSE KIDNEY FAILURE IN CHILDREN!?!)</i> https://www.youtube.com/watch?v=xSFOBE_5dGE | 1549 |
| PARAH! OBAT SIRUP ANAK INI MASIH DIJUAL?!?! PADAHAL BUAT GAGAL GINJAL!?! <i>(TERRIBLE! THIS CHILDREN'S SYRUP MEDICINE IS STILL BEING SOLD?!?! EVEN THOUGH IT CAUSES KIDNEY FAILURE?!?!)</i> https://youtu.be/yiffzzl7EFY | 371 |

C. Preprocessing

Preprocessing is a process in classification that aims to clean and prepare data so that it can be processed in other classification stages [16]. Preprocessing is performed to convert the text into term indices representing the document. The following are the preprocessing steps in Figure 2.

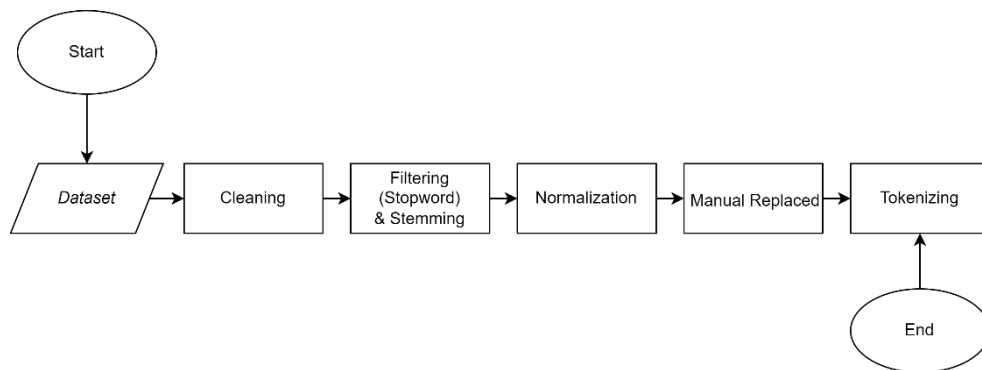


Fig. 2. Preprocessing Steps in Text Classification

In Figure 2, the first step of preprocessing involves cleaning, where characters in the text, such as hashtags, URLs, mentions, and symbols, are eliminated, resulting in the raw tweet data. This is followed by filtering, which involves stopword removal to get rid of unimportant words or those that appear in the stopword list, and the stemming process of removing affixes, prefixes, and suffixes from words to change the word into its basic form with StemmerFactory from Sastrawi library. Subsequently, any slang words are removed or replaced, and manual replacement is done to standardize certain words according to a standard dictionary. Finally, tokenization is carried out, breaking the text into individual tokens.

D. Labelling

Labelling data is assigning labels to the collected syrup medication issue data. Labelling is done manually by one human annotator repeated three times, where labels are divided into three types: label -1 as negative, 0 as neutral, and 1 as positive. A label 0 denotes a neutral sentiment, meaning the comment or text does not contain a significant positive or negative view. A -1 label refers to negative sentiment, indicating that the comment or text contains an unfavorable or pessimistic view. Conversely, label 1 indicates positive sentiment, indicating that the comment or text reflects an optimistic outlook. Table II shows some examples of data labelling.

TABLE II
EXAMPLES OF DATA LABELLING

| Number | Text | Label |
|--------|---|-------|
| 1 | Cara media massa cari uang.... Memainkan keresahan, kebingungan, kekacauan.... Sengajakah, atau kah tehnik kaliimat beirsayap? <i>(The way mass media makes money... Playing on fears, confusion, chaos... Is it deliberate, or is it a technique of sensationalism?)</i> | -1 |
| 2 | Haii Dokter!! Mau tanya apakah Vitamin Curcumapulus sudah bisa diikonsumsi utk anak2? <i>(Hello Doctor!! I wanted to ask if the Vitamin Curcumapulus can be consumed by children2?)</i> | 0 |
| 3 | Nonton ini karena anak lg demam , Kemarin ² g terlalu meingikuti peimbeiritaannya. Teirimakasih atas share peingetahuannya dokter 🙏 <i>(I watched this because my child has a fever, I didn't follow the news closely before. Thank you for sharing your knowledge, doctor 🙏)</i> | 1 |

E. Class Label Distribution

From the data collection stage, through preprocessing and data exploration stages, and after labelling for this study, there are three classes with 2625 negative labels (-1), 1245 neutral labels (0), and 906 positive labels (1). The distribution of these data labels can be seen in the following Figure 3.

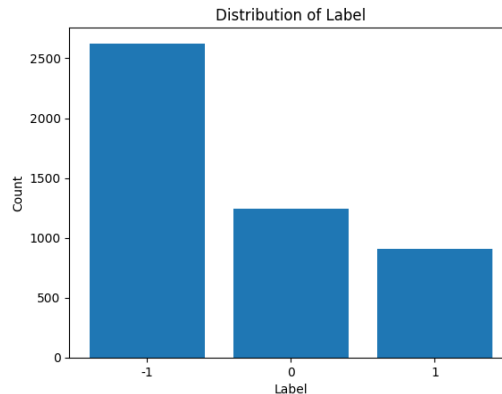


Fig. 3. Label Distribution

As seen from the context of Figure 3 Label Distribution, the results show a tendency towards negative compared to neutral and positive. There is also an imbalance of labels or imbalanced classes in these labels. In this study, to address this imbalance, resampling techniques are employed. Data-level methods to address class imbalance include oversampling and undersampling. These strategies adjust the training data distribution to mitigate the imbalance severity or diminish noise, such as mislabeled samples or outliers. At its most basic, random under-sampling removes arbitrary samples from the overrepresented category, whereas random over-sampling replicates arbitrary samples from the underrepresented category [17].

This study uses both oversampling and undersampling techniques to achieve class balance in the dataset. The reason for using these different dataset techniques is to ensure that the model can learn well from both classes. Doing this can improve the model's performance in classifying data.

F. Feature Extraction (Word2Vec)

Feature Extraction is the process of extracting a list of words from text data and then transforming them into features that classification algorithms can use. The feature extraction method used in this research is Word2Vec [18]. Word2Vec is a word embedding model that can convert words into representations of an N-length vector, where the vector is represented syntactically and semantically. Word2Vec works with a neural network whose architecture consists of an input layer, a projection (hidden layer), and an output in its architectural design [19]. The Word2Vec model consists of Skip-gram and CBOW models. The Skip-gram model is an efficient way to examine the extent of vector representation in unstructured text. The architecture of the Skip-gram word embedding model works by trying to make predictions on the context after or before the current word where the input also comes from the current word. At the same time, The CBOW model anticipates the current word solely relying on its surrounding context. [19].

The creation and training of Word2Vec CBOW and Skip-gram models with word vector dimension parameters of 100, a context window of 2, and a learning rate decrease of 0.002 per epoch for 30 epochs were performed. After training, the Word2Vec model is saved for future use, saving time and resources. This code leverages all CPU cores, accelerating training, and uses data shuffling techniques, hierarchical softmax, and negative sampling for optimization. These two Word2Vec architectures can generate word vector representations that capture contextual and semantic meaning.

G. Resampling and Splitting data

Resampling uses both oversampling and undersampling techniques to achieve class balance in the dataset. Undersampling involves reducing the number of samples in the majority class to match the number in the minority class. This helps to balance the class distribution and reduce the bias towards the majority class. Conversely, oversampling involves increasing the number of samples in the minority class by duplicating or generating synthetic samples. This helps to balance the class distribution by providing more examples of the minority class [17]. Resampling uses both oversampling and undersampling, using the RandomUnderSampler and RandomOverSampler techniques from the imblearn library. The following is Figure 4 of the dataset results after using the two resampling techniques.

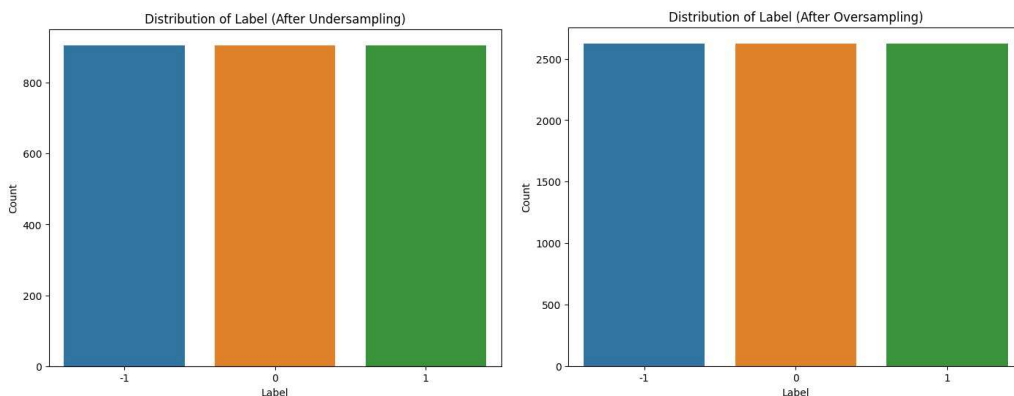


Fig. 4. Label Distribution (Resampling)

The subsequent step involves partitioning the data into sets for training, validation, and testing. The training and validation sets are parts of the dataset used to train and make predictions or execute the functions of an

algorithm. Meanwhile, the testing set is a part of the dataset used to evaluate the accuracy or performance of the algorithm based on the former model. The specified data splitting ratio is 90:10.

H. Modelling

1) *Convolutional Neural Network*: or CNN Can detect information with high accuracy, and CNN is a part of an artificial neural network (ANN) and multi-layer network, where the output of one layer becomes the input for the subsequent layers [12]. The details of the CNN architecture are as follows [20].

- Sentence Representation, the words inputted into a sentence are represented as vectors with dimension k .
- Convolutional Layer, the convolutional layer process uses filters and is applied to a window to generate new features. The filter functions to produce a feature map.
- Max-Pooling, this process serves to take the largest value from the feature map.
- Fully Connected Layer, this process functions to calculate the weight matrix, which is the output of the fully connected layer.
- Sigmoid, this process connects the output from the fully connected layer to the sigmoid layer and produces an output consisting of 2 classes, where if \hat{y} has a large value, the output will be 1, and vice versa, if \hat{y} has a small value, the output will be 0.
- Softmax, this process connects the output from the fully connected layer to the softmax layer and provides the largest probability output.

Classification using CNN, this model will be built using a Sequential architecture consisting of convolution and pooling layers. The process of building the model entails incorporating numerous layers. These include an Embedding layer, specifying embedding dimension and input length, two Conv1D layers with diverse filters and kernel sizes, two MaxPooling1D layers, a GlobalMaxPooling1D layer, a Dense layer with 256 units and relu activation function, and a final Dense output layer with 3 units using a softmax activation function. The architecture of the Convolutional Neural Network (or CNN) can be visualized in Figure 5.

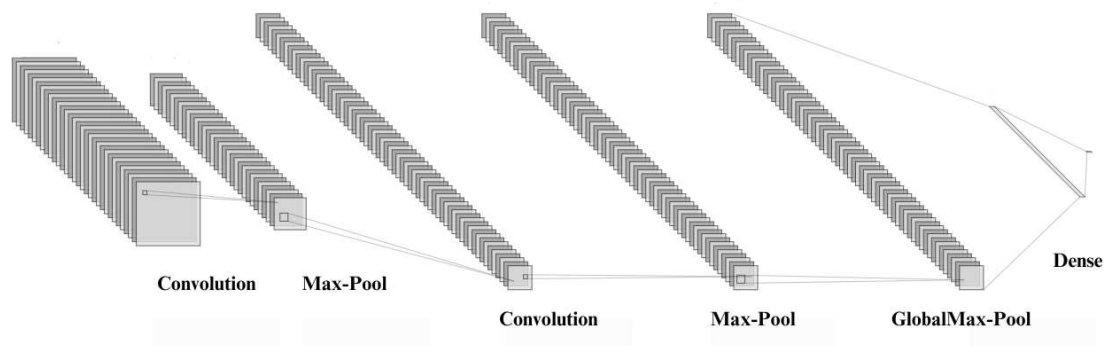


Fig. 5. Convolutional Neural Network (CNN) Architecture

2) *Long Short-Term Memory*: LSTM has demonstrated notable accuracy when working with text data, and LSTM is a summary of the deep learning RNN method that has the advantage of being able to process relatively long data (long-term dependency) [13]. LSTM has three types of gates, namely the forget gate (ft), input gate (it), and output gate (ot). The input gate functions to decide the input value to be updated on the memory state. The output gate functions to decide whether the output corresponds to the input and memory in the cell [21]. The four activation functions applied to each input within neurons are called gate units. Each piece of input data is evaluated in the forget gates, and decisions are made on what information will be kept or discarded in the memory cells. The sigmoid activation function is utilized in these forget gates as the activation function, which produces outputs ranging between 0 and 1. If the output is 1, all data will be retained; conversely, if the output is 0, all data will be removed [22].

Classification using LSTM, This model will be built using a Sequential architecture that consists of layers such as Embedding, embedding dimensions, and input length, a Dropout layer to reduce overfitting, an LSTM layer with 150 units, and a final Dense output layer that comprises three units, utilizing a softmax activation function. The LSTM architecture can be seen in Figure 6.

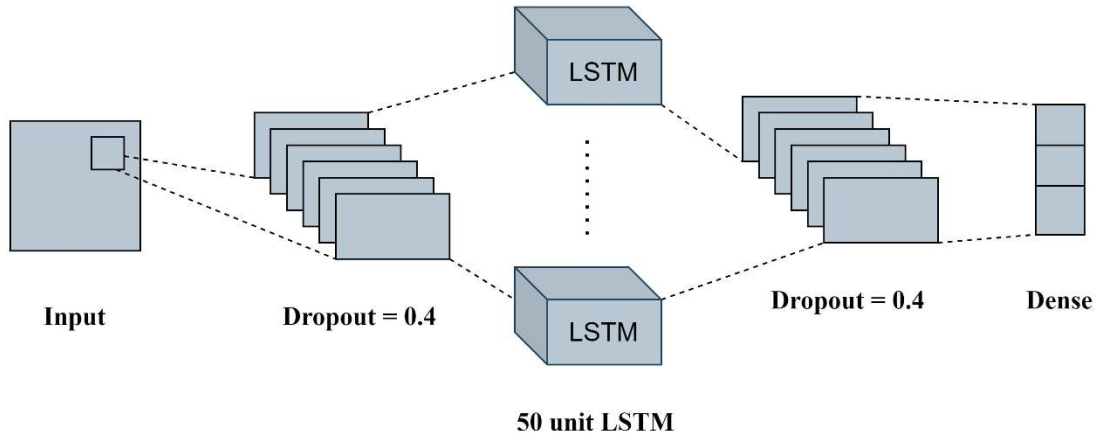


Fig. 6. Long Short-Term Memory (LSTM) Architecture

I. Evaluation

Evaluation metrics serve as measuring tools for the classification outcomes to assess the efficacy of the constructed model. In this context, the Confusion Matrix is used. The Confusion Matrix is a table that records the number of test data instances that the classification model predicts accurately and inaccurately. This table plays a crucial role in evaluating the effectiveness of a classification model. This method is typically utilized in multiple classifiers or classes that are more than binary. As such, this method is apt for this research, as it measures the accuracy of the classification results of the developed model [23].

Table III. represents a performance measurement table using the confusion matrix. There are four primary terminologies in the confusion matrix [24]:

TABLE III
CONFUSION MATRIX

| Class | Classified Positive | Classified Neutral | Classified Negative |
|----------|---------------------|---------------------|---------------------|
| Positive | TP (True Positive) | FN (False Negative) | FN (False Negative) |
| Neutral | FP (False Positive) | TN (True Negative) | TN (True Negative) |
| Negative | FP (False Positive) | TN (True Negative) | TN (True Negative) |

In Table III the following brief explanation:

- True Positive (TP) pertains to situations where the system successfully recognizes positive data.
- True Negative (TN) pertains to situations where the system successfully categorizes negative data.
- False Positive (FP) pertains to situations where the system mislabels data as positive.
- False Negative (FN) pertains to situations where the system mislabels data as negative.

IV. RESULTS AND DISCUSSION

A. Test Results

In this subsection, after going through the predetermined modelling stage, we move on to the evaluation stage, which aims to test the performance of the CNN and LSTM models with Adam optimizer (learning_rate=0.001) dan loss_function (label_smoothing=0.1) using three types of datasets from the original, undersampling, and oversampling, with feature extraction (Word2Vec). The test results include performance measures like accuracy, precision, recall, and the F1 score. Table IV shows the test results of the model on the original dataset.

TABLE IV
MODEL TEST RESULTS (ORIGINAL)

| Model | Precision | Recall | F1 Score | Accuracy |
|-------|-----------|--------|----------|-------------|
| CNN | 0.71 | 0.71 | 0.71 | 0.71 |
| | 0.72 | 0.72 | 0.72 | 0.72 |
| | 0.76 | 0.75 | 0.75 | 0.75 |
| | 0.75 | 0.73 | 0.74 | 0.73 |
| LSTM | 0.73 | 0.73 | 0.73 | 0.73 |
| | 0.73 | 0.74 | 0.72 | 0.74 |

Table IV compares the performance of the models (CNN and LSTM) in classifying sentiment on the original dataset before resampling. The CNN model performs better, achieving 0.75 accuracy, slightly superior to the LSTM, even though the difference is insignificant from the overall tests.

TABLE V
MODEL TEST RESULTS (UNDERSAMPLING)

| Model | Precision | Recall | F1 Score | Accuracy |
|-------|-----------|--------|----------|-------------|
| CNN | 0.71 | 0.71 | 0.71 | 0.71 |
| | 0.73 | 0.73 | 0.73 | 0.73 |
| | 0.76 | 0.76 | 0.76 | 0.76 |
| | 0.73 | 0.73 | 0.73 | 0.73 |
| LSTM | 0.74 | 0.74 | 0.74 | 0.74 |
| | 0.75 | 0.74 | 0.74 | 0.74 |

Table V shows the performance results on the undersampling dataset type, which slightly increases from the original dataset, with the CNN model having the highest accuracy of 0.76. However, there is a substantial difference in the oversampling dataset type, as seen in Table VI.

TABLE VI
MODEL TEST RESULTS (OVERSAMPLING)

| Model | Precision | Recall | F1 Score | Accuracy |
|-------|-----------|--------|----------|-------------|
| CNN | 0.88 | 0.88 | 0.88 | 0.88 |
| | 0.89 | 0.89 | 0.89 | 0.89 |
| | 0.93 | 0.92 | 0.92 | 0.92 |
| | 0.87 | 0.87 | 0.87 | 0.87 |
| LSTM | 0.88 | 0.88 | 0.88 | 0.88 |
| | 0.90 | 0.90 | 0.89 | 0.90 |

Table VI compares the performance of the models (CNN and LSTM) in classifying sentiment on the oversampled dataset. Overall, the results show a significant performance improvement for the models compared to the original and undersampled datasets. The CNN model achieved the highest performance in the accuracy

evaluation metric, with a result of 0.92. However, the LSTM model also demonstrated a significant increase in the evaluation metrics, indicating that oversampling techniques can enhance the model's performance in sentiment classification.

In the context of sentiment analysis, it's crucial to acknowledge the potential for errors in the process. While advanced models have shown promising accuracy, inherent complexities in natural language can pose challenges. Notably, the model might encounter difficulties when confronted with comments containing ambiguous phrases, which can contribute to the misclassification of sentiment labels. For instance, consider the comment: "Apa kerjaannya BPOM? (*What does BPOM work?*)" In this case, the word "bad" might lead the model to classify the sentiment as negative, even though the overall sentiment is nuanced and leans towards neutrality.

B. Analysis of Test Results

From the test results, the two models show good evaluation performance in all three types of data sets. In addition, it is seen that the performance of both models improves on the oversampling data set. For more details, Table VII shows the performance results on the oversampling dataset type for each sentiment.

TABLE VII
MODEL TEST RESULTS (OVERSAMPLING FOR EACH SENTIMENT)

| Model | Class | Precision | Recall | F1 Score | Accuracy |
|-------|-------|-----------|--------|----------|-------------|
| CNN | -1 | 0.96 | 0.83 | 0.89 | |
| | 0 | 0.86 | 0.96 | 0.91 | 0.92 |
| | 1 | 0.96 | 0.98 | 0.97 | |
| LSTM | -1 | 0.92 | 0.80 | 0.86 | |
| | 0 | 0.86 | 0.92 | 0.89 | 0.90 |
| | 1 | 0.91 | 0.96 | 0.94 | |

In particular, the CNN model performs better than the LSTM model. Due to local feature extraction capabilities, data processing scalability, and text pattern recognition. This analysis shows how superior the CNN model is in sentiment analysis regarding syrup medicine and acute kidney failure in YouTube comment data. The CNN, utilizing Word2Vec features on an oversampling dataset, achieves the highest Accuracy, Precision, Recall, and F1 Score performance.

Therefore, oversampling is a crucial technique that effectively mitigates class imbalance issues, particularly in datasets such as YouTube comment data. By oversampling the minor classes, their representation within the dataset is significantly increased. As a result, the model gains a deeper understanding of patterns and variations associated with these underrepresented classes, thereby enhancing its ability to identify the corresponding sentiments accurately. To elaborate further on the mechanics, it's important to clarify that the oversampling technique is exclusively applied to the training data. This approach ensures that the model is exposed to a more balanced representation during the learning phase, contributing to improved performance. The test data, however, remains unaltered to maintain an accurate evaluation of the model's generalization capabilities.

The Convolutional Neural Network (CNN) model stands out among the approaches tested, exhibiting exceptional performance when trained on Word2Vec features derived from the oversampled dataset. This superiority could be attributed to CNN's inherent capability to capture local and hierarchical features within textual data, aligning well with the intricacies of sentiment analysis. In conclusion, the synergistic combination of oversampling to address the class imbalance, a 90:10 training-to-testing ratio to provide ample training data while maintaining robust evaluation, and the utilization of the CNN model leveraging Word2Vec features culminate in the most promising outcomes for sentiment analysis on our research dataset.

The following are the Confusion Matrices for the best-performing models (CNN and LSTM), as shown in Tables VIII and IX below.

TABLE VIII
CONFUSION MATRIX (CNN-OVERSAMPLING)

| Actual | Predicted | | |
|--------|-----------|-----|-----|
| | -1 | 0 | 1 |
| -1 | 109 | 19 | 3 |
| 0 | 3 | 127 | 2 |
| 1 | 1 | 2 | 128 |

Table VIII, the model correctly classified 109 as negative (-1), while 19 of them were predicted as neutral (0) and 3 as positive (1). For the neutral label (0), the model correctly classified 127, with 3 as negative (-1) and 2 as positive (1). Lastly, for the positive label (1), the model correctly classified 128, with 1 as negative (-1) and 2 as neutral (0). The model performed sentiment classification well, especially for positive and neutral labels.

TABLE IX
CONFUSION MATRIX (LSTM-OVERSAMPLING)

| Actual | Predicted | | |
|--------|-----------|-----|-----|
| | -1 | 0 | 1 |
| -1 | 105 | 18 | 8 |
| 0 | 6 | 122 | 4 |
| 1 | 3 | 2 | 126 |

Table IX, in this case, the model correctly classified 105 as negative (-1), while 18 of them were predicted as neutral (0) and 8 as positive (1). For the neutral label (0), the model correctly classified 122. However, 6 were predicted as negative (-1) and 4 as positive (1). Lastly, for the positive label (1), the model correctly classified 126, with 3 predicted as negative (-1) and 2 as neutral (0). The LSTM model also performed well in sentiment classification, especially for positive and neutral labels.

V. CONCLUSION

Observing public comments on YouTube channels related to the issue of syrup medication causing acute kidney problems, there is a trend of predominantly negative sentiments. The results of the testing and analysis of this study indicate that techniques for handling data imbalances positively impact the performance of the sentiment analysis model. The best models found are the CNN and LSTM models, using Word2Vec on an oversampling type dataset, showing the best performance with an accuracy of 0.92, while for LSTM, it is 0.90. The CNN model performs better than the LSTM model.

As a suggestion for future research, one could try other feature extraction methods, combine models, evaluate model performance on a wider dataset and in different contexts, and consider using hyperparameter optimization techniques to find the best combination of model parameters to improve performance further.

REFERENCES

- [1] d. P. "Gagal Ginjal Akut," 25 Juli 2022. [Online]. Available: <https://www.alodokter.com/gagal-ginjal-akut>.
- [2] A. W. Finaka, "WASPADA Gagal Ginjal Akut Pada Anak," 24 Oktober 2022. [Online]. Available: <https://indonesiabaik.id/infografis/waspada-gagal-ginjal-akut-pada-anak>.

- [3] M. J. Firmansyah and F. , "Menkes Pastikan Penyebab Gagal Ginjal Akut Karena Obat Sirup," 24 Oktober 2022. [Online]. Available: <https://nasional.tempo.co/read/1648837/menkes-pastikan-penyebab-gagal-ginjal-akut-karena-obat-sirup>.
- [4] R. Salsabilla, "Pengumuman! Kemenkes Larang Apotek Jual Obat Sirup," 19 Oktober 2022. [Online]. Available: <https://www.cnbcindonesia.com/lifestyle/20221019101043-33-380821/pengumuman-kemenkes-larang-apotek-jual-obat-sirup>.
- [5] R. Fatmasari, V. M. Ayu, H. B. Pratama and W. Gata, "Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan," *Building of Informatics, Technology and Science (BITS)*, p. 395–404, 2022.
- [6] A. F. Pulungan and D. Selvida, "Kombinasi Metode Sampling pada Pengklasifikasian Data Tidak Seimbang Menggunakan Algoritma SVM," *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 6, no. 2, 2022.
- [7] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami and S. Sudioanto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube “Kekerasan Seksual”),” *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, 2022.
- [8] D. H. Qudsi and Yuliska, "Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Convolutional Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 5, pp. 1067-1076, 2019.
- [9] H. Jelodar, Y. Wang, R. Orji and S. Huang, "Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, 2020.
- [10] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," *Jurnal Informatika Universitas Pamulang* , vol. 5, pp. 544-552, 2020.
- [11] Sartini, "Analisis Sentimen Twitter Bahasa Indonesia Menggunakan Algoritma Convolutional Neural Network," *Unnes Repository*, Agustus 2020.
- [12] S. N. Listyarini and D. A. Anggoro, "Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan Convolution Neural Network (CNN)," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, vol. 1, no. 7, pp. 261-268, Juli 2021.
- [13] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, pp. 1018-1026, 2021.
- [14] M. Ihsan, B. S. Negara and S. Agustian, "Metode LSTM (Long short term memory) untuk Klasifikasi Sentimen Vaksin Covid-19 pada Twitter," *Jurnal Teknologi Informasi & Komunikasi Digital Zone*, vol. 13, pp. 79-89 , 2022.
- [15] F. Z. Ahmad, M. F. S. Arifandy, M. R. Caesarardhi and N. A. Rakhmawati, "Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat pada Media Sosial Youtube Menggunakan Algoritma Deep Learning Sekuensial dan LDA," *Jurnal Linguistik Komputasional*, vol. 4, no. 2, 2021.
- [16] R. D. Putra, "Klasifikasi Komentar Terhadap Polisi Pada Media Sosial Youtube Menggunakan Support Vector Machine," *repository*, Februari 2021.
- [17] T. M. K. J. M. Johnson, "Survey on deep learning with class imbalance," *Springer*, p. 7, 2019.

- [18] I. N. Husada and H. Toba, "Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, pp. 400-413, 2020.
- [19] J. Ipawati, Kusri and E. T. Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *IJNS - Indonesian Journal on Networking and Security*, vol. 6, no. 1, pp. 28-36, 2017.
- [20] Yuliska, D. H. Qudsi, J. H. Lubis, K. U. Syaliman and N. F. Najwa, "Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Convolutional Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 5, pp. 1067-1076, Oktober 2021.
- [21] D. I. Afida, D. S. F. Handayani and R. W. Pratiwi, "Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, 2021.
- [22] P. A. Riyantoko, T. M. Fahrudin, K. Maulida, Hindrayani and E. M. Safitr, "Analisis Prediksi Harga Saham Sektor Perbankan Menggunakan Algoritma Long-Short Terms Memory (LSTM)," *Seminar Nasional Informatika 2020 (SEMNASIF 2020)*, pp. 427-435, 2020.
- [23] M. Rivki and A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia," *Jurnal Sistem Informasi (Journal of Information Systems)*, vol. 1, no. 13, pp. 31-37, 2017.
- [24] K. Wabang, O. D. Nurhayati and F. , "Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 872-876, 2022.
- [25] U. D. Nuswantoro, "Pertemuan Ke 8," 2021. [Online]. Available: https://repository.dinus.ac.id/docs/ajar/pert_ke_7_sistem_cerdas.pdf.