JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 1, June 2025, pp. 362~375
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: 10.36378/jtos.v8i1.4450

A Deep Learning-Based Sentiment Classification for Identifying Advertorial Content in Online News

Brian Rizqi Paradisiaca Darnoto[1], Dony Bahtera Firmawan[1], Fahrobby Adnan[2]
[1] Informatics, Universitas Jember, Indonesia
[2] Information System, Universitas Jember, Indonesia

Article Info

Article history:

Received 06 01, 2025
Revised 06 10, 2025
Accepted 06 25, 2025

Keywords:

BiGRU
BiLSTM
CNN
Online News
Sentiment Classification

ABSTRACT

The rapid advancement of technology and the widespread use of the internet have brought significant positive and transformative impacts across various aspects of human life, including finance, healthcare, education, and the media industry. One notable consequence of information transparency is the vast availability and large-scale exchange of data. However, this also presents new challenges, particularly in the spread of misleading content such as disguised advertorials that resemble genuine news. This threatens the objectivity of the information received by the public. To address this issue, an automated solution is needed to identify the distinguishing characteristics of advertorials in online news content. This study proposes a deep learning approach using the Convolutional Neural Network (CNN) model to detect sentiment as an indicator of advertorial content. CNN is a widely used deep learning model for processing sequential and spatial data, capable of automatically learning features from text. The dataset comprises news articles categorized by advertorial traits, such as positive or neutral sentiment, persuasive language, and promotional content highlighting specific entities. The data undergo several processing stages, including text preprocessing, tokenization, padding, and CNN model training. Model performance is evaluated using accuracy, precision, recall, and F1-score. The experimental results show a validation accuracy of 84%, although overfitting issues were observed. Despite ongoing limitations, such as restricted data and suboptimal parameter tuning, the findings suggest that the CNN model has potential for automatically detecting advertorial content and can serve as a basis for future research using more advanced models and refined parameter adjustments.

Corresponding Author:

Brian Rizqi Paradisiaca Darnoto
Informatics
Universitas Jember
Jember, Indoneisa
Email: brianrizqi@unej.ac.id

1. Introduction

Initially, online news platforms in Indonesia served as replicas of print media and were not fully leveraged to mold public opinion or alter public perception [1]. Native advertising is a widely used form of online advertising with a style and functionality similar to the original content on online platforms, including news, sports, and social media sites [2]. Native ads are designed to blend in with journalistic content in news publications, making it challenging for average readers to distinguish them from standard reporting. This phenomenon prompts significant concerns regarding the objectivity of news reports, particularly when news coverage influences people's emotional perceptions of a product or establishment. Camelia notes that a key feature of native ads is that the news they contain is typically positive or neutral [3]. Identifying potential emotional bias in news reporting is crucial when analyzing sentiment in articles that include native advertising. This detection can facilitate transparency in information dissemination among the public, regulatory bodies, and media outlets.

The principal obstacle in researching native advertising is its structural and linguistic resemblance to editorial content, which hampers audiences' ability to discern sponsored material [4]. Traditional classification methods rely on set rules and are less accurate in identifying whether content is advertorial or non-advertorial. Research into trends based on news article content, particularly in boldness, sentiment, and events, is gaining momentum within the data mining and text analysis community [5]. Identifying emotional feelings in the text is a primary objective of natural language processing (NLP), specifically within sentiment analysis, encompassing positive, neutral, and negative emotional tendencies. Advanced deep learning techniques provide more adaptable and complex methods for identifying and capturing news articles' underlying semantic structures and emotional nuances. Specifically, models like BiLSTM, CNN, and BiGRU can acquire intricate context representations without manually engineered features. This functionality enables classification systems to improve their accuracy in identifying the emotional undertone of promotional and neutral news articles.

Past research has concentrated on identifying native advertising or categorizing news according to its content [6][7] and persuasive content [8]. Despite a lack of extensive research, few studies have explicitly examined the disparities in sentiment patterns between news articles that incorporate advertising and those that do not. In reality, sentiment in advertorials can be manipulated to create favorable impressions without allowing for a balanced presentation of information. The gap created provides an opportunity to investigate sentiment analysis techniques as indicators of the potential for advertorial content in online news sources. This study aims to address this requirement via a quantifiable and methodical computational methodology.

To resolve this problem, a comparison of several widely used deep learning models that have demonstrated effectiveness in text classification is undertaken, including BiLSTM, CNN, and BiGRU. The three models were selected because they exhibit distinct architectural features in processing sequential and spatial information and have been extensively utilized in NLP applications, including sentiment analysis. This aligns with prior research that examined several deep learning architectures, including CNN, LSTM, GRU, BiGRU, and BiLSTM, in the context of sentiment analysis using text [9].

The dataset employed comprises online Indonesian news articles, which have been classified as either advertorial or non-advertorial content and further divided into three categories based on sentiment: positive, neutral, and hostile. Comparing the performance of these models on sentiment classification is expected to identify the model that best suits the task of uncovering emotional inclinations in advertising news. This approach not only provides a technical solution but also enhances media literacy and the transparency of public information.

This analysis holds both practical and academic value, offering a foundation for developing automated systems capable of detecting advertising content based on emotional tone or sentiment. Such a system can act as an early warning tool for readers, helping them recognize news articles that may contain subtle promotional intent. From an ethical standpoint, the findings encourage media platforms to adopt more transparent and responsible methods when publishing advertorials, ensuring that audiences are not misled by hidden marketing messages. Academically, this study strengthens sentiment classification techniques by integrating the contextual understanding of advertorial content, marking a significant step toward understanding sentiment bias in digital journalism.

The research is positioned at the intersection of sentiment analysis, native advertisement identification, and the application of deep learning in natural language processing (NLP). It investigates the comparative performance of three deep learning models—BiLSTM (Bidirectional Long Short-Term Memory), CNN (Convolutional Neural Network), and BiGRU (Bidirectional Gated Recurrent Unit)—in detecting advertorial characteristics in online news. These models are evaluated using key performance metrics, including accuracy, precision, recall, and F1-score, to provide a balanced and comprehensive analysis.

Beyond proposing a technical solution, this study contributes to filling an academic gap by exploring the emotional influence of promotional content within online news platforms. It aims to uncover how emotional tone can serve as a clue to the presence of advertorials, thus reinforcing digital media literacy. The outcomes of this research are expected to advance the development of AI-driven tools for monitoring and analyzing digital content. Ultimately, this study offers a novel contribution by integrating sentiment analysis with deep learning approaches and comparing multiple model architectures to improve the detection of hidden advertising in news media.

2. Research Method

This research utilizes an annotated database comprising 12,088 entries from six online news websites in Indonesia. Experts have annotated the dataset to categorize sentiment as positive, neutral, or negative and to determine the occurrence of advertorial content within news articles. This study employs three deep learning techniques, namely BiLSTM, BiGRU, and CNN, to construct a text-based sentiment classification model. This study's research process encompasses eight stages: data preparation, data preprocessing, word embedding creation, the allocation of training and test data, model training, result examination, model performance assessment, and drawing conclusions.

The initial stage commences with gathering and preparing data, which is then followed by text pre-processing techniques, including case conversion, breaking down the text into individual words, stemming or reducing words to their base form, and removing common words like 'the' and 'and.' The cleaned data is subsequently transformed into word embeddings representing words as numerical vector forms. The data is then split into two categories - training data and test data - to assess the model's ability to generalize.

The model was next trained with three different deep learning architectures, BiLSTM, BiGRU, and CNN, utilizing the training data. The training outcomes were assessed by calculating accuracy, precision, recall, and F1-score evaluation metrics. This study aimed to identify the most effective model for categorizing opinions on news stories, considering whether or not they included promotional content. The development process used Python programming, with support from libraries such as TensorFlow, Keras, Scikit-learn, and NLTK. The last stage of the study involved drawing inferences from the outcome of the model evaluation and offering suggestions for future research. The flowchart presented in Figure 1 illustrates the various stages undertaken in this research.

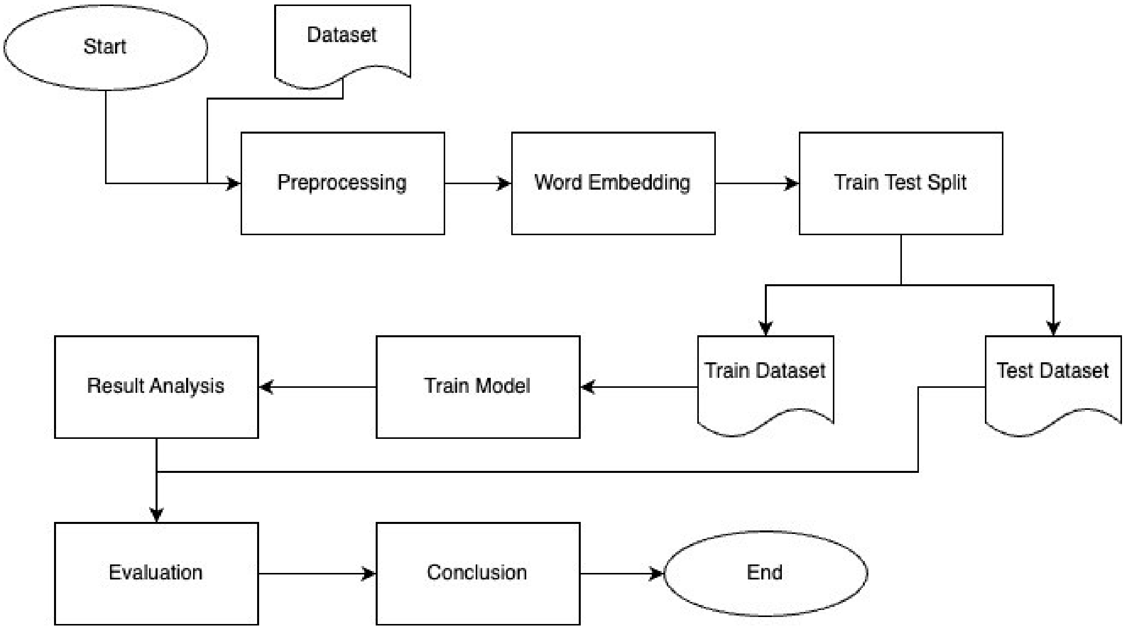

Figure 1 Research Methodology

2.1 Problem Identification

This study used an experimental stage to assess the ability of deep learning models to classify sentiments and effectively detect advertorial content in online news. This research focuses on a comparative analysis of three distinct model architectures: a BiLSTM model, a BiGRU model, and a CNN. The primary challenge is understanding how these models identify sentiment patterns, encompassing positive, neutral, and negative sentiments, which can indirectly indicate the presence of concealed advertising or promotional material within news articles. One difficulty is validating the model's ability to differentiate between advertorial and non-advertorial material, frequently employing comparable journalistic writing styles. This research aims to assess the effectiveness of the three deep learning models in providing a dependable method for identifying sentiment as a preliminary indicator of advertorial content in online news articles.

2.2 Literature Review

This study involves a literature review encompassing previous sentiment classification research, deep learning models like BiLSTM, BiGRU, and CNN, and methodologies for identifying advertorial content in online news articles. This study also explores the role of text representation methods and data annotation techniques in facilitating the classification process. This literature review stage aims to discover and clarify topics relevant to the current research, thereby refining the research direction and augmenting existing knowledge. This study also aids in grasping the benefits and drawbacks of the methods employed in prior studies, serving as a foundation for crafting more effective experiments. This research can yield new insights, thereby enhancing the development of more effective methods for sentiment classification within the context of advertorial content identification.

2.3 Data Analysis and Processing

2.3.1 Data Preparation

This study drew upon 12,088 news entries gathered from six prominent online news websites in Indonesia. News entries were randomly chosen based on various topics and the likelihood of advertorial material appearing. Experts in media and communication annotated news content to assign sentiment labels (positive, neutral, negative) and identify whether or not it contained hidden advertising elements. The annotation process is carried out by media and communication experts to provide sentiment labels (positive, neutral, negative) and identify the presence of advertising elements in news content. Sentiment labels are applied by considering that advertisements are not always negative sentiment but can also be neutral or positive depending on the context of the advertisement delivery. This annotation is designed to verify the accuracy and consistency of the training and test data that will be utilized in the sentiment classification model. The data obtained not only has contextually relevant information but also underwent a validation process conducted by experts to ensure the quality of the experiment.

2.3.2 Data Preprocessing

The pre-processing stage is carried out to clean and prepare text data so that it can be processed optimally by the deep learning model. The first step is case folding, which is changing all letters in the text to lowercase to reduce word duplication due to differences in capitalization. Next, punctuation is removed to remove symbols that are not needed in the analysis. After that, the text is broken down into word units using tokenization; then, the lemmatization process is applied to return the words to their basic form. Finally, stopwords, which are common words that do not have essential meanings in the context of classification, are removed, such as "yang," "dan," "adalah," and so on. This stage aims to simplify data representation and reduce noise affecting model performance.

2.3.3 Word Embedding

Word embedding is a distributed dense format and refers to a collection of language modeling and feature learning methods in NLP [10]. The text representation process in this study is undertaken by applying a word embedding method derived from Bidirectional Encoder Representations from Transformers (BERT). BERT is a language model developed by Google that has demonstrated exceptional performance in a range of NLP tasks [11]. Unlike traditional embedding techniques like Word2Vec or GloVe, which generate static representations for each word, BERT can produce contextual representations that allow the meaning of a word to change based on the sentence's context. BERT employs a transformer encoder architecture that enables the model to comprehend word relationships within a sentence in both directions (bidirectional). As a result, words with multiple possible interpretations can be more clearly understood. This investigation utilized the pre-trained IndoBERT model [12] , tailored explicitly for Indonesian text, which had previously undergone training on a substantial corpus of the Indonesian language. The primary objective of IndoBERT is to refine the semantic representation of online news texts utilized as research data.

The first step in processing the text involves tokenizing it using a BERT-compatible tokenizer, which splits the sentence into individual tokens the model can understand. The tokens are subsequently converted into fixed-dimensional vectors. The sentences in the dataset will be represented as sequences of high-dimensional vectors, capturing both the semantic and syntactic aspects of the text. As examined in this research, BERT's embedding vectors are subsequently utilized as input attributes in deep learning-based classification models, including BiLSTM, BiGRU, and CNN. Implementing BERT as the embedding stage will enhance the model's accuracy in sentiment classification and advertorial content detection within online news articles.

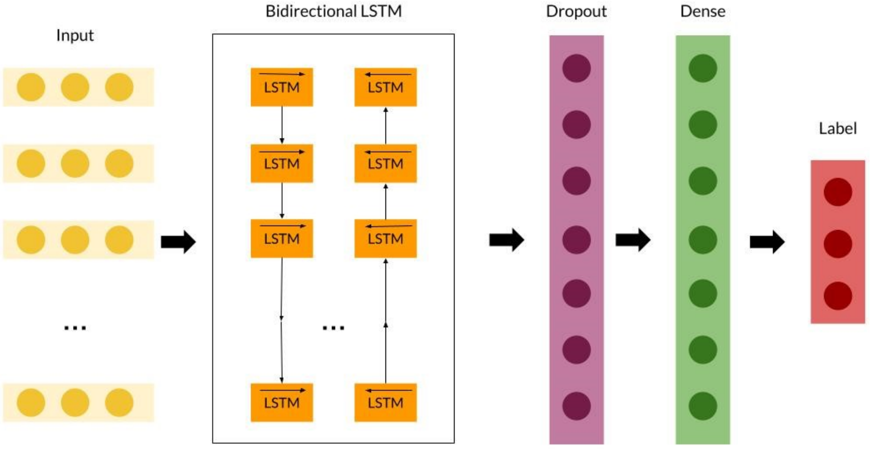

Figure 2 Architecture BiLSTM

2.3.4 BiLSTM

The model employed in this research is a Bidirectional Long Short-Term Memory (BiLSTM) network, a variant of the Long Short-Term Memory (LSTM) architecture. A key recurrent neural network (RNN) type is the Long Short-Term Memory (LSTM) architecture. It was specifically designed to handle sequential data and address the issue of vanishing gradients in standard RNNs. By contrast, BiLSTM incorporates a dual processing mechanism, processing information not only from the start to the end of a sequence (forward) but also from the end to the start (backward). Traditional RNNs' issue with vanishing gradients is addressed by incorporating gating units into the LSTM architecture, enabling LSTMs to more effectively identify and leverage long-range data dependencies and improve their capacity for capturing long-term relationships [13].

The BiLSTM architecture of this study is illustrated in Figure 2. Input data is converted into vector form via word embedding and fed into the BiLSTM layer. In this layer, two parallel LSTM paths are employed, one processing the word sequence in a forward direction and the other in a backward direction. The results from both sides are subsequently merged. The output of the BiLSTM is then passed to the dropout layer, which mitigates overfitting by temporarily turning off a specified number of neurons during the training phase. The output from the dropout is subsequently processed through a dense layer, also known as a fully connected layer, resulting in the final representation. The output from the dense layer is then passed on to the output layer, which generates the classification label.

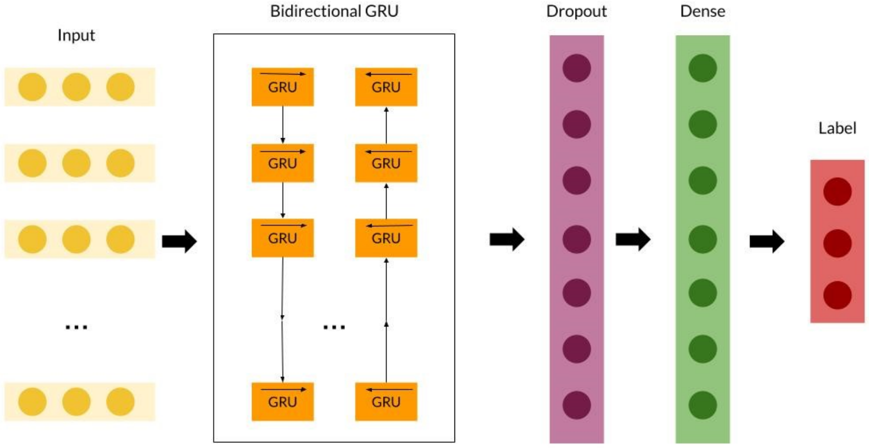

Figure 3 Architecture BiGRU

2.3.5 BiGRU

One of the recurrent neural network architectures employed in this research is the Bidirectional Gated Recurrent Unit (BiGRU) model, a substitute for BiLSTM. The GRU model is a variant of the LSTM architecture, where the forget gate and input gate are combined into a single update gate [14]. The performance of GRU is comparable to long-term and short-term memory networks, but it has fewer parameters and lower computational complexity [15].

The BiGRU model employs two identical GRUs that process data sequences in tandem, one proceeding in a forward direction and the other in a backward direction, enabling the model to incorporate contextual information from both perspectives. The architecture of the BiGRU model is illustrated in Figure 3. Input, a sequence of words embedded as vectors, is then sent to the BiGRU layer for processing. This layer comprises two GRU paths that operate in opposing directions. The outputs of the two GRU paths are then combined and sent to the dropout layer to mitigate overfitting during the training phase. Following the dropout layer, the output is fed into a dense layer, which produces the final feature representation through its fully connected nodes. The output layer then utilizes these representations to determine classification labels.

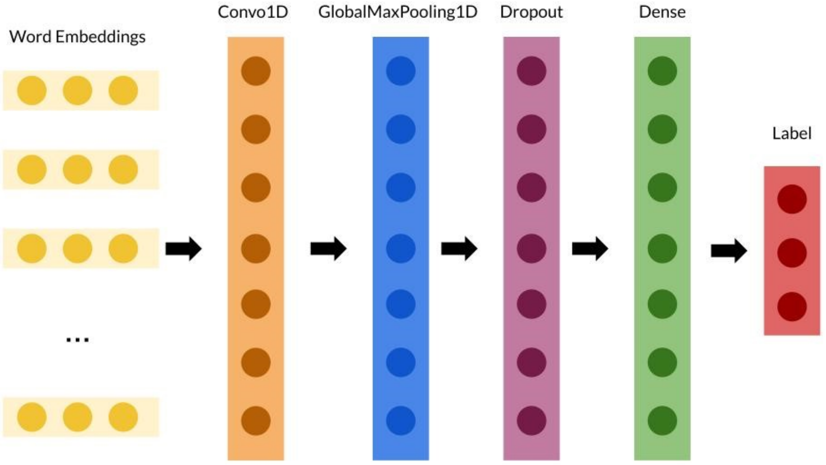

Figure 4 Architecture CNN

2.3.6 CNN

A Convolutional Neural Network is a deep-learning artificial neural network frequently employed for computer vision tasks or image classification [16]. Another approach used in this study for text-based classification tasks is the CNN model. In natural language processing, word embeddings are input into a one-dimensional convolutional neural network, replacing traditional image pixels, and this network can effectively detect and identify significant characteristics with high precision [17]. The CNN model architecture depicted in Figure 4 is initialized with an input represented by a word embedding of the text. The input sequence is then processed by a 1D Convolutional (Conv1D) layer, which extracts local spatial features from the input sequence. The filters in Conv1D move along the word sequence to identify key patterns crucial for the final prediction.

Following convolution, the output is fed into a 1D Global Max Pooling layer, whose purpose is to select the most significant features from each filter, achieved by identifying the highest value in each filter's output. This method helps decrease the complexity of the data, preserving the most significant details. This layer is succeeded by a dropout layer, which reduces overfitting during training by temporarily turning off several neurons at random. The dropout outcome is subsequently fed into a dense, fully connected layer to synthesize the features and generate the final representation. The final output layer generates the classification label.

2.3.7 Evaluation Measurement Tools

In classification research, performance evaluation tools are employed to assess the effectiveness of the classification model. Frequently employed evaluation measurement tools include the confusion matrix. A confusion matrix measures accuracy, precision, recall, and F-1 score performance. The precision of a system is determined by dividing the number of relevant documents retrieved by the total number of documents retrieved. The recall metric is the proportion of relevant documents retrieved again relative to the total number of relevant documents in the collection. Accuracy is calculated by dividing the total number of correct predictions (both positive and negative) by the overall size of the dataset. The minimum value limit for the three values is zero, and the maximum is 1. The confusion matrix for sentiment analysis in news classification comprises four key performance indicators that evaluate the model's accuracy in determining the sentiment of news articles.

a. True Positive (TP): News that has positive/neutral/negative sentiment and is successfully classified correctly by the model into the appropriate sentiment class.
b. False Positive (FP): News that is not actually included in a particular sentiment class, but is

incorrectly classified by the model as belonging to that class.

c. True Negative (TN): News that is not included in a particular sentiment class and is successfully not classified into that class by the model.
d. False Negative (FN): News that is actually included in a particular sentiment class, but fails to be classified by the model into that class.
e. The precision of a model is calculated as the True Positive (TP) rate compared to the total number of data points predicted to be positive [18]. The calculation of precision is based on the following formula:

$$precision = \frac{TP}{TP+FP} \ (1)$$

A high level of precision implies that most of the documents retrieved are relevant and valuable, resulting in a lower incidence of false positive results. While precision is an essential factor, it does not account for the number of relevant documents that are not retrieved. Recall is defined as the percentage of relevant documents that are correctly retrieved.

$$recall = \frac{TP}{TP+FN} \quad (2)$$

High recall means that the system retrieves most of the relevant documents but also some irrelevant ones. Recall is the proportion of positive examples correctly detected by the classifier [19]. To balance precision and recall, the F1-score is often used. It is the harmonic mean of precision and recall:

$$F1\ score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (3)$$

The F1-score combines the precision and recall values produced by the proposed framework by taking their harmonic mean values [20]. One key metric is Accuracy, which assesses the percentage of all correctly classified documents, including both relevant and irrelevant ones, out of the total number of documents reviewed.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

The closeness of a prediction to its actual value is a measure of accuracy, determined by dividing the number of accurate predictions by the total number of predictions made. The accuracy may be misleading in scenarios where the quantity of irrelevant documents greatly surpasses that of relevant documents.

3. Result and Discussion

3.1. Dataset

One of the essential components for enabling the analysis is the creation of comprehensive datasets. The objective is to gather at least 12,000 Indonesian-language electronic news articles, consistent with the quantity of data in the LIAR dataset [21], encompassing approximately 12,800 news items. However, we only used half of the dataset that we collected, which was 6044. The collected news must adhere to specific requirements: each item must comprise a minimum of 100 words and be comprised solely of written text, excluding any accompanying images or headlines. News that consists solely of images with no accompanying text will also be excluded from this dataset.

The study identifies key news websites, gathers articles, and examines the attributes of the collected dataset. This sub-chapter will showcase the findings from our investigation of news websites to compile the dataset essential for this research. The period under consideration spans from February to August 2022. The news categories included are Economy, Lifestyle, Entertainment, International, National, Others, Automotive, Education, Sports, and Technology. The data presented in Table 1 illustrates the distribution of categories for native ads and news labels.

Table 1 Total data per category

Category	Total Data
Economy	1914
Lifestyle	1807

Entertainment	635
International	359
National	1262
Others	3565
Automotive	474
Education	274
Sports	823

By selecting this news portal, the collected dataset includes relevant and diverse information related to the research topic. Table 2 explains the number of articles and other statistics that were successfully collected from the news portal.

Table 2 Statistics data from six news portals

Source	Total News	Average Words
News portal 1	1415	300
News portal 2	4026	280
News portal 3	1290	270
News portal 4	2384	310
News portal 5	2006	250
News portal 6	967	304

All data collected in this study are publicly available news articles from official online news portals and are used solely for research purposes. The content is accessed without bypassing any paywalls or violating platform terms of service. Therefore, no special licensing or permissions were required for data access. Regarding ethical considerations, this study does not involve any personal or sensitive information, nor does it include human subjects. However, the annotation process for labeling news items as ‘positive' or ‘neutral’ or 'negative' involved trained annotators. These annotators were informed about the research objectives and agreed to participate voluntarily without any coercion or conflict of interest. The annotation guidelines were designed to ensure consistency and reduce bias in labeling.

3.2. BiLSTM

After completing data preprocessing and transforming the textual data into word embeddings, a classification model is constructed using the Bidirectional Long Short-Term Memory (BiLSTM) architecture. BiLSTM is particularly effective for sentiment analysis tasks because it can capture contextual meaning from both preceding and succeeding words in a sequence, allowing for deeper understanding of sentence structure and sentiment cues. This bidirectional capability makes BiLSTM more accurate in interpreting the emotional tone of a sentence compared to traditional models.

The preprocessed dataset is then used to train the BiLSTM model with specific hyperparameters, including the number of neurons in the hidden layer, batch size, number of training epochs, and optimization using the Adam algorithm—a widely used optimizer for deep learning tasks due to its adaptive learning rate capabilities. The training configuration and parameter settings used in this model are detailed in Table 3.

Once trained, the performance of the BiLSTM model is evaluated using a separate test dataset to ensure generalizability. To assess the effectiveness of the model, it is compared against other deep learning models using evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics offer a comprehensive view of the model’s performance in identifying advertorial or sentiment-laden content within news articles.

The model encounters substantial overfitting, as the data in Figure 5 indicates. The model's training loss graph reveals a steep drop-off to almost zero, signifying its proficiency in mastering the training dataset. The validation loss demonstrates an upward trend as the number of epochs escalates, particularly exhibiting significant fluctuations around the 10th epoch. During training, the model experiences a decline in its capacity to generalize effectively to new, untested data sets. The training accuracy curve also mirrors this phenomenon, reaching 100% in a short period, whereas the validation accuracy plateaus around 81% and then unexpectedly drops off sharply at a particular point. The results demonstrate that although the BiLSTM model excels at identifying patterns in the training data, its performance on the validation data becomes increasingly unstable and tends to degrade after multiple epochs. Implementing a regularization method or an early stopping technique is crucial to combat this overfitting issue.

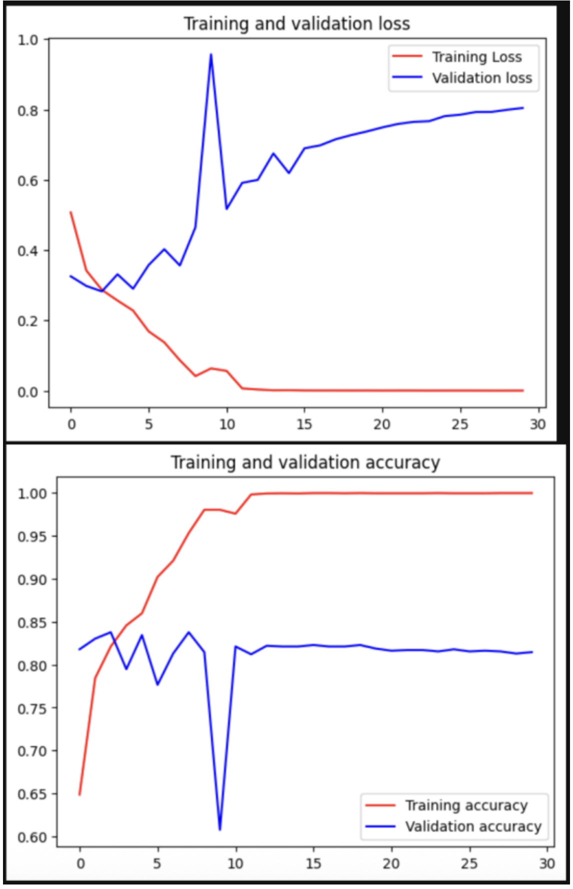

Figure 5 Accuracy and loss curves of the BiLSTM model

3.3. *BiGRU*

Compared to BiLSTM, the BiGRU model offers improved computational efficiency due to its more streamlined architecture while retaining the capacity to process sequential data and consider the two-way context. The training procedure for the BiGRU model is performed using the same dataset and data preprocessing protocol as that for the BiLSTM model. The BiGRU model is constructed with a comparable architecture to the BiLSTM model but incorporates GRU units, and training is conducted under similar conditions to ensure consistency in model comparison. The training parameters and evaluation results for the BiGRU model are presented in Table 3. To compare the performance of different methods fairly, the model's results are evaluated using the same metrics.

The BiGRU (Bidirectional Gated Recurrent Unit) model exhibits clear signs of overfitting, as reflected in the loss and accuracy graphs presented in Figure 6. This is evident from the training loss, which consistently decreases to nearly zero, while the training accuracy climbs sharply and reaches close to 100% by the 10th epoch. Such results indicate that the model is learning the training data too well, potentially memorizing patterns rather than learning to generalize.

However, the validation loss behaves differently—it begins to rise steadily after the early epochs, and the validation accuracy stagnates around 83%, showing no further improvement despite continued training. This widening gap between training and validation performance strongly suggests that the model is overfitting; it performs exceptionally well on the training data but struggles to maintain accuracy when applied to new, unseen data. This overfitting likely stems from the model’s complexity relative to the dataset's size or variability. The BiGRU architecture, while powerful, may be too sophisticated for the current dataset or may have been trained for too many epochs without sufficient regularization strategies in place.

To mitigate this, several corrective actions are recommended. Implementing early stopping can prevent unnecessary training once validation performance begins to degrade. Increasing the dropout rate can help reduce model complexity by randomly deactivating neurons during training, encouraging more robust feature learning. Furthermore, hyperparameter tuning—such as adjusting learning rates, batch sizes, or the number of hidden units—can enhance model performance and help the BiGRU generalize more effectively.

These adjustments are essential for developing a more balanced and reliable model capable of accurately detecting sentiment in diverse text data.

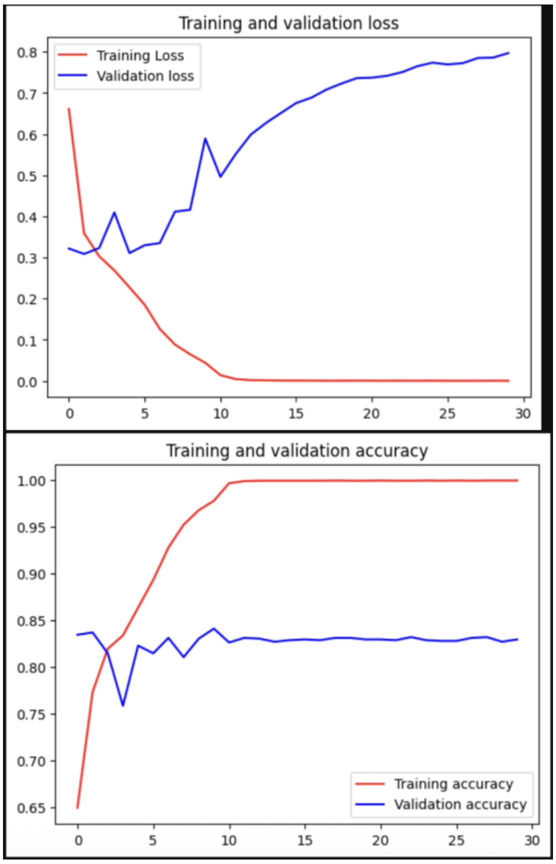

Figure 6 Accuracy and loss curves of the BiGRU model

3.4. CNN

The CNN (Convolutional Neural Network) model implemented in this study represents a recent approach in the field of text classification. Although CNNs are traditionally known for their success in image recognition tasks, they have also proven to be effective in analyzing textual data. By applying convolutional filters to sequences of embedded words, CNNs can identify local patterns and extract spatial features relevant to sentiment and advertorial content. This makes CNN a valuable model for tasks such as detecting emotional tone or promotional intent in news articles.

In this research, the CNN model is specifically designed to capture critical linguistic patterns associated with sentiment and subtle advertising cues. The architecture consists of convolutional layers to detect features, followed by pooling layers that reduce dimensionality and prevent overfitting, and fully connected dense layers that serve as the classifier. Before entering the CNN network, all text data is preprocessed and transformed into numerical word embeddings, enabling the model to process textual input effectively.

The experimental setup, including parameter configurations and evaluation metrics, is detailed in Table 3. These results are presented alongside those of the BiLSTM and BiGRU models to enable a comprehensive comparison. Overall, CNN contributes an efficient, lightweight solution for identifying key textual signals, and its performance in this context highlights its potential for future use in sentiment-based advertorial detection systems.

Examining the training and validation loss graphs in Figure 7 reveals that the model exhibits an overfitting effect. The model's training performance is reflected in the steady decline of the training loss graph to almost zero levels, suggesting that it acquires knowledge effectively from the provided training data. The validation loss does not exhibit a consistent trend, as it initially decreased at the start of the epoch but then began to increase after approximately the 6th to 8th epoch and continued to fluctuate without substantial improvement. After this point, the model's ability to make accurate predictions on unseen validation data is severely impaired. The training accuracy graph saw a considerable boost to 100% by the 15th epoch, and it remained stable at that level, which supports the notion that the model is overfitting to the training data. In contrast, validation accuracy did not see a substantial increase and instead plateaued at approximately 84% after experiencing initial fluctuations during training. A notable disparity exists between the model's training and validation results, indicating that the CNN model has achieved high accuracy in classifying sentiment on the training data yet has struggled to sustain its performance on unseen validation sets.

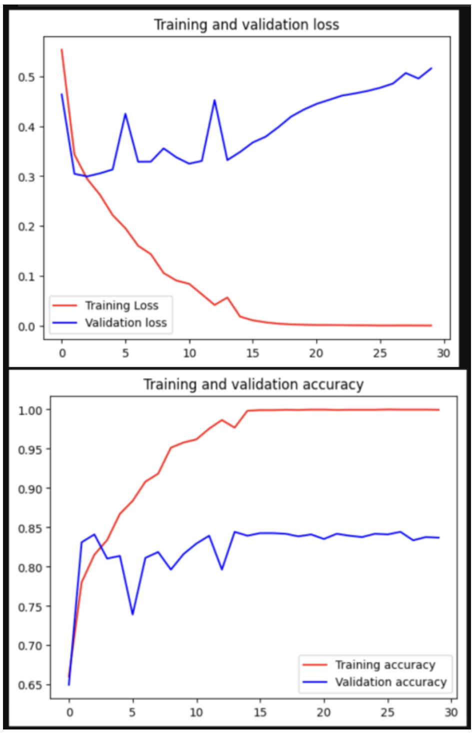

Figure 7 Accuracy and loss curves of the CNN model

3.5. Model Evaluation

This study's model was evaluated using various methods, including the confusion matrix and specific evaluation metrics - accuracy, precision, recall, and F1-score. These metrics are utilized to assess the model's performance in classifying data and the efficiency of the inference process in terms of time. Furthermore, a weighted average approach was employed, calculating the average evaluation metric based on the weight assigned to each class's data. This method yields more equitable outcomes, particularly when there is a disparity in the quantity of data between classes.

The models evaluated utilized a test dataset, and each model's performance was illustrated through a confusion matrix, which was subsequently assessed using the metrics previously described. Furthermore, the time it took for the model to process and classify data was calculated to evaluate its computational efficiency. The training and validation loss images, along with the training and validation accuracy for each model (BiLSTM, BiGRU, and CNN), are depicted in Figures 5 to 7, and the full results of the model evaluation metrics can be found in Table 3.

Table 3 Result of confusion matrix

Model	Accuracy	Precision	Recall	F1-score
BiLSTM	0.81	0.68	0.61	0.63
BiGRU	0.83	0.69	0.62	0.63
CNN	0.84	0.60	0.59	0.58

3.6. Regularization Strategy Using Class Weights and L2 Regularization

To reduce overfitting observed in earlier model evaluations, this study applies regularization techniques by adjusting class weights and using L2 regularization. The issue of class imbalance is handled by calculating class weights based on the inverse frequency of each class in the training set. This approach gives more emphasis to the minority class during training, helping the model become more sensitive to underrepresented examples. L2 regularization, also known as weight decay, is also used to control model complexity. This technique works by adding a penalty term to the loss function, which discourages large weight values. In this study, the L2 penalty is applied to key trainable layers including the Bidirectional LSTM, dense layers, and the output layer. A regularization value of λ = 0.001 is used. By limiting how large the weights can grow, L2 regularization helps the model generalize better and reduces overfitting.

The results of applying class weights and L2 regularization are shown in Figure 8. All three models—BiGRU, BiLSTM, and CNN—show more stable validation loss curves compared to previous training results. The gap between training and validation loss is noticeably smaller, which indicates that the models perform more consistently on both seen and unseen data. This improvement is especially clear in the BiGRU and BiLSTM models, where validation loss no longer rises sharply after a few epochs. The CNN model also benefits, with reduced fluctuations in validation loss across epochs. These results show that using class weights and L2 regularization successfully improves model robustness and generalization while addressing the overfitting issue found earlier.

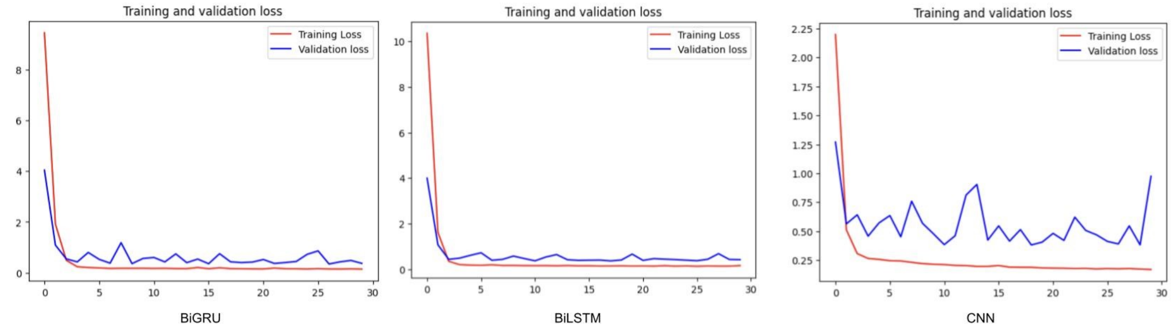

Figure 8 Training and validation loss after using regularization

4. Discussion

The evaluation results for three deep learning models applied in the sentiment classification task to detect advertorial content in online news are presented in Table 3. These models are BiLSTM, BiGRU, and CNN. The assessment was conducted based on the accuracy, precision, recall, and F1-score criteria. The chosen metrics are suited because they can capture the model's overall performance in binary classification. This task faces difficulties in dealing with label imbalance and complex linguistic context.

The CNN model achieved the highest accuracy of 0.84 was obtained on validation data, outpacing BiGRU with an accuracy of 0.83 and BiLSTM with 0.81. However, it should be noted that the highest F1-score values were obtained by BiLSTM and BiGRU, which reached 0.63, compared to CNN, which only reached 0.58. This higher F1-score indicates that the recurrent architecture-based models (BiLSTM and BiGRU) are better at balancing precision and recall, making them more reliable in classifying whether content contains advertorial elements.

The BiGRU model achieved a precision rate of 0.69, surpassing BiLSTM's rate of 0.68 by a small margin and substantially exceeding that of CNN at 0.60. According to the analysis, BiGRU performs better in pin-pointing advertorial content and yields fewer incorrect results. BiGRU's recall rate was 0.62, slightly higher than BiLSTM's 0.61, suggesting that the model is better at identifying authentic advertorial content.

While the CNN model is highly accurate, it still exhibits lower precision and recall rates. The results suggest that CNNs are less adept at incorporating sequential context, a characteristic prevalent in news articles. Instead, they tend to be more effective at processing spatial features. Still, they may fall short in capturing temporal context or word order, factors that are crucial for accurate sentiment classification in natural language.

These outcomes suggest that methods relying on recurrent neural network architectures, specifically BiGRU and BiLSTM models, are more efficient for sentiment analysis in online news, particularly in detecting advertorial content masked by neutral or positive language. The significance of employing architectures that can grasp the nuances of word order and sentence context is thus underscored.

Combining BiGRU architecture with a suitable activation function and loss function, such as the tanh-hinge or sigmoid-hinge combination, which has been demonstrated to yield the most effective results in prior studies, should be investigated to enhance the model's performance. Analysis of inference time is also crucial if this model is to be integrated into a real-time detection system on an online news platform.

To ensure long-term sustainability, the model should be periodically retrained using updated datasets from online news sources to adapt to changes in language, advertorial patterns, and content strategies. This process can be supported by automated data pipelines and model monitoring tools to detect performance drift and schedule retraining when needed. In addition, the model's relatively lightweight architecture allows for efficient retraining without requiring extensive computational resources.

For real-world implementation, the model can be deployed as a backend service in news platforms or editorial systems using REST APIs or inference servers such as TensorFlow Serving. It can operate in real-time or batch mode, depending on platform requirements. Integration with tools like Apache Kafka enables scalable processing, while a human-in-the-loop review mechanism can improve reliability and trust. These strategies provide a practical and adaptable framework for integrating the model into production environments.

5. Conclusion

This study uses an online news dataset categorized based on indicators of advertorial content, such as news sentiment, to develop a CNN-based method for automatically identifying advertorial content. The research process encompassed several stages, commencing with text preprocessing, tokenization, and pad-ding before training and evaluating the CNN model.

The study's findings indicate that the constructed CNN model effectively classifies sentiment, which, in turn, aids in identifying advertorial content. The training results graph shows that the model suffered from overfitting after several iterations: the training loss value continued to decrease, and its accuracy neared 100%, whereas the validation loss increased, and the validation accuracy tended to plateau at around 84%. The model's strong performance in identifying patterns in the training data does not necessarily translate to its ability to apply this knowledge to new, unseen data.

The study employed various modeling stages: data cleansing, tokenization, padding to standardize the input length, and training a CNN model using its initial parameters. The CNN model is utilized because it can extract spatial characteristics from text transformed into vector representations. The training process was conducted over 30 iterations, and the evaluation outcomes represented via loss and accuracy graphs indicated that the model was highly responsive to the training data.

The current investigation is hampered by several constraints, notably data scarcity and the lack of dropout or early termination methods, which can mitigate overfitting. Furthermore, the model has not been directly compared to other methods, including LSTM, BiLSTM, or transformer-based models like BERT, which can yield more effective outcomes in natural language processing.

Future research can be advanced by expanding the dataset, refining the model's structure, and evaluating the performance of CNNs against other deep-learning methodologies. Beyond single-label classification, a multi-label approach can also be employed, given that a single article may possess multiple characteristics. The CNN-based sentiment classification strategy examined here demonstrates potential as a starting point for automated advertorial content detection in online news, but further refinement is necessary to enhance the system's reliability and usability in real-world settings.

References

[1] P. Widjanarko and L. Hariyani, “Media Convergence-Deconvergence-Coexistence Triad in Indonesia: The Case of Liputan6.com,” *Jurnal ASPIKOM*, vol. 7, no. 2, 2022, doi: 10.24329/aspikom.v7i2.1134.

[2] N. Sahllal and E. M. Souidi, “A Comparative Analysis of Sampling Techniques for Click-Through Rate Prediction in Native Advertising,” *IEEE Access*, vol. 11, pp. 24511–24526, 2023, doi: 10.1109/ACCESS.2023.3255983.

[3] C. C. Pasandaran, “Political Advertising Camouflage As News,” *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia*, vol. 3, no. 2, Dec. 2018, doi: 10.25008/jkiski.v3i2.239.

[4] M. Carvajal and I. Barinagarrementeria, “The Creation of Branded Content Teams in Spanish News Organizations and Their Implications for Structures, Professional Roles and Ethics,” *Digital Journalism*, vol. 9, no. 7, 2021, doi: 10.1080/21670811.2021.1919535.

[5] K. Z. Ye, Y. M. Naing, Y. Naung, K. P. Nyein, and T. Z. Lin, “Implementation of Burmese Language News Classification System by Using SVM and LSTM Machine Learning Algorithm,” in *2023 IEEE 6th International Conference on Computer and Communication Engineering Technology, CCET 2023*, 2023. doi: 10.1109/CCET59170.2023.10335115.

[6] B. R. P. Darnoto, D. Siahaan, and D. Purwitasari, “A Comprehensive Ensemble Deep Learning Method for Identifying Native Advertising in News Articles,” in *8th International Conference on Software Engineering and Computer Systems, ICSECS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 164–169. doi: 10.1109/ICSECS58457.2023.10256392.

[7] B. R. P. Darnoto, D. Siahaan, and D. Purwitasari, "Deep Learning for Native Advertisement Detection in Electronic News: A Comparative Study," in *2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, 2022, pp. 304–309. doi: 10.1109/EECCIS54468.2022.9902953.

[8] B. R. P. Darnoto, D. Siahaan, and D. Purwitasari, "Automated Detection of Persuasive Content in Electronic News," *Informatics*, vol. 10, no. 4, Dec. 2023, doi: 10.3390/informatics10040086.

[9] N. Hicham, H. Nassera, and S. Karim, "Enhancing Arabic E-Commerce Review Sentiment Analysis Using a hybrid Deep Learning Model and FastText word embedding," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024, doi: 10.4108/eetiot.460

[10]. V. F. Pacheco, Microservice Patterns and Best Practices: Explore Patterns Like CQRS and Event Sourcing to Create Scalable, Maintainable, and Testable Microservices, Packt Publishing, 2018.

[11]. M. Fowler, "Microservices," MartinFowler.com, 2014. [Online]. Available: https://martinfowler.com/articles/microservices.html

[12]. S. Newman, Building Microservices, O'Reilly Media, 2021.

[13]. Nofri Wandi Al-Hafiz, Helpi Nopriandi, and Harianja, "Design of Rainfall Intensity Measuring Instrument Using IoT-Based Microcontroller", JTOS, vol. 7, no. 2, pp. 202 - 211, Dec. 2024.

[14]. N. W. Al-Hafiz and H. Harianja, "Design of an Internet of Things-Based automatic cat feeding control device (IoT)", Mandiri, vol. 13, no. 1, pp. 161–169, Jul. 2024.

[15]. PutriD. and Al-HafizN., "SISTEM INFORMASI SURAT KETERANGAN GANTI RUGI TANAH PADA KECAMATAN KUANTAN TENGAH MENGGUNAKAN WEBGIS", Biner : Jurnal Ilmiah Informatika dan Komputer, vol. 2, no. 2, pp. 112-121, Jul. 2023.

[16]. H. Harianja, N. W. Al-Hafiz, and J. Jasri, "Data Analysis of Informatics Engineering Students of Islamic University of Kuantan Singingi", JTOS, vol. 6, no. 1, pp. 23 - 30, Jan. 2023.

[17]. Siregar, M., and N. Al-Hafiz. "Design of Cloud Computer to Support Independent Information System Servers Universitas Islam Kuantan Singingi." Journal of Information System Research (JOSH) 3.2 (2022)

[18]. Apri Denta, H. Nopriandi, and E. Erlinda, "Information System Design Realization and Performance Achievements of the Manpower Office of Kuantan Singingi District", JTOS, vol. 7, no. 1, pp. 01 - 09, Nov. 2024.

[19]. AP Putra, E. Erlinda, and M. Yusfahmi, "Sales System in the Endocell Mobile Phone Business Using the CRM (Customer Relationship Management) Method) in Kompe Berangin Village, Cerenti District", JTOS, vol. 7, no. 1, pp. 10 - 21, Nov. 2024.

[20]. D. Setiawan, F. Haswan, and J. Jasri, "Design of School Bell Scheduling Application Based on Arduino Uno on MTs Babussalam Simandolak", JTOS, vol. 7, no. 1, pp. 22 - 30, Jul. 2024.

[21]. D. Juniarti, A. Aprizal, and S. Chairani, "Information System for Analyzing Disease Trends in the Region Cerenti Health UPTD", JTOS, vol. 7, no. 1, pp. 31 - 43, Jun. 2024.

[22]. F. Restuadi, H. Nopriandi, and A. Aprizal, "ANALISIS QOS JARINGAN INTERNET FAKULTAS TEKNIK UNIVERSITAS ISLAM KUANTAN SINGINGI MENGGUNAKAN WIRESHARK 4.0.3", JTOS, vol. 7, no. 1, pp. 44 - 54, Jun. 2024.

[23]. J. Jasri and N. W. Al-hafiz, "Designing a mobile-based infaq application markazul quran wassunnah foundation (MQS)Kuantan Singingi", J. Teknik Informatika CIT Medicom, vol. 15, no. 5, pp. 247–254, Nov. 2023.

[24]. R. Nazli, A. Amrizal, H. Hendra, and S. Syukriadi, "Modeling User Interface Design E-Business Applications for Marketing Umkm Products in Payakumbuh City Using Pieces Framework", JTOS, vol. 7, no. 2, pp. 55 - 66, Nov. 2024.

[25]. Siti Saniah and Mhd. Furqan, "Classification Of Rice Plant Diseases Using K-Nearest Neighbor Algorithm Based On Hue Saturation Value Color Extraction And Gray Level Co-Occurrence Matrix Features", JTOS, vol. 7, no. 2, pp. 212 - 223, Dec. 2024.