Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Data tweet clustering using bidirectional gated recurrent unit and k-prototype for the Indonesian political year

Ibnu Try Rosadi[a], Anindya Apriliyanti Pravitasari[b*] and Yudhie Andriyana[b]

[a]*Post-Graduate Program in Applied Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*
[b]*Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Padjadjaran, Sumedang 45363, Indonesia*

CHRONICLE

Article history:
Received: November 10, 2023
Received in revised format: November 20, 2023
Accepted: December 15, 2023
Available online: December 15, 2023

Keywords:
Election
Fake News
Hate Speech
BiGRU
K-Prototype

ABSTRACT

As time passes, social media, which was formerly used as a means of communication between users, is experiencing a transition as a means for broadcasting information, conducting business, advertising, and even political campaigning. In elections, social media is also used to discredit political opponents to reduce the electability of opposing candidate. Spreading hate speech and fake news to undermine the electability of opposing candidate is a common violation of the law committed by supporters of one candidate over another. Considering that the number of social media users increases annually at a very rapid rate, the hazard of social media abuse has the potential to grow. In 2022, Indonesia had 191 million social media users in January 2022. Obviously, this will make the election situation more tumultuous and has the potential to cause societal divisions. The government must have a control system in place to screen social media content that can be considered illegal. In this study, fake news and hate speech are classified using the Bidirectional Gated Recurrent Unit (BiGRU). Lastly, K-Prototype was used to do clustering based on categorization dimensions and probable distribution to identify which clusters had the greatest risk of breaking the law, creating confusion, and dispersing broadly throughout society. It is hoped that the clusters that are created will represent the levels of priority of tweet data that requires prompt attention from the government to prevent it from spreading and inciting social unrest. Based on the results of the analysis, the BiGRU fake news model yields a F1-score of 95%, while the BiGRU hate speech model yields a F1-score of 90%. Clustering data using K-Prototype in this research can reduce the number of tweet data from 13,183 to 1,791 data. These new data are considered as a priority that must be pursued in preventing social media disputes.

1. Introduction

The development of Internet technology has instigated a paradigm shift in societal dynamics, wherein the essence of daily existence has transitioned from tangible, real-world engagements to virtual activities within the realm of cyberspace. The utilization of cyberspace bears multifaceted consequences, both positively and negatively. On one hand, it facilitates the expeditious and precise dissemination of information; however, on the other hand, it can be used for illegal activities such as the dissemination of radical ideology, pornography, drug trafficking, and organised crime. These adverse elements pose a potential threat to national resilience and the cohesive unity of the Republic of Indonesia's Unitary State (Amilin, 2019; Ananda, 2022).

The threat of social media misuse has the potential to increase, given that the number of social media users grows annually at an astounding rate (Birjali et al., 2021). According to research conducted by “We are social” and “Hootsuite” in 2022, Indonesia had 191 million social media users in January 2022. Compared to 2021, this number has increased by 21 million social

* Corresponding author.
E-mail address: anindya.apriliyanti@unpad.ac.id (A. A. Pravitasari)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.12.015

media users (We are social, 2022). As the number of social media consumers increases, the process of validating information is increasingly neglected. Ethical disorientation is the primary cause of the rapid spread of fake news in Indonesia (Susanti, 2022). In addition, the concept of anonymity in social media is one of the factors contributing to the growth of fake news (Masrudi, 2019; Cho et al., 2014; Saksesi et al., 2018). The shield of anonymity empowers social media users to express themselves freely, without the apprehension of revealing their identities, even if their posts bear the potential to transgress legal boundaries.

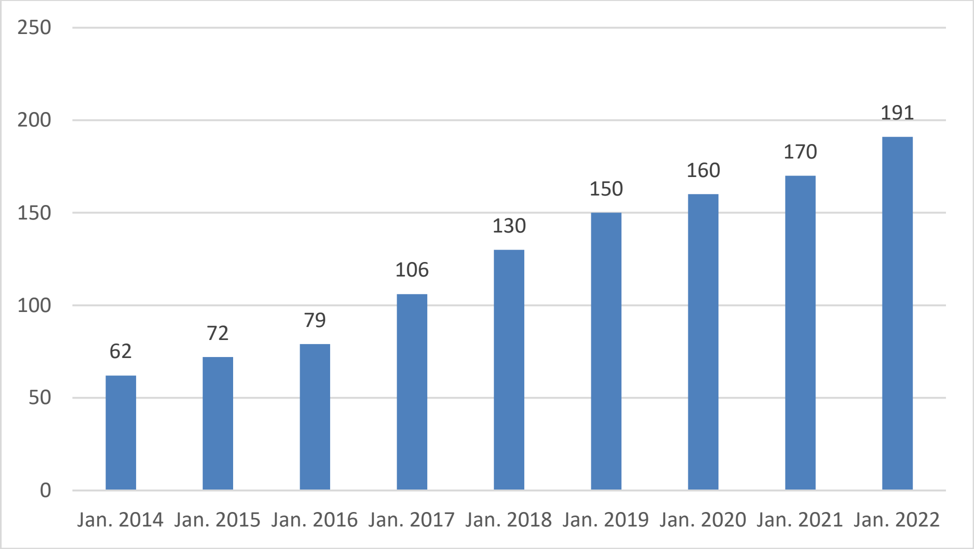


Fig. 1. The growth of social media user numbers between 2014 and 2022

According to Fig. 1, the number of social media users prior to the implementation of the Regional Head Election and Presidential Election from 2017 to 2019 increased significantly. This is certainly an illustration of the increase in social media consumers preceding the 2024 elections. In the post-truth era, and particularly as the political year begins, the spreading of fake news has become a very dangerous issue in Indonesian national and social life. Exploiting issues related to ethnicity, religion, ancestry, and social groups not only undermines national resilience but also poses a risk of national disintegration and jeopardises the Republic of Indonesia's integrity (Amilin, 2019; Chung et al., 2014).

As the 2024 Indonesian election approaches, the government is confronted with the pervasive dissemination of fake news and hate speech on social media platforms. Fake news (also known as junk news, pseudo-news, or hoax news) is a form of news consisting of deliberate disinformation or hoaxes spread via traditional news media (print and broadcast) or online social media (Mohsin, 2020). Hate speech, on the other hand, refers to actions by individuals or groups involving instigation, provocation, or insults directed towards others, spanning various dimensions such as ethnicity, religion, race, gender, skin colour, disability, sexual orientation, and related factors (Febriansyah & Purwinarto, 2020; Jeong, 2023). This poses a significant challenge for the Indonesian government to filter out this content to prevent it from spreading further. The government needs to quickly decide which content needs to be reported to social media platform providers for takedown. The utilization of deep learning models can be used to rapidly classify whether a piece of content falls under the categories of hoaxes or hate speech or not. Many previous studies have utilized machine learning and deep learning as classification methods to detect fake news and hate speech (Hao et al., 2016; Liao, 2005).

Triyono et al. (2023) compared five machine learning methods used to detect fake news: Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), and Random Forest (RF). Based on their research, the lowest accuracy was 75.33% for the DTC method, and the highest accuracy was 83.55% for the SVM method. Alameri (2021) compared machine learning methods: Naïve Bayes (NB), Support Vector Machine (SVM), and deep learning methods: Long Short-Term Memory (LSTM), Neural Network with Keras (NN-Keras), and Neural Network with TensorFlow (NN-TF) to detect fake news. The results of the study indicated that deep learning models achieved higher accuracy compared to machine learning models.

Sentiment analysis conducted by Omara et al. (2022) showed that BiGRU-based methods had better accuracy compared to LSTM and CNN-based methods. AitechTrend (2023) stated that based on several empirical studies conducted in speech recognition, language modeling, and sentiment analysis, LSTM and GRU outperformed each other, with their performance depending on the task and data used. Therefore, this research used the BiGRU algorithm as a method for detecting fake news and hate speech because it has a simpler structure than BiLSTM, making the computation process more efficient. This is important since the implementation of the model is expected to be used for real-time streaming data from Twitter.

Previous research has mostly focused on demonstrating the accuracy levels of classification models. In practical application, if a significant number of social media posts are classified as fake news, it becomes a significant task for the government to follow up on all of these posts. Additionally, the government may be perceived as limiting freedom of expression if too many posts are taken down on social media. Therefore, this research is conducted to further filter the classification results of the fake news model into smaller groups. The grouping is done by considering the results of the hate speech model classification and the potential spread of these posts.

The contributions of this research are as follows:

- This research not only demonstrates the accuracy level of models based on training and testing data but also shows how these models perform when implemented with actual Twitter data.
- By incorporating both supervised and unsupervised learning methods, the research innovatively reclassifies tweets initially labeled as fake news and hate speech into new categories, considering priority scale and urgency level. This nuanced approach adds depth and context to the classification process.
- This research demonstrates how to reduce a large amount of data into smaller groups to facilitate the filtering of content classified as fake news and hate speech.

2. Materials and Methods

2.1 Data Source

Data source used in this research is employed for a two-stage analysis, namely the modeling stage and the implementation stage for clustering. During the phase of developing the fake news classification model, this study utilised a total of 5,384 Indonesian language datasets acquired from three distinct sources, which are:

- Cekfakta.com Dataset (4,284 data points). Acquired from cekfakta.com, this dataset spans the period from February 2018 to June 2022. This dataset was categorised into 2,094 fake news data and 2,190 actual news data (Sadida, 2023).

- Mendeley Dataset (600 data points), was obtained from the website https://data.mendeley.com/datasets/p3hfgr5j3m/1. These data were categorised into 372 real news and 228 fake news (Nayoga et al., 2021).

- GitHub Repository Dataset (500 data points), was obtained from the online repository located at https://github.com/pierobeat/Hoax-News-Classification. This dataset was evenly split into two subsets, with one subset containing 250 real news and the other subset containing 250 fake news (Nayoga et al., 2021).

During the hate speech classification model development phase, this study utilised a dataset consisting of 13,169 instances in the Indonesian language. These instances were further categorised into 5,561 instances labelled as hate speech and 7,608 instances labelled as non-hate speech. The data categorised as hate speech is categorised into three distinct levels: weak, moderate, and strong. The weak category consists of 2,283 data points, the moderate category has 1,705 data points, and the strong hate speech category consists of 473 data points. The categorization of hate speech levels is determined through the analysis of data obtained from Focus Group Discussions and talks with personnel from the Cyber Crime Directorate of the Criminal Investigation Agency of the Indonesian National Police (Bareskrim Polri), as well as linguistic specialists (Ibrohim & Budi, 2019).

Subsequently, the entirety of the news will be partitioned into two distinct subsets, which are the training data and the testing data. In accordance with established empirical studies, the recommended data distribution for training and testing purposes involves allocating 80% of the data for training and reserving the remaining 20% for testing (Gholamy et al., 2018). The training data is utilised to train the model that will be constructed, whereas the testing data is employed to assess the model's performance by examining its accuracy.

Furthermore, an additional analysis is conducted involving clustering, employing a total of 13,183 tweets obtained through scraping activities using the keyword "pilpres" during the period from March 23rd to March 26th, 2023.

2.2 Experimental Setting

This research is generally divided into four stages: data preprocessing, model formation, model implementation, and clustering. As visualized in Figure 2, the initial step involves pre-processing to ready the data for model development. In pre-processing, special consideration is given to word embedding, a method that imparts information on the structure, sequence, semantics, and context surrounding words. The model development in this study adopts the deep learning architecture of Bi-GRU. The resulting model is applied to Twitter data for the prediction of hate speech and fake news. The predictions serve as variables for clustering Twitter data using the K-prototype method. The methods underlying the analysis process are comprehensively discussed in sub-chapters 2.3 to 2.6.

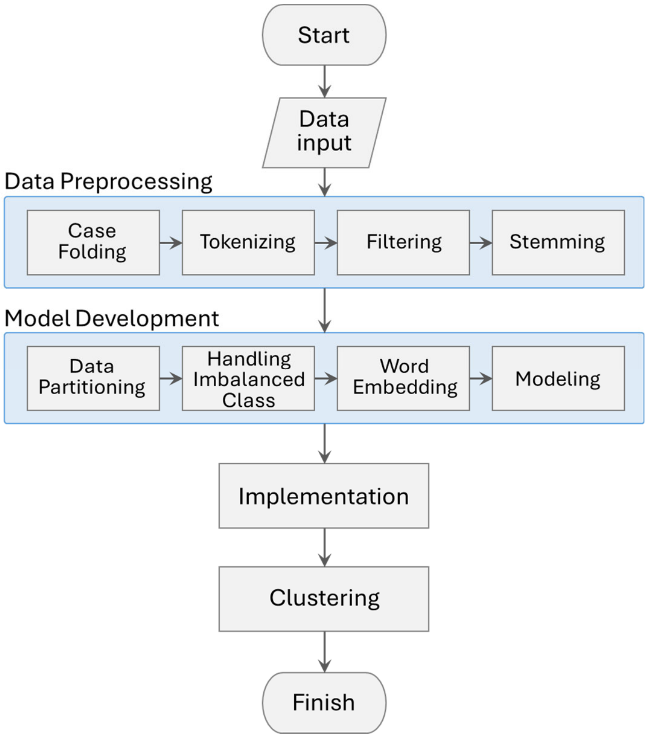


Fig. 2. Experimental Setting

2.3 Pre-processing

Data preprocessing is performed to ensure that the data used is free of noise, has lower dimensions, and is more structured, allowing for further processing. Data preprocessing encompasses various phases, including case folding, tokenization, filtering/stop word removal, and stemming. In word processing, the use of word embeddings is also taken into consideration.

Word embedding is a computational procedure that involves the conversion of words represented in alphanumeric format into vectorized form. The word2vec algorithm will be employed as the word embedding technique in this study. The Word2vec model employs a neural network architecture to obtain vector representations. The architecture of word2vec comprises three distinct layers, specifically the input layer, the projection layer (also known as the hidden layer), and the output layer. There exist two distinct types of architecture inside the word2vec framework, namely skip-gram and continuous Bag of Word (CBOW). The comparison between two models is as visualized as Fig. 3. The skip-gram architecture employs a one-hot encoded vector as input data and an n-hot encoded vector as output data. To clarify, the skip-gram model employs a single word as its input and aims to predict a set of n words in the vicinity of the input. In contrast, the CBOW model utilizes input data represented as an n-hot encoded vector, while the output data is represented as a one-hot encoded vector. The CBOW model utilizes a set of n words as input to make predictions on a single word from within the input set. Hence, the skip-gram design of word2vec is deemed appropriate for implementation in this study.

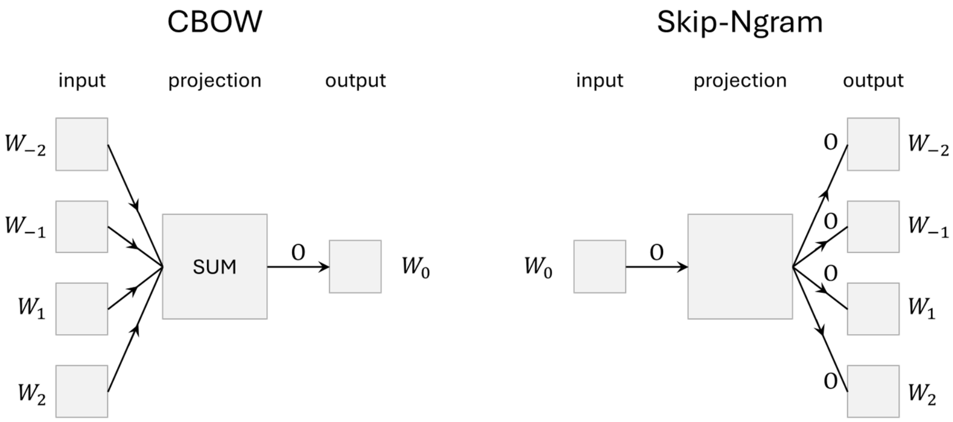


Fig. 3. Comparison of CBOW and skip-gram architectures

2.4 Bidirectional Gated Recurrent Unit (BiGRU)

In this study, we employ BIGRU for to create a classification model for fake news and hate speech. The Bidirectional Gated Recurrent Unit (BiGRU) is a neural network architecture that processes data in a sequential process, capturing information in both the forward and backward directions (Yu et al., 2019). The architecture of Bi-GRU describe in Fig. 4.

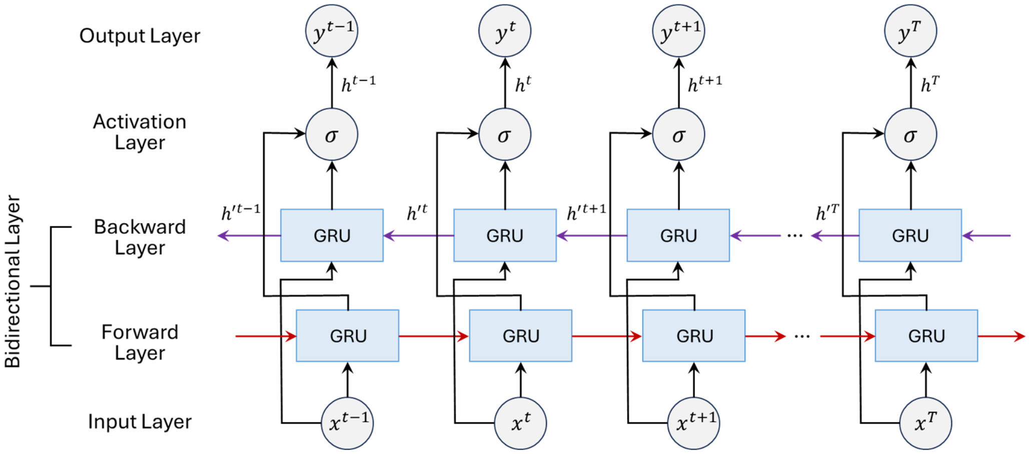


Fig. 4. Bidirectional Gated Recurrent Unit (BiGRU) Architecture

The forward layer of the BiGRU model effectively captures sequential data in the following sequence, meanwhile, the backward layer in BiGRU sequentially gathers data from the previous sequence.

$$\overrightarrow{h_t} = \sigma(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}) \quad (1)$$
$$\overleftarrow{h_t} = \sigma(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \quad (2)$$
$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \quad (3)$$

The stages of the BiGRU process involve sequential input $(x_1, x_2, x_3, \dots, x_t)$, which result in the generation of forward hidden states $(\overrightarrow{h_1}, \overrightarrow{h_2}, \overrightarrow{h_3}, \dots, \overrightarrow{h_t})$ and backward hidden states $(\overleftarrow{h_1}, \overleftarrow{h_2}, \overleftarrow{h_3}, \dots, \overleftarrow{h_t})$. The final hidden state$(h_t)$ is obtained by applying the element-wise addition operation ⊕ to both the forward BiGRU output $\overrightarrow{h_t}$ and the backward BiGRU output $\overleftarrow{h_t}$.

The accuracy of deep learning predictions is significantly influenced by hyperparameters. The task of identifying the optimal hyperparameters is commonly referred to as hyperparameter optimization or hyperparameter tuning (Wu, 2019). This research will employ an automated hyperparameter search technique utilizing the random search approach. The subsequent hyperparameters employed in this study are as follows:

- Dropout
  The dropout technique is employed in the training phase to prevent overfitting by randomly deactivating units inside the hidden layer. This research will employ a dropout rate of 0.2 for the purpose of model building.
- Batch Size
  The process of training a deep learning model must be carried out effectively. Hence, it is imperative to partition a substantial quantity of datasets into many batches of reduced dimensions. The batch size refers to the quantity of data samples that are processed simultaneously within a neural network. This research will employ batch sizes of 16, 32, 48, and 64 for the purpose of evaluation.
- Epoch
  The epoch is a parameter utilized to quantify the number of complete iterations that a deep learning algorithm performs on the full dataset. An epoch is considered complete when the entire batch has been successfully processed once by the neural network.
- Optimizer
  The optimizer is utilized to minimize the loss function, which is used as a measure of the difference between the predicted output and the actual output. The methodology employed in this study will involve the utilization of the Adamax optimization algorithm.

2.5 K-Prototype

In this research, we utilize K-Prototype to cluster tweet data based on variables such as fake news classification, hate speech classification, number of followers, number of retweets, and hashtag usage. The K-Prototype method was chosen because it has the capability to handle heterogeneous data that includes both numeric and categorical attributes. Unlike the K-Means algorithm, which is limited to processing numeric data, and the K-Modes algorithm, which only applies to categorical data. The K-Prototype method possesses an advantageous characteristic in that its algorithmic complexity is not too complex, rendering it capable of effectively managing substantial datasets (Huang, 1997).

K-Prototype procedure will consist of the following steps:

(1) Determine the cluster centroids of $k < n$ $(C_1, C_2, \ldots, C_k)$, where n is representing the number of observation data.

(2) Calculate the distance and similarity of data points to the cluster centroid. Then, cluster the data according to its proximity to the centroid.

$$d(X,Y) = \sum_{j=1}^{p} (x_{jn} - y_{jn})^2 + \gamma \sum_{j=P+1}^{m} d\left(x_{jc}, y_{jc}\right) \tag{4}$$

where,

$d(X,Y)$: The proximity distance between objects X and Y

p : The number of numeric variables

m : The number of categorical variables

x_{jn} : The j-th numerical variable of object x

y_{jn} : The j-th numerical variable of object y

x_{jc} : The j-th categorical variable of object x

y_{jc} : The j-th categorical variable of object y

$d\left(x_{jc}, y_{jc}\right)$: The distance between objects x and y of categorical variable j, where

$$d\left(x_{jc}, y_{jc}\right) = \begin{cases} 0, & x_{jc} = y_{jc} \\ 1, & x_{jc} \neq y_{jc} \end{cases}$$

γ : The weights of category attributes obtained by calculating the average standard deviation of the numerical variables

(3) After all data has been entered into the clusters, the new centroid should be recalculated. Then, regroup the data into new centroids. In the case of numeric data, the determination of a new centroid is achieved by computing the mean value of the numeric variable values inside each cluster.

$$C_{in}^{(t+1)} = \frac{1}{\left|K_i^{(t)}\right|} \sum_{x_{jn} \in K_i^{(t)}} x_{jn} \tag{5}$$

(4) In the context of categorical type data, the updated centroid is determined by calculating the mode of the categorical variable values inside each cluster.

$$C_{ic}^{(t+1)} = mod\left(x_{jc}\right), \qquad x_{jc} \in K_i^{(t)} \tag{6}$$

where,

$C_{in}^{(t+1)}$: Centroid of the i-th numeric variable at time $t+1$

$C_{ic}^{(t+1)}$: Centroid of the i-th categorical variable at time $t+1$

$K_i^{(t)}$: The i-th cluster at time t

(5) Continue to execute procedures (2) and (3) iteratively until there is no further alteration observed in either the centroid or the members. The iteration is halted and the process is considered to have converged if no changes are seen.

2.6 Model Evaluation

The evaluation of the model should be conducted subsequent to its development. This process is conducted to evaluate the efficacy of the developed model in accurately classifying data. The confusion matrix is a tool that can be employed to assess the performance of machine learning models. The confusion matrix facilitates the comparison of classification outcomes generated by the model with the true classification outcomes. This research used a binary class confusion matrix to classify instances of fake news, whereas a multiclass confusion matrix is utilized for the categorization of hate speech. Table 1 and Table 2 show the confusion matrix for binary class and multi-class (four classes) respectively.

Table 1
Confusion Matrix for binary class

Predicted Values	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	TP (True Positive)	FP (False Positive) Type I Error
Negative (0)	FN (False Negative) Type II Error	TN (True Negative)

Table 2
Confusion Matrix for four classes

Predicted Values	Actual Values			
	a	***b***	***c***	***d***
a	T_{aa}	F_{ab}	F_{ac}	F_{ad}
b	F_{ba}	T_{bb}	F_{bc}	F_{bd}
c	F_{ca}	F_{cb}	T_{cc}	F_{cd}
d	F_{da}	F_{db}	F_{dc}	T_{dd}

This research employs three calculation metrics that can be conducted with the confusion matrix, specifically: $Precision$, $Recall$, and $f1-score$.

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

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$f1-score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \tag{9}$$

3. Results

3.1 Data Preprocessing

3.1.1. Case Folding

The process of case folding is employed in order to standardize the characters inside the dataset. This process involves converting all letters to lowercase. The characters from A to Z that are present in the data will be transformed into lowercase characters from a to z. Table 3 is an example of a news story that has endured a process of case folding.

Table 3
Results of the Case Folding Process

News	Case Folding Output
Vote changes when the KPU count reaches 100%. Coincidentally, vote number 3 was reduced by 7000 votes. Meanwhile, vote no.2 increased by 50 votes.	vote changes when the kpu count reaches 100%. coincidentally, vote number 3 was reduced by 7000 votes. meanwhile, vote no.2 increased by 50 votes.

3.1.2. Tokenizing

Following the completion of the case folding process in the preceding stage, the subsequent step involves the tokenization process, wherein all sentences are transformed into individual words, commonly referred to as tokens. In addition, numerals, punctuation marks, and symbols are eliminated. The results of the tokenizing procedure are illustrated in Table 4.

Table 4
Results of the Tokenizing Process

Case Folding Output	Tokenizing Output
vote changes when the kpu count reaches 100%. coincidentally, vote number 3 was reduced by 7000 votes. meanwhile, vote no.2 increased by 50 votes.	‘vote’, ‘changes’, ‘when’, ‘the’, ‘kpu’, ‘count’, ‘reaches’, ‘coincidentally’, ‘vote’, ‘number’, ‘was’, ‘reduced’, ‘by’, ‘votes’, ‘meanwhile’, ‘vote’, ‘no’, ‘increased’, ‘by’, ‘votes’

3.1.3. Filtering

Subsequently, a filtering process is conducted to extract significant or pertinent terms from the token derived from the tokenization procedure. In the current phase, stop words are employed to eliminate words that often occur but lack relevance within a phrase. The utilization of this filtering stage can prove to be advantageous in terms of minimizing the computational time required for modeling. The results of the filtering procedure are illustrated in Table 5.

Table 5
Results of the *Filtering* Process

Tokenizing Output	Filtering Output
‘vote’, ‘changes’, ‘when’, ‘the’, ‘kpu’, ‘count’, ‘reaches’, ‘coincidentally’, ‘vote’, ‘number’, ‘was’, ‘reduced’, ‘by’, ‘votes’, ‘meanwhile’, ‘vote’, ‘no’, ‘increased’, ‘by’, ‘votes’	‘vote’, ‘changes’, ‘kpu’, ‘count’, ‘reaches’, ‘coincidentally’, ‘vote’, ‘number’, ‘reduced’, ‘votes’, ‘meanwhile’, ‘vote’, ‘no’, ‘increased’, ‘votes’

3.1.4. Stemming

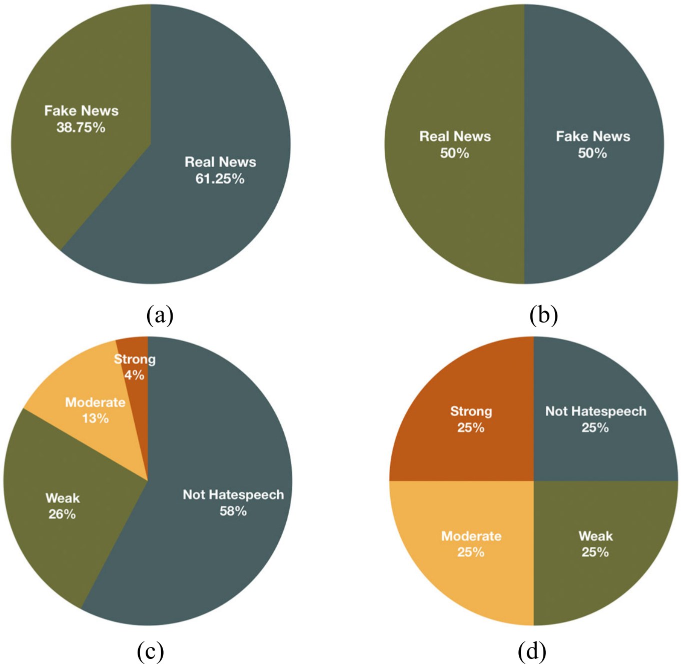
During this phase, words or tokens are transformed into their fundamental word forms. This process eliminates prefixes, suffixes, or both from a given word. Table 6 presents an illustrative depiction of the outcomes derived from the application of the stemming procedure.

Table 6
Results of the Stemming *Process*

Filtering Output	Stemming Output
'vote', 'changes', 'kpu', 'count', 'reaches', 'coincidentally', 'vote', 'number', 'reduced', 'votes', 'meanwhile', 'vote', 'no', 'increased', 'votes'	'vote', 'chang', 'kpu', 'count', 'reach', 'coincidental', 'vote', 'number', 'reduce', 'vote', 'meanwhile', 'vote', 'no', 'increase', 'vote'

3.2 Handling Imbalanced Class

Models constructed with imbalanced data can lead to reduced accuracy in predicting outcomes for the minority class. This study aims to handle the issue of imbalanced classes by employing the random over sampling technique. Random oversampling is a resampling approach wherein more sample data is added to the minority class without adding variations to the class data.



(a) (b) (c) (d)

Fig. 5. Proportion of dataset labels before and after random oversampling:
(a) fake news dataset before oversampling
(b) fake news dataset before oversampling
(c) hate speech dataset before oversampling
(d) hate speech dataset after oversampling

3.3 Word Embedding

The process of converting textual data into vector representation is performed during this stage, enabling its utilization as input during the modeling phase. Initially, a distinct index value will be assigned to each individual word, with the sorting criterion being the word's frequency. The dataset utilized in this study for fake news has a total of 28,456 distinct words, while the dataset used for hate speech contains a total of 18,860 distinct words..

Furthermore, it is known that the longest sentence in the used dataset contains 923 words in the fake news dataset and 71 words in the hate speech dataset. In order to ensure data consistency, it is imperative to perform a padding procedure wherein a value of zero is appended to news articles that include fewer than 923 words in the fake news dataset and fewer than 71 words in the hate speech dataset. The subsequent step involves the creation of a word embedding matrix using the distinct word index values that have been previously established. The utilized methodology involves the application of a Fasttext pretrained model with a dimensionality of 100.

3.4 Model Development

The proposed study would employ a network architecture consisting of five (5) distinct layers, namely the embedding layer, BiGRU layer, dropout layer, dense layer, and output layer, as shown in Fig. 6.

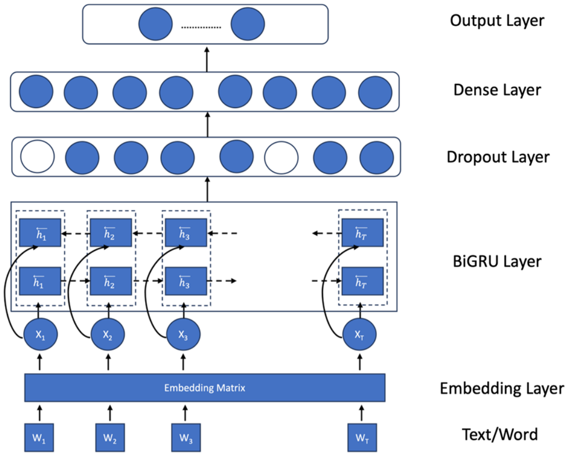


Fig. 6. Design of Architectural Modeling

(1) The modeling process starts with a 100-dimension Word2vec embedding layer. The embedding matrix generated within this layer will be used as the input for the subsequent BiGRU layer.

(2) Next, the BiGRU layer is utilized to analyze patterns within the dataset by considering both the sequential forward and backward aspect. The quantity of neurons within this layer has been confirmed to be 8.

(3) Thirdly, the use of the dropout layer provides the purpose of reducing overfitting in the model. The dropout rate is established at 0.2.

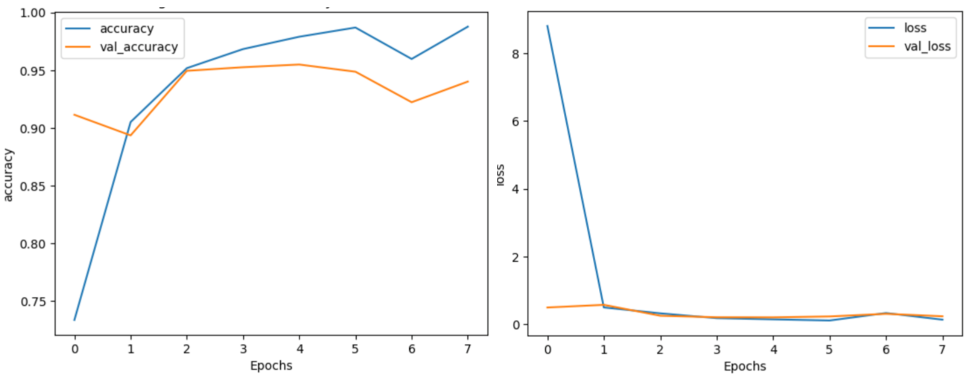
(4) Furthermore, the utilization of the Rectified Linear Unit (ReLU) activation function will be employed in the dense layer to establish non-linearity inside the model.

(5) In the end, the softmax activation function is employed in the output layer to provide probability values across different classes, hence facilitating the process of categorization. In addition to this, the model employs the Adam optimizer with a learning rate chosen via random search, and utilizes the categorical crossentropy loss function.

The process of training the model involves the utilization of hyperparameters such as batch size, number of neurons, learning rate, and dropout. Table 7 presents the optimal hyperparameter values for the models employed in the detection of fake news and hate speech. Fig. 8 displays the graphical representation of the accuracy and loss value during the training and validation process.

Tabel 7
The Optimal Hyperparameter Values for The Models

Model	Hyperparameter			
	Batch Size	**Learning Rate**	**Dense Dim**	**Dropout**
Fake News	48	0.001	8	0.2
Hate speech	16	0.01	8	0.2



(a) (b)

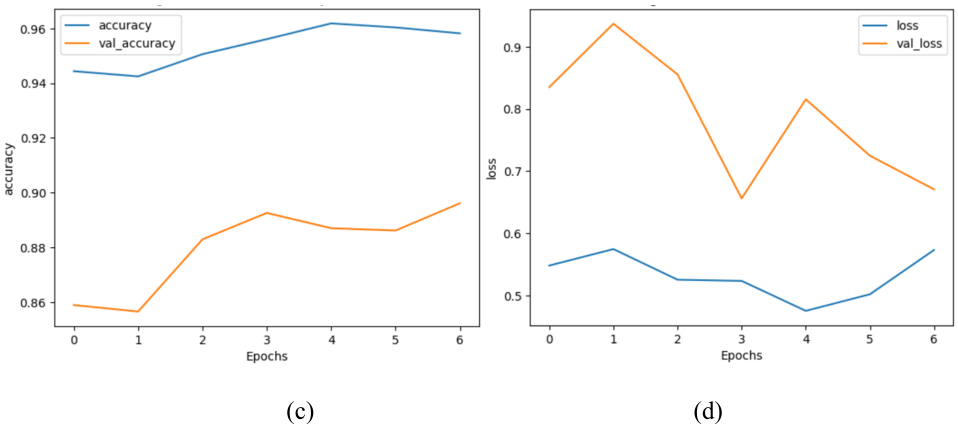


Fig. 7. The accuracy and loss value during the training and validation process:
(a) the accuracy of fake news model
(b) loss value of fake news model
(c) the accuracy of hate speech model
(d) loss value of hate speech model

The data depicted in Fig. 7 indicates a noticeable trend wherein the accuracy of the model consistently enhances with an increase in the number of iterations (epochs), ultimately reaching a point of convergence. This pattern implies that a higher number of iterations correlates with improved predictive performance of the model. Concurrently, the loss graph illustrates a reduction in the loss value as the epoch count increases, culminating in convergence at a specific point. Consequently, the model developed in this study exhibits commendable proficiency in generating accurate predictions.

3.5 Model Evaluation

The final stage that needs to be done after the model is formed is to carry out an evaluation using a confusion matrix. The confusion matrix shows the comparison between actual data and predicted data. The confusion matrix results can be seen in Table 8 and Table 9**,** respectively**.**

Table 8
Confussion Matrix Model Fake News

Predicted Values	**Actual Values**	
	Valid (0)	**Fake News (1)**
Valid (0)	604	39
Fake News (1)	19	624

Table 9
Confussion Matrix Model Hate speech

Predicted Values	**Actual Values**			
	Weak (0)	**Not Hate speech (1)**	**Moderate (2)**	**Strong (3)**
Weak (0)	1,302	115	46	15
Not Hate speech (1)	253	1,147	64	14
Moderate (2)	76	14	1,371	17
Strong (3)	0	0	0	1,478

Following this, to assess the effectiveness of the generated model, precision, recall, and f1-score metrics are calculated, and the results are detailed in Table 10 and Tabel 11.

Tabel 10
Calculation of Performance Levels of Fake News Model

Classification	**Precision**	**Recall**	**F1-Score**
Valid	97%	94%	95%
Fake News	94%	97%	96%
Overall	96%	95%	95%

Tabel 11
Calculation of Performance Levels of Hate speech Model

Classification	**Precision**	**Recall**	**F1-Score**
Weak *(0)*	80%	88%	84%
Not Hate speech	90%	78%	83%
Moderate (2)	93%	93%	93%
Strong (3)	97%	100%	98%
Overall	90%	90%	90%

From the information provided in Table 10 and Table 11, it is evident that the fake news model attains an impressive F1-score of 95%, whereas the hate speech model achieves a commendable F1-score of 90% overall. This showcases the robust predictive capabilities of the model when applied to unclassified data.

3.6 Implementation and *Clustering*

The model for detecting fake news and hate speech was subsequently applied to a dataset consisting of 13,183 tweets that had the phrase "pilpres" (presidential election) over the time frame of March 23 to March 26, 2023. The data obtained from applying the model to the twitter data included in this study is presented in Table 12.

Table 12
Model Implementation Result

Obs	Tweet	Fake News Classification	Hate speech Classification
1	The influence of Mr. SBY (Susilo Bambang Yudhoyono) in the 2024 presidential election is not particularly strong. Of course, this might not be able to boost his son, AHY (Agus Harimurti Yudhoyono), who will face difficulties competing against @erickthohir as a vice presidential candidate later.	Fake News	Weak
2	Absolutely unbeatable in debates! It's difficult to defeat any survey institution that only has a few thousand respondents. Our polling results show that Anies is certain to win the 2024 presidential election with the support of over 140,000 respondents!	Fake News	Not Hate speech
…			
13,182	With high electability, Ahy is considered the key to victory in the 2024 presidential election #ahypolitikusmendunia	Fake News	Not Hate speech
13,183	Buzzers in despair, Ahy emerges as the key figure for winning the 2024 presidential election #ahypimpinperubahan	Fake News	Not Hate speech

Fig. 8. shows the proportion of outcomes obtained from the application of the fake news and hate speech model on the dataset of 13,183 tweets. The findings of the classification indicate that the distribution of each label tends to be rather balanced, with no statistically significant differences seen.

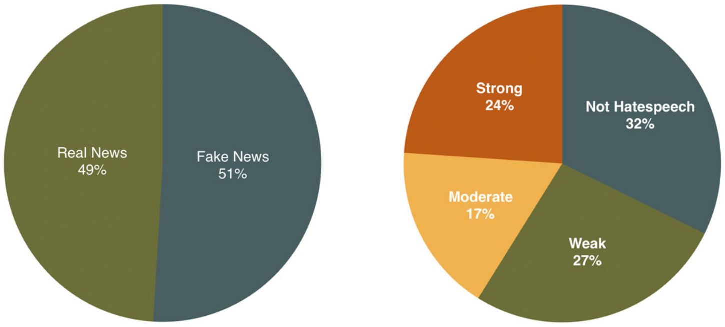


Fig. 8. Proportion of tweet data labeling results

Table 13
Contingency Table of The Result

Fake News Classification	Hate speech Classification			
	Weak	Not Hate speech	Moderate	Strong
Fake News	1,458	1,936	1,516	1,806
Real News	2,052	2,317	749	1,348

Subsequently, the dataset consisting of 13,183 tweets was categorized based on five variables: fake news classification, hate speech classification, number of followers, number of retweets, and hashtag usage. The application of the K-Prototype approach for clustering twitter data relies on the utilization of mixed type variables. The process of clustering is exclusively used to tweets that have been categorized as both fake news and hate speech inside the strong classification. This approach serves the purpose of narrowing down the range of tweets that require monitoring and attention. The silhouette approach is employed to determine the optimal number of clusters that can be produced. The findings derived from the silhouette analysis indicate that the optimal number of clusters that can be generated is three. The boxplot of cluster representation and the cluster profile are sequentially illustrated in Fig. 9 and Table 14, respectively.

Table 14
Clusters Profile

Cluster	Number of Cluster Members	Average Number of Followers	Average Number of Retweet	Percentage Members Using Hashtag
1	1,791	3,834	76.5	7%
2	12	250,421	32.7	12%
3	3	3,800,918	0.3	3%

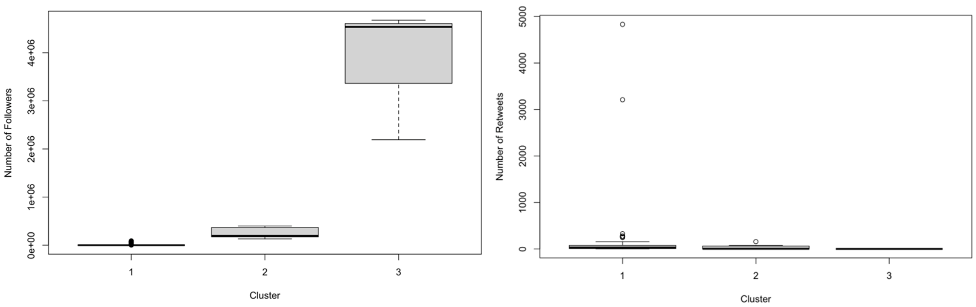


Fig. 9. Boxplot Representing the Clustering Results

4. Discussion

The Model produced by BiGRU yields excellent results, with an F1-score of 95% for the fake news model and an F1-score of 90% for the hate speech model. In forming this model, additional analysis is carried out, including data balancing, as the dataset falls into the category of an imbalanced dataset. The optimal BiGRU model is then used to predict tweet data, resulting in each label having proportions that tend to be balanced and not significantly different. Subsequently, clustering is performed using K-Prototype based on variables such as fake news classification, hate speech classification, number of followers, number of retweets, and hashtag usage. The findings from the categorization of tweet data, as presented in Table 14, indicate an opposite relationship between the mean number of followers and the mean number of retweets. Cluster 1 exhibits the lowest mean follower count, although displays the highest mean retweet count in comparison to the remaining clusters. This is due to the fact that the tweet data used in this study has a tendency to indicate that accounts with a large number of followers receive a small amount of retweets. Based on the utilized data, it is recommended to prioritize and closely monitor cluster 1 in order to mitigate the occurrence of divides and disagreements on social media platforms. This is shown by the greatest average retweet rate in cluster 1 and the second-highest hashtag usage rate. It is widely recognized that hashtags on the social media platform Twitter serve as a strategic instrument for facilitating the dissemination of various concerns, enabling them to gain traction and achieve trending status.

Clustering may need to be applied to larger data sets and incorporate additional keywords so that the characteristics of the tweet data are more comprehensive. According to the tweet data used in this study, accounts with a large number of followers tend to receive a small number of retweets, and vice versa. The objective of the cluster analysis is to group tweet data that exhibits characteristics such as being classed as fake news, containing strong hate speech, having a substantial number of followers and retweets, and utilizing hashtags in these tweets.

5. Conclusions

The fake news model was developed using 5,384 data, which were divided into 2,572 fake news data and 2,812 real news data, resulting in a F1-score of 95%. In the meantime, the hate speech model was developed using 13,169 data, which were divided into 8,708 not-hate-speech data, 2,283 weak data, 1,705 moderate data, and 473 strong hate-speech data, yielding a F1-score of 90%. This indicates that the model exhibits strong predictive capabilities. The successful use of the K-Prototype algorithm in this study resulted in a significant reduction in the volume of tweet data, from an initial count of 13,183 to 1,791. This outcome has considerable importance and warrants further attention, as it can contribute to the mitigation of divisions and conflicts on social media platforms.

Acknowledgements

The author extends sincere gratitude to the Department of Statistics Universitas Padjadjaran and the Research Center for AI and Big Data (AIDA) Universitas Padjadjaran for their support in completing this research.

References

Aitechtrend.com. (2023). LSTM vs GRU Which One Is Better for Recurrent Neural Networks. Access on 10 Juli 2023, from https://aitechtrend.com/lstm-vs-gru-which-one-is-better-for-recurrent-neural-networks.

Alameri., M. Mohd. (2021). Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques. *3rd International Cyber Resilience Conference (CRC)*, pp. 1–6, doi: 10.1109/CRC50527.2021.9392458.

Amilin. (2019). The Impact of Political Fake News in the Post-Truth Era on National Resilience and Its Implications for Sustainable National Development (published in Bahasa). *Jurnal Kajian Lemhanas RI, 39*, 5-11.

Ananda, D. (2022). Sentiment Analysis of Social Media Users Towards Government Policies in the Covid-19 Vaccination Program in Indonesia Using the Bidirectional Gated Recurrent Unit (BiGRU) Method (published in Bahasa). Skripsi Sarjana, Universitas Padjadjaran.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Datareportal. Digital 2022:Indonesia. Access on March 14th 2023, from https://datareportal.com/reports/digital-2022-indonesia.

Dewanpers. Berita Dewan Pers ETIKA. Access on April 19th 2023, from https://dewanpers.or.id/assets/ebook/buletin/646-AGUSTUS%202017.pdf.

Febriansyah, F., & Purwinarto, H. (2020). Criminal Liability For Hate Speech Actors In Social Media (published in Bahasa). *Jurnal Penelitian Hukum De Jure, 20*, 177-188.

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation.

Hao, X., Zhang, G., & Ma, S. (2016). Deep learning. *International Journal of Semantic Computing*, *10*(03), 417-439.

Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. *Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore: World Scientific 1997a*, pp. 21–34.

Ibrohim, M. O., & Budi, I. (2019, August). Multi-label hate speech and abusive language detection in Indonesian Twitter. In *Proceedings of the third workshop on abusive language online* (pp. 46-57).

Jeong, H. (2023). Hate Speech, Subject Agency and Performativity of Bodies. The Criticism and Theory Society of Korea. *Criticism and Theory, 28*(1), 271-313. 10.19116/theory.2023.28.1.271.

Kumparan.com Lebih dari 1,1 Juta Tweet Ramaikan Debat Keempat Pilpres 2019. Access on March 14th 2023, from https://kumparan.com/kumparantech/lebih-dari-1-1-juta-tweet-ramaikan-debat-keempat-pilpres-2019-1554031898897425099/4.

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, *38*(11), 1857-1874.

Masrudi. (2019). Hoaxes, New Media, and Our Literacy Power (published in Bahasa). *Orasi Jurnal Dakwah dan Komunikasi, 10*(2).

Mohsin, K. (2020). Defining 'Fake News'. *SSRN Electronic Journal*. doi:10.2139/ssrn.3675768.

Nayoga, B. P., Adipradana, R., Suryadi, R., & Suhartono, D. (2021). Hoax analyzer for Indonesian news using deep learning models. *Procedia Computer Science*, *179*, 704-712.

Omara, E., Mousa, M., & Ismail, N. (2022). Character gated recurrent neural networks for Arabic sentiment analysis. Sci Rep. National Library of Medicine.

Sadida, H. Q. (2023). Classification of Indonesian Political Hoax News Using the Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) Methods (published in Bahasa). Skripsi Sarjana, Padjadjaran University.

Saksesi, A. S., Nasrun, M., & Setianingsih, C. (2018, December). Analysis text of hate speech detection using recurrent neural network. In *2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC)* (pp. 242-248). IEEE.

Susanti, Indah., & Nurmiati. (2022). Mitigating the Impact of Social Media Hoaxes to Achieve National Unity (published in Bahasa). Ahmad Dahlan Legal Perpective, 2, 153-168.

Triyono, L., Gernowo, R., Prayitno, P., Rahaman, M., & Yudantoro, T. R. (2023). Fake News Detection in Indonesian Popular News Portal Using Machine Learning For Visual Impairment. *JOIV: International Journal on Informatics Visualization*, *7*(3), 726-732.

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.

Yu, Q., Zhao, H., & Wang, Z. (2019, August). Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. In *Proceedings of the 2nd international conference on artificial intelligence and pattern recognition* (pp. 116-119).