

Classification of Review Text using Hybrid Convolutional Neural Network and Gated Recurrent Unit Methods

Fiqih Fathor Rachim¹, Auli Damayanti^{1,*} & Edi Winarko¹

¹Mathematics Department, Faculty of Science and Technology, Universitas Airlangga, Indonesia

*Corresponding author: auli-d@fst.unair.ac.id

Abstract. Consumer reviews are opinions from buyers to sellers based on service satisfaction or product quality. The more consumer reviews cause the process of analyzing manually will be difficult. Therefore, an automated sentiment analysis system is needed. Each review will be grouped into a sentiment class which is divided into positive and negative classes. This study aims to classify review texts using the Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) methods. The research stages in this study include collecting data on Tokopedia review texts, extracting hidden information from review texts using CNN, conducting learning on review texts using GRU. A total of 1000 review texts were divided into 80% training data and 20% test data. The review text is converted into matrix using One Hot Encoding algorithm and then extracted using CNN. The CNN process includes the convolution calculation, the calculation of the Rectified Linear Unit (ReLU) activation function, and the pooling stage. The extraction results in the CNN process are continued in the GRU process. The GRU process includes initializing parameters, GRU feed forward, Cross-Entropy Error calculation, GRU feedback, and updating weights and biases. The optimal weight is obtained when the error value in the training is less than the expected minimum error or the training iteration has reached the specified maximum iteration. Optimal weight is used for validation test on test data. The implementation of review text classification using the hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) method was made using the python programming language. The accuracy of the validation test is 88.5%.

Keywords: *artificial neural network, convolutional neural network, gated recurrent unit, sentiment classification.*

1 Introduction

Advances in technology have developed various businesses, including buying and selling online or e-commerce. With a population of more than 269 million, 133 million connected to the internet and more than 70 million smartphone users, Indonesia is indeed the largest market share in Southeast Asia for e-commerce users [1]. In e-commerce transactions, buyers cannot touch, try, or see directly the product to be purchased. Buyers can only rely on product descriptions and reviews. This review will play an important role for both the seller and the buyer. According to data from Bright Local in its 2016 report Local Consumer Review Survey, around 84% of people trust reviews from other customers [2].

Consumer reviews are opinions from buyers given to sellers based on service satisfaction and quality of goods purchased. These consumer reviews can influence a person's decision to buy goods or services use services. Consumer reviews can contain positive or negative sentiment [3]. Consumer reviews can be used by sellers to improve the quality of goods and services. With the large number of consumer reviews, the manual analysis process will make it difficult for sellers to know the opinions of their consumers. Therefore, it takes a sentiment analysis system from consumer reviews automatically [4].

Sentiment analysis is the process of extracting and analyzing text documents to obtain information with a specific purpose. Each review will be grouped into a sentiment class. Sentiment class is divided into positive and negative classes so that users can read and choose the opinion they want. Positive sentiment indicates consumer satisfaction with the product or service. Meanwhile, negative sentiment indicates consumer disappointment with the product or service [5]. One of the methods used to analyze sentiment is Convolutional Neural Network (CNN).

In previous studies, Convolutional Neural Network (CNN) has been successfully used in heart rate classification, sentence modeling, and semantic representation learning [6],[7],[8]. Convolutional Neural Network (CNN) has the advantage of detecting and extracting important information automatically or without human supervision [9]. This method also does not require knowledge of the syntactic and semantic structure of a language. If this method works only by using characters in the text representation, the problem of word writing errors can still be studied [10]. However, the Convolutional Neural Network method is not able to study the relationship of several words or sentences [11]. This problem can be solved by the Gated Recurrent Unit (GRU) method.

The Gated Recurrent Unit (GRU) introduced by Cho is a method that studies text by paying attention to the relationship of several words or sentences. In a previous study Gated The recurring unit has been successfully used in the translator system [13]. With the advantages of Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU), it is interesting to use the hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) method in the review text classification system.

2 Overview

This chapter discuss some literature review about sentiment analysis, artificial neural network, one hot encoding, convolutional neural network, gated recurrent unit, cross-entropy error, back propagation and confusion method.

2.1 Sentiment Analysis

Sentiment analysis is the process of understanding, extracting and processing textual data to obtain information. It aims to find out opinions on an issue, or to identify trends in

things in the market. There are many benefits of sentiment analysis from various perspectives, including being able to get a general picture of public perceptions of service quality, monitoring a product, predicting sales, politics and making investors' decisions [13].

2.2 Artificial Neural Network

According to Gurney [14], an artificial neural network is a series of simple processing elements that are interconnected and modeled to imitate a human neural network. The ability to store information lies in the weights that connect between nodes or units. This ability is generated by adapting to each lesson. The artificial neural network consists of 3 layers, namely Input Layer, Hidden Layer and Output Layer, as shown in Fig. 1. This layer is tasked with performing functions to complete the system. This structure is based on a modification of the three-layer software architecture model, namely: data layer, service layer, and presentation layer [15]

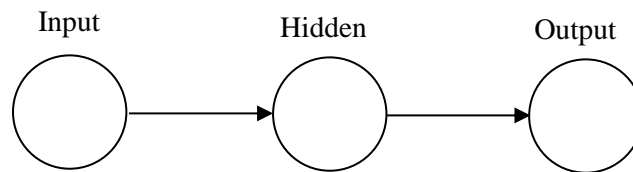


Figure 1 Neural Network Structure

According to Haykin [16] the components in the Artificial Neural Network are as follows:

1. Input Units

The input unit is the point of receiving the input signal in the form of data to be trained. In an artificial neural network, the input points are numerical data in the interval $[0,1]$ or $[-1,1]$. This point is a source of information that will be processed in the processing unit.

2. Weight.

The weight is the connecting component between the input unit and the processing unit. The weights will keep hidden information by adapting to each lesson.

3. Enhancer (Processing unit)

Used to add or add up the input signals and weights.

4. Activation function.

The activation function aims to limit the allowable output of neurons.

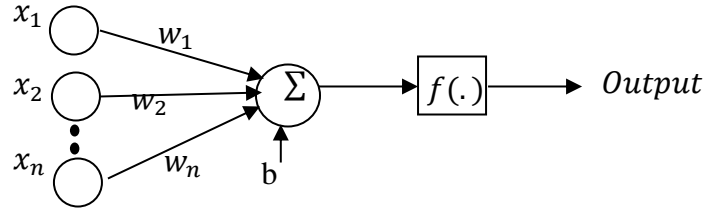


Figure 2 Basic Principles of Artificial Neural Networks

Fig. 2 is the basic principle of an artificial neural network. In the artificial conditional network there is a processing unit in the form of a mathematical model shown in Eq.(1) while the output unit is shown in Eq.(2).

$$\Sigma = \sum_{i=1}^n x_i \cdot w_i + b \quad (1)$$

$$y = f(.) \quad (2)$$

where x_i and w_i are inputs, and weights, $\Sigma \in \mathbb{R}$ is the processing unit operating result of $x_i \in \mathbb{R}$ and $w_i \in \mathbb{R}$ plus bias $b \in \mathbb{R}$. While $y \in \mathbb{R}$ is the output, and $f(.) \in \mathbb{R}$ is the activation function [16].

2.4 One Hot Encoding

Neural networks cannot process categorical data directly. Therefore, the input and output variables that will be processed by machine learning must be numeric. One technique to convert categorical variables into numeric variables is one hot encoding. One hot encoding is the process of changing one word variable in n word dictionaries with the value d , d is binary with n word dictionaries each. Each d will have a value of 1, if it is contained in the index of n data dictionaries and the remainder will be 0 [18].

2.5 Convolutional Neural Network

Convolutional Neural Network is a mathematical construct composed of 3 types of layers, namely convolution, pooling, and fully connected layers. Convolutional and pooling are tasked with performing feature extraction, while the Full Connected layer places the extracted features into the final output in the form of classification. Convolutional Neural Network architecture can be seen in Fig. 3 [19].

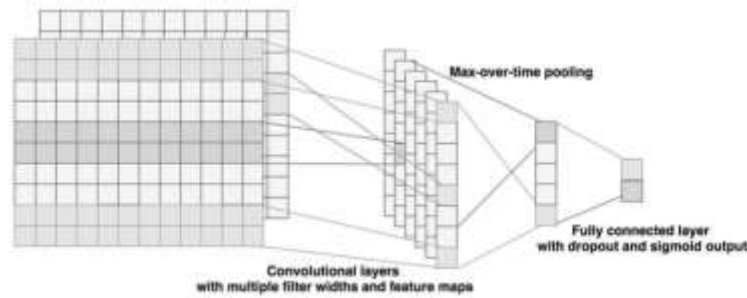


Figure 3 Convolutional Neural Network Architecture

The first stage in the Convolutional Neural Network (CNN) is convolution. In this process the sum of the elements in the input matrix is multiplied by the kernel (weight on CNN), after that it is continued with the ReLU (Rectifier Linear Unit) activation function. Then proceed to the pooling process (reducing the dimensions of the feature matrix). The final process on the Convolutional Neural Network is Fully Connected Layer and is continued at the prediction stage using the Softmax activation function [20].

2.6 Gated Recurrent Unit

Gated Recurrent Unit (GRU) is a network with a repeating layer proposed by Cho to make each recurring unit able to adaptively capture dependencies from different timescales. Gated Recurrent Units have unit gates that modulate the flow of information within the unit. The Gated Recurrent Unit (GRU) has two main components called Update Gate and Reset Gate which control the flow of information through each hidden unit. Broadly speaking, the update gate will determine how much past information should be stored, while the reset gate will combine past information and new information [21].

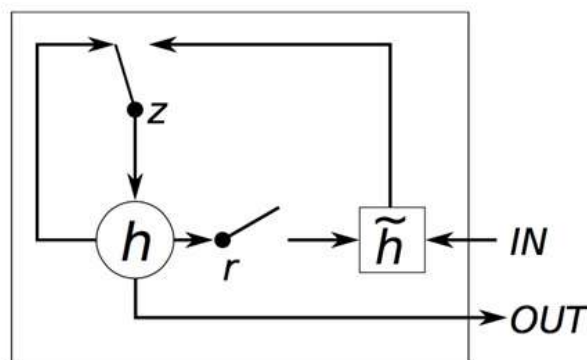


Figure 4 Gated Recurrent Unit Architecture

Fig. 4 shows that r is the reset gate and z is the update gate. While h is final memory which is used to calculate how much previous information needs to be stored, and h is

new memory or often known as candidate final memory which calculates candidate information that needs to be stored. [21]

Update Gate on Gated Recurrent Unit (GRU) can be seen in Eq.(3). Reset Gate can be seen in Eq.(4). Candidate final memory can be seen in Eq.(5). While the Final Memory can be seen in Eq.(6). Where x_t is the input vector at time t . h_{t-1} is past information prior to t . While W and U are the weights of the new input vector and the weight of the past information vector, respectively is a binary sigmoid activation function, \tanh is a hyperbolic tangent activation function [21]

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (3)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (4)$$

$$\tilde{h}_t = \tanh(W_h x_t + r_t \odot U_h h_{t-1}) \quad (5)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (6)$$

2.7 Cross-entropy Error

Cross-entropy error is used to measure errors or losses in learning artificial neural networks. If it is defined that $E_t \in R$ is the error value at time t , and $y_t \in R$ is the target or actual output value at time t while $\hat{y}_t \in R$ is the output of training results at time t , then the error can be defined according to Eq.(7) [22]

$$E_t = -y_t \log(\hat{y}_t) \quad (7)$$

2.8 Backpropagation

Backpropagation is one of the procedures for updating weights with optimization methods such as Gradient descent. Backpropagation will calculate the gradient from the loss function to all weights and biases in the Artificial Neural Network. Furthermore, the gradient will be forwarded to the optimization method to minimize the cost function. The gradient descent equation is described in Eq.(8).

$$W_{ij}^{new} = W_{ij}^{old} - \alpha \frac{\partial E}{\partial W_{ij}} \quad (8)$$

where W_{ij} is the weight element in the i row and j column, α is learning rate and E is error function [23]. *Gated Recurrent Unit* also use backpropagation in train.

2.9 Confusion Matrix

Confusion Matrix is a method that can measure accuracy in data mining concepts and decision-making systems. This method provides information on the comparison of the classification results carried out by the system with the actual classification results. In this method there are terms used, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Negative (TN) is the number of negative data detected correctly, while False Positive (FP) is negative data but detected as positive data. Meanwhile, True Positive (TP) is positive data that is detected correctly. False Negative (FN) is the opposite of True Positive, so the data is positive, but is detected as negative data. The Confusion Matrix table can be seen in Table 1. [24]

Tabel 1 Confusion Matrix

		Target	
		True	False
Prediction	True	TP	FP
	False	FN	TN

Accuracy describes how accurately the model can classify correctly. Thus, accuracy is the ratio of correct predictions (positive and negative) to the overall data. In other words, accuracy is the level of closeness of the predicted value to the actual (actual) value. The accuracy value can be obtained by Eq.9 [24]

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (9)$$

3 Research Method

In this research, the following methods will be used:

- (1) Conduct a literature search related to sentiment classification, Neural Networks, and their supporting algorithms, namely Convolutional Neural Network and Gated Recurrent Units. As well as other references related to this research.
- (2) Perform data retrieval. The data taken in this study are Tokopedia reviews (one of the e-commerce applications in Indonesia) in the Playstore application on December 3, 2020. The data obtained is 1000 data with 500 positive sentiment data and 500 negative sentiment data
- (3) Perform data design. The data obtained are review texts and scores in the Tokopedia application in the Google Playstore. The review text has a positive sentiment if the score in the playstore gets 4-5 stars. While the review text has a negative attitude if the score in the Playstore gets 1-2 stars. The review text with a 3-star score will be removed from the training. Then the data enters preprocessing stage. The preprocessing stages are as follows:

- i. Performing punctuation removal, which is the process of removing punctuation marks or characters that are not used. The allowed characters in this study were 38, the characters used can be seen in Table 2.
- ii. Doing Case Folding, which is the process of changing all letters in the dataset to be the same, in this study all data will be converted to lowercase.

Table 2 Character used in training

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	0	1
2	3	4	5	6	7	8
9	(spasi)	/n (Enter)				

- (4) Divide the data into training data and test data. The proportion of this distribution is 80% training data and 20% test data.
- (5) Forming a matrix representation of the review text using the One Hot Encoding algorithm so as to form a M matrix of size $m \times n$, where m is the number of characters in the review text and n is the number of index characters represented. In this study, the review text characters are 500 characters and 38 index characters so that each review will be represented in a 500×38 matrix. M_{ij} is 1 if the i th character in the review text is the same as the j character in the represented index character.
- (6) Develop an algorithm to classify data using a hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) algorithm. The outline of the CNN-GRU hybrid algorithm is as follows:
 - 1) Divide the data into training data and test data. The proportion of this distribution is 80% training data and 20% test data.
 - 2) Forming a matrix representation of the review text using the One Hot Encoding algorithm so as to form a M matrix of size $m \times n$, where m is the number of characters in the review text and n is the number of index characters represented. In this study, the review text characters are 500 characters and 38 index characters so that each review will be represented in a 500×38 matrix. M_{ij} is 1 if the i th character in the review text is the same as the j th character in the represented index character.
 - 3) Develop an algorithm to classify data using a hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) algorithm. The outline of the CNN-GRU hybrid algorithm is as follows:
 - i) Initializes the parameters of the Convolutional Neural Network process. The number of filters/kernels is 38, measuring 5×38 with real number values taken randomly at intervals $[0,1]$. The number of convolution layers is 1, stride (the amount of shift in each step) = 1
 - ii) Perform the process on the convolution layer.

- iii) Transforming the output matrix value with the ReLU activation function.
 - iv) Performs Max pooling on each output matrix.
 - v) Combining the output matrices in one vector that will be used as input in the Gated Recurrent Unit process. This results in a vector measuring 1×38 .
 - vi) Initializes the parameters used in the Gated Recurrent Unit (GRU). The parameters used are the maximum number of iterations (epochs), the minimum expected error (e), the learning rate and the number of neurons in the hidden layer as many as k . Initialize weights W_r, W_z, W_o, U_r, U_z , bias b_z, b_r, b_y and h_0 . Weights and biases are initialized randomly with real numbers in the interval $[0,1]$. Initialize iteration = 1 and $t = 1$
 - vii) Check conditions. If the iteration is less than or equal to the maximum iteration and the error is greater than the expected minimum error (e) then a feed-forward process is carried out. If not then take steps xv
 - viii) Calculates the value of z_t (Update Gate).
 - ix) Calculates the value of r_t (Reset Gate).
 - x) Calculates \tilde{h}_t .
 - xi) Calculating h_t which is Final memory as information for the next data.
 - xii) Calculates y (output prediction).
 - xiii) Counting Error (E).
 - xiv) Updating the weights with the Backpropagation method. If $t < a$ lot of review text then t is increased by 1 and return to step viii. If not then iteration is increased by 1 and return to step vii.
 - xv) Saving optimal weight
- (7) Testing the test data and measuring its accuracy with the Confusion Matrix. The process carried out in this step is as follows:
- i. Load test data, optimal weights and biases generated in step 6 and initialize $t = 1$, TP (True Positive) = TN (True Negative) = FP (False Positive) = FN (False Negative) = 1.
 - ii. Calculates the value of z_t (Update Gate).
 - iii. Calculates the value of r_t (Reset Gate).
 - iv. Calculating \tilde{h}_t
 - v. Calculating h_t which is Final memory as information for the next data.
 - vi. Calculates y (output prediction).
 - vii. Check conditions. If $y = 1$ and $y = 1$ where y is the target output then TP is added 1. If $y = 0$ and $y = 0$ then TN is added 1. If $y = 1$ and $y = 0$ then FP is added 1. If $y = 0$ and $y = 1$ then FN plus 1.
 - viii. Increase t by 1 and return to step (iii)
 - ix. Counting Accuracy.
- (8) Create a program and apply to the test data.

4 Result and Discussion

Solving the problem of classification of review texts using the hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) method is implemented in the program using the Python programming language with Spyder software.

The parameters used in the Convolutional Neural Network layer are as follows: variations in the number of kernels $k = \{128,256\}$, each kernel has a kernel size of $H_{5 \times 38}$, the number of characters of text review $T = 500$. From the process carried out on the Convolutional Neural Network layer, feature extraction will be generated from the review text. The next step is review text training with the Gated Recurrent Unit (GRU). The initialization of parameters in the Gated Recurrent Unit (GRU) layer is as follows: variation in learning rate value = $\{0.01,0.001\}$, maximum value variation of epoch iteration = $\{25,50,100,200\}$, minimum expected error value $minError = 0.00001$. The following variations of parameter values with error and accuracy values from the Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) training processes can be seen in Table 3.

Table 3. Results of training and test data

No	Convolutional Neural Network	Gated Recurrent Unit			Error	Accuracy
	Kernel (k)	Learning rate	Epoch	Hidden node		
1	128	0,01	25	87	0,4932	77,37%
2				128	0,4727	82%
3			50	87	0,5099	80,5%
4				128	0,4604	61,87%
5			100	87	0,4889	79,87%
6				128	0,4998	53,25%
7			200	87	0,5062	55,75%
8				128	0,4970	53,25%
9		0,001	25	87	0,5579	85%
10				128	0,5534	75%
11			50	87	0,5640	64,75%
12				128	0,5319	75,25%
13			100	87	0,5531	86,87%
14				128	0,5072	83,87%
15			200	87	0,4785	83,75%
16				128	0,4893	87,12%
17	256	0,01	25	172	0,3457	92%
18				256	0,3134	92,5%
19			50	172	0,3367	89,37%
20				256	0,3398	56,75%
21			100	172	0,3245	53,87%

No	Convolutional Neural Network	Gated Recurrent Unit			Error	Accuracy
	Kernel (k)	Learning rate	Epoch	Hidden node		
22			200	256	0,3418	53,62%
23				172	0,3228	54,75%
24				256	0,3877	51,62%
25		0,001	25	172	0,4796	82,37%
26				256	0,4993	81%
27			50	172	0,4622	81,87%
28				256	0,3781	84%
29			100	172	0,3606	89%
30				256	0,3315	78%
31		200	172	0,3167	89,5%	
32			256	0,2879	91,12%	

From the validation test, the hyperparameters with the best accuracy will be taken. The selected hyperparameters and weights and biases will be used as models in the formation of the review text classification application using the Hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) methods. The results of the validation test can be seen in Table 4.

Table 4 Validation Test Results Against Test Data

No	Convolutional Neural Network	Gated Recurrent Unit			Accuracy
	Kernel (k)	Learning rate	Epoch	Hidden node	
1	256	0,01	25	256	88.5%
2	256	0,01	25	172	88%
3	256	0,001	200	256	84,5%
4	256	0,001	200	172	84,5%
5	256	0,01	50	172	87%
6	256	0,001	100	172	82,5%
7	128	0,001	200	128	84,5%
8	128	0,001	100	87	83,5%
9	128	0,001	25	87	75%
10	256	0,001	50	256	85,5%

Based on the results of the validation test on the test data listed in Table 11.2, the best accuracy is 88.5% with the following hyperparameters: many kernels (k) = 256, learning rate (α) = 0.01, the number of iterations = 25, and the number of hidden nodes = 256. Because the accuracy value is 92.5% for training data, and 90% for test data, the training model obtained can be used in review text classification applications with hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU).

5 Conclusion

Hybrid Convolutional Neural Network (CNN) and Gate Recurrent Unit (GRU) can be used to solve review text classification problems. The processes carried out in the classification of review texts using the hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) methods include: inputting data and converting the review text into matrix input and the sentiment label of the review text being the target, initializing the Convolutional Neural Network (CNN) parameters.), generate Convolutional Neural Network (CNN) kernel, Convolutional Neural Network (CNN) feed forward, initialize Gated Recurrent Unit (GRU) parameters, generate Gated Recurrent Unit (GRU) weights and biases, Gated Recurrent Unit (GRU) feed forward, calculate Cross-entropy Error values, perform Gated Recurrent Unit (GRU) feedback, update weights and biases.

The implementation of the program as an example of a sentiment classification case uses the Tokopedia application review data located in the Google Playstore application. Data obtained as much as 1000 data reviews. Completion of the review text classification with the data obtained an accuracy of 88%. The use of different parameters makes the accuracy values vary. The number of kernels initialized in the Convolutional Neural Network (CNN) layer greatly affects the results of the training. Initialization of learning rate that is too large makes the program fail in training. The number of iterations affects the training results, the number of iterations that are too large can reduce the accuracy obtained. The number of hidden nodes also affects the training results with hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU).

6 References

- [1] Deny, S., 2019. *Daftar E-Commerce dengan Transaksi Tertinggi di RI, Siapa Saja?* - *Bisnis Liputan6.com*, <https://www.liputan6.com/bisnis/read/4048211/daftar-e-commerce-dengan-transaksi-tertinggi-di-ri-siapa-saja> (23 April 2020)
- [2] Debora, Y., 2016, *Berapa Besar Pengaruh Ulasan Pembeli Saat Berbelanja Online?* - *Tirto.id*. Diambil kembali dari <https://tirto.id/berapa-besar-pengaruh-ulasan-pembeli-saat-berbelanja-online-b7Gm> (23 April 2020).
- [3] Fang, X., dan Zhan, J., 2015, Sentiment analysis using product review data. *Journal of Big Data*.
- [4] Liu, B., 2010, Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, Second Edition*. Taylor and Francis Group, Boca.
- [5] Pang, B., dan Lee, L., 2008, Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, vol. 2 : 1-135. <https://doi.org/10.1561/1500000011>

- [6] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., dan Tan, R. S., 2017, A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, vol. 89 : 389-396.
- [7] Kalchbrenner, N., Grefenstette, E., dan Blunsom, P., 2014, A Convolutional Neural Network for Modelling Sentences. *Computation and Language*.
- [8] Shen, Y., He, X., Gao, J., Deng, L., dan Mesnil, G., 2014, Learning Semantic Representations Using Convolutional Neural Networks for Web Search. *Proceedings of the 23rd International Conference on World Wide Web*, 373–374.
- [9] Dertat, A., 2017, *Applied deep learning-Part 4: convolutional neural networks*, Toward DataScience: <https://towardsdatascience.com/applied-deeplearning-part-4-convolutional-neural-networks-584bc134c1e2>.
- [10] Zhang, X., Zhao, J., dan LeCun, Y., 2015, Character-level convolutional networks for text classification. *arXiv preprint arXiv:1509.01626*.
- [11] Zhou, C., Sun, C., Liu, Z., dan Lau, F. C., 2015, A C-LSTM Neural Network for Text Classification. *arXiv preprint arxiv: 1511.08630*.
- [12] Chung, J., Gulcehre, C., Cho, K., dan Bengio, Y., 2014, Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [13] Ipmawati, J., 2016, Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *IJNS-Indonesian Journal on Networking and Security*, 6(1).
- [14] Gurney, K., 2014, *An introduction to neural networks*. CRC press.
- [15] Vivian F., López, Santo L., Medina, 2012, Taranis: Neural networks and intelligent agents in the early warning against floods, *Expert Systems with Applications*. 39, 10031–10037.
- [16] Haykin, S. S., 2009, *Neural networks and learning machines*.
- [17] Bendersky, E., 2016, The softmax function and its derivative. Diambil kembali dari <https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative/> (diakses tanggal 23 April 2020)
- [18] Vasudev, 2019, “What Is One Hot Encoding? Why And When Do You Have to Use It?”, Hackernoon: <https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f>
- [19] Yamashita, R., Nishio, M., Do, R. K., dan Togashi, K., 2018, Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4), 611-629.
- [20] Zhang, Y., Pezeshki, M., Brakel, P., Zhang, S., Bengio, C. L., dan Courville, A., 2017, Towards end-to-end speech recognition with deep convolutional neural networks. *arXiv preprint arXiv:1701.02720*.
- [21] Chung, J., Gulcehre, C., Cho, K., dan Bengio, Y., 2014, Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [22] Nasr, G. E., Badr, E. A., dan Joun, C., 2002, Cross entropy error function in neural networks : Forecasting Gasoline Demand. *Proceedings of the Fifteenth*

International Florida Artificial Intelligence Research Society Conference. Pages 381–384

- [23] Rojas, R., 1996, *The backpropagation algorithm*. In *Neural networks*. Springer, Berlin, Heidelberg.
- [24] Visa, S., Ramsay, B., Ralescu, A. L., dan Van Der Knaap, E., 2011, Confusion Matrix-based Feature Selection. *MAICS*, 710, 120-127.