IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.12, No.1, January 2018, pp. 1~10
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.19237

Sentiment Analysis of Movie Opinion in Twitter Using Dynamic Convolutional Neural Network Algorithm

Fajar Ratnawati*[1], Edi Winarko[2]
[1]Master Program of Computer Science; FMIPA UGM, Yogyakarta
[2]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta
e-mail:*[1] **fajar.ratnawati@gmail.com**,[2] ewinarko@ugm.ac.id

Film memiliki karakteristik yang unik. Ketika seseorang menulis opini suatu film, tidak hanya unsur film itu sendiri yang ditulis, tetapi juga orang-orang yang terkait dalam film juga ditulis. Untuk mendapatkan kecenderungan opini terhadap film apakah cenderung beropini positif, negatif atau netral maka dibutuhkan analisis sentimen. Opini film biasa ditulis dimedia sosial terutama twitter. Penelitian ini bertujuan untuk melakukan pengklasifikasian terhadap sentimen positif, netral dan negatif dari opini film berbahasa Indonesia serta mencari akurasi dari metode yang digunakan yaitu Dynamic Convolutional Neural Network yang nantinya akan dibandingkan dengan metode Naive Bayes. Hasil pengujian pada sistem yang dibangun memperlihatkan bahwa algoritme Dynamic Convolutional Neural Network memberikan hasil akurasi yang lebih baik daripada metode Naive Bayes, dengan nilai akurasi sebesar 85,56 % sedangkan nilai akurasi yang dihasilkan Naive Bayes sebesar 82,92 %.

Kata kunci— *analisis sentimen, opini film, twitter, Dynamic Convolutional Neural Network*

Abstract

The movie has unique characteristics. When someone writes an opinion about a movie, not only the story in the movie itself is written, but also the people involved in the movie are also written. Opinion ordinary movie written in social media primarily Twitter.To get a tendency of opinion on the movie, whether an opinion is likely positive, negative or neutral, it takes a sentiment analysis. This study aims to classify the sentiment is positive, negative and neutral from opinions Indonesian language movie and look for the accuracy, precision, recall, and f-measure of the method used is Dynamic Convolutional Neural Network. The test results on a system that is built to show that Dynamic Convolutional Neural Network algorithm provides accurate results better than Naive Bayes method, the value of accuracy of 80,99%, the value of precision 81,00%, recall 81,00%, f-measure 79,00% while the value of the resulting accuracy Naive Bayes amounted to 76,21%, precision 78,00%, recall 76,00%, f-measure 75,00%.

Keywords— *sentiment analysis, movie opinion, twitter, Dynamic Convolutional Neural Network*

Received February 8th,2017; Revised September 15th, 2017; Accepted January 3rd, 2018

1. INTRODUCTION

The rapid development of information can not be separated from web service providers that provide diverse information. The more rapid the development of technology, the more people write their opinions about products or services in social media. Opinion is central to almost all human activity and is a major influence on human behavior. Confidence and perception about the product or service and the choices we make, so far conditioned on how others see and evaluate them. When it comes to making decisions, we often seek or seek the opinions of others. This not only applies to individuals but also applies to organizations [1]. In addition, social media is also used to write opinions about favorite movies ever watched. One of the social media used is twitter.

Twitter is an online social networking and microblogging service that allows its users to send and read text-based messages of up to 140 characters, known as tweets. Twitter was founded in March 2006 by Jack Dorsey, and its social networking site was launched in July. Since its launch, Twitter has become one of the ten most visited sites on the internet and is dubbed with a short message from the internet. On Twitter, unregistered users can only read tweets, while registered users can write tweets through the website interface, short messages (SMS) or through various apps for mobile devices.

In 2012 Twitter has more than 100 million users worldwide and continues to increase every day as many as 300,000 users and Twitter every day to get more than 3 million requests. Of that number, Indonesia became a country that ranked 8th in accessing the Twitter site. As many as 34% of Twitter users use computer devices and 55% access Twitter via mobile [2]. Thus we can utilize twitter data about film opinion to get sentiment information contained therein. Do those opinions tend to be positive, negative or neutral [3]?

Natural Language Processing (NLP) is the application of computer science, especially computational linguistics (computational linguistics), to examine the interaction between computers with (natural) human language. NLP seeks to solve problems to understand human natural language, with all its grammatical and semantic rules, and transforms the language into a formal representation that can be processed by a computer [4].

Sentiment analysis or opinion mining is the process of understanding, extracting and processing textual data automatically to get sentiment information contained in an opinion sentence. Sentiment Analysis is done to view opinions or tendency of opinion to a problem or object by someone, whether tends to view or opinionated negative or positive [5]. Sentiment Analysis can be differentiated based on its data source, some of the level used in Sentiment Analysis is Sentiment Analysis at the document level and Sentiment Analysis at sentence level [6].

Text Mining is a process of pattern extraction (useful information and knowledge) from a large amount of unstructured data. Text mining has a purpose and uses the same process with data mining, but has different inputs. Inputs for text mining are data that is not (or less) structured, such as documents, word, PDF, text quotes etc. whereas inputs for data mining are structured data [7].

Deep Learning is about learning multiple levels of representation and abstraction that help to understand data such as images, sounds, and text [8]. In this research, the method used is Dynamic Convolutional Neural Network (DCNN). DCNN is one of machine learning method for sentence modeling which is included in Deep learning method. The layers of DCNN are formed by convolution operations [9] followed by pooling operations [10].

Deep Learning is about learning multiple levels of representation and abstraction that help to understand data such as images, sounds, and text [8]. In this research, the method used is Dynamic Convolutional Neural Network (DCNN). DCNN is one of machine learning method for sentence modeling which is included in Deep learning method. The layers of DCNN are formed by convolution operations [9] followed by pooling operations [10].

2. METHODS

2.1 System Analysis

This research begins with do tweet data collection by scraping tweets containing hashtag # title of the desired film. This data collection is done by utilizing the Twitter Search API. Tweet data collected include id, username, index, data_id, URL, date and time of the tweet, tweet content, original_tweet.

The next stage is the labeling of tweets in accordance with predetermined sentiments of positive sentiment for positive film opinions, negative sentiments for negative film opinions and neutral sentiment for a neutral film opinion.

After all the data already has each class then separated into three parts of training data, data validation and data testing. Furthermore, the third part is subject to preprocessing process. The preprocessing stages include:

1. Change all the letters to lowercase.
2. Delete username.
3. Removing URL.
4. Delete special characters (# $% * etc).

After the training data, data validation and net testing data then performed the next stage of the conversion of sentences into numbers with the conversion stage, among others:

1. Create a word dictionary.
2. Determine the number of longest sentences.
3. Determining the padding number.
4. Change the sentence into numbers.
5. Perform the padding according to the longest sentence.

After the conversion phase of the sentence is completed then implemented dynamic convolution neural network algorithm in the training process to build the probability model of training data and validation data. And then testing the classification model sentiment generated in the training process by using new data tweet (data testing). Testing will also be done by other methods using the same data as a comparison. This test is done by using 5 fold cross-validation technique. To calculate the value of its accuracy is done by using Equation (1).

$$\text{accuracy} = \frac{\text{the number of sentiments is correct}}{\textit{the number of test data}} \times 100\% \tag{1}$$

To measure the performance of classification the way used in addition to calculating accuracy is to calculate the precision, recall, and f-measure. To calculate the precision value can be done by using Equation (2), the recall calculation can use Equation (3) and the f-measure calculation using Eq. (4).

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \tag{2}$$

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \tag{3}$$

$$\text{f} - \text{measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

For more details, step research can be seen in Figure 1

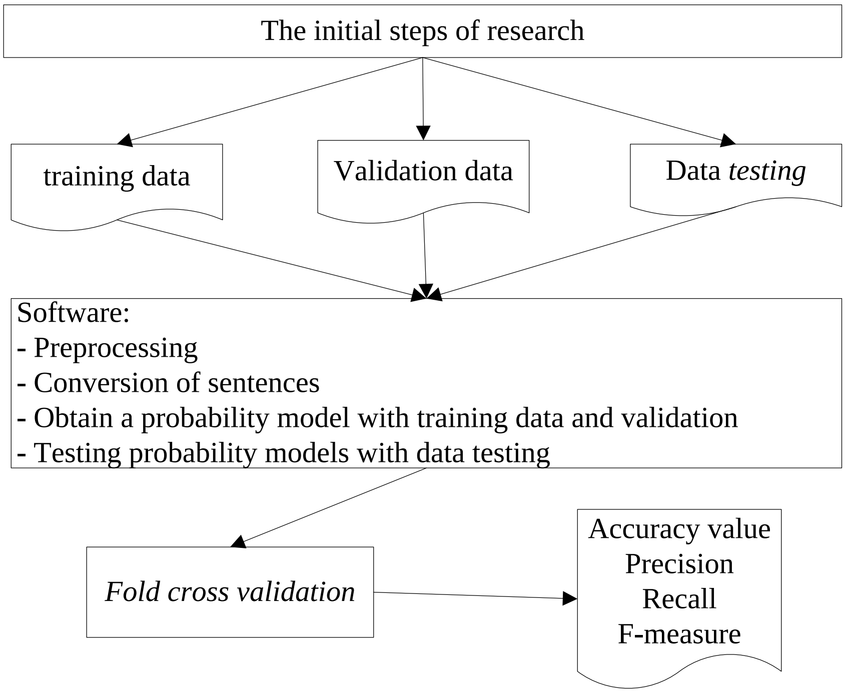
The initial steps of research
training data
Validation data
Data *testing*
Software:
- Preprocessing
- Conversion of sentences
- Obtain a probability model with training data and validation
- Testing probability models with data testing
Fold cross validation
Accuracy value
Precision
Recall
F-measure

Figure 1 Flow Research

2.2 System architecture

In general, this system consists of five parts of which are data retrieval, preprocessing tweet, sentence conversion, tweet sentiment classification and sentence accuracy results. The architecture design of the system can be seen in Figure 2.

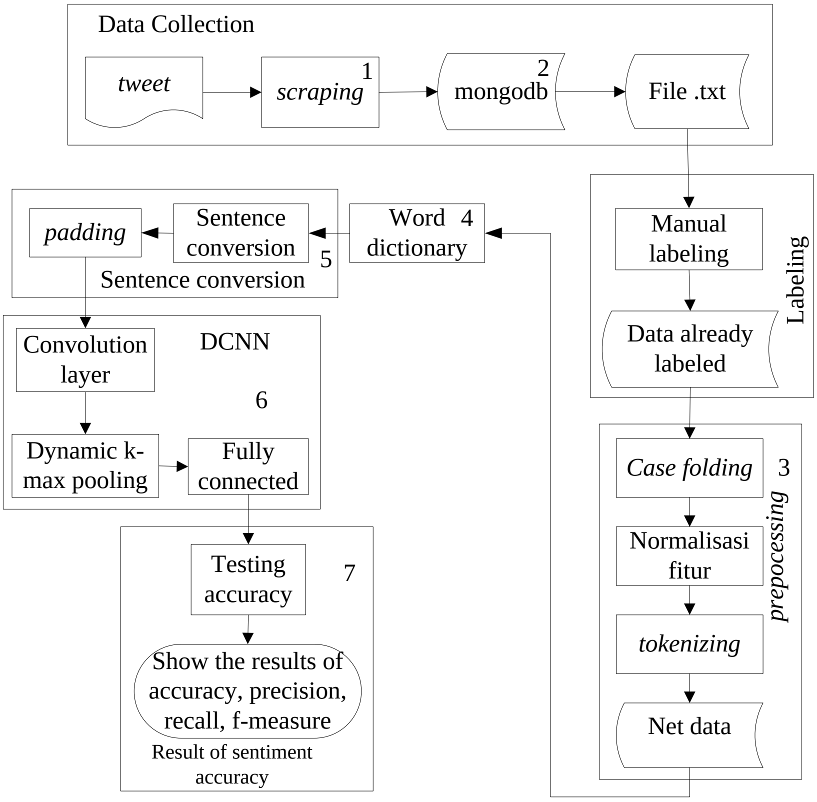

Figure 2 System Architecture

2.3 Data and Data Class

Tweets that are collected are tweets that contain hashtag # movie titles. Tweet data retrieval that will be used in this research starting from January 1, 2015, to December 31, 2015. The data used as many as 3600 sentences.

Data is divided into three sentiments: positive sentiments for positive film comments or opinions, negative sentiments for negative film comments or opinions and neutral sentiments for commentary or film opinion are neutral.

2.4 Data Modeling

Data modeling is made with the purpose to test the accuracy of system predictions based on model data already created. The training process to build the probability model in this research required training data and validation data. Furthermore, testing the classification of sentiment model that is produced in the training process by using new tweet data (data testing). System overview can be seen in Figure 3.

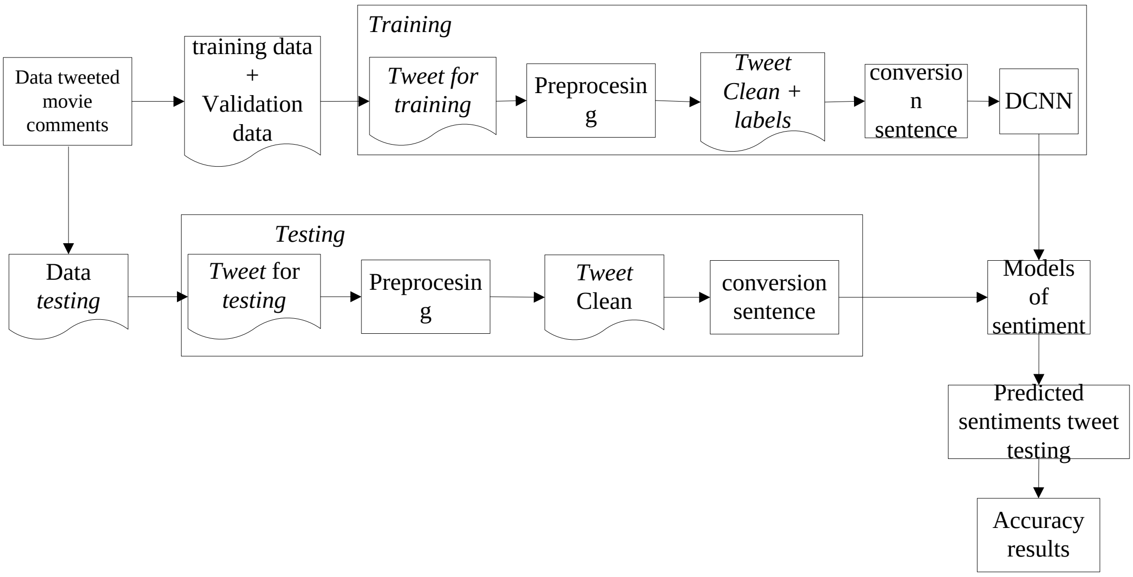

Figure 3 Data Modeling

2.5 Dynamic Convolutional Neural Network

Dynamic Convolutional Neural Network is an algorithm that uses a convolution architecture that alternates between wide convolution layers with dynamic pooling layers determined by dynamic k-max pooling. Since the feature map width of the middle layer varies according to the length of the input sentence the resulting architecture is called Dynamic Convolutional Neural Network [11].

2.5.1. Wide Convolution

To obtain the first layer matrix from DCNN requires embedding word $w_i \in R^d$ for each word in sentence and input sentence matrix $s \in R^{d\times s}$. The value of embedding word w_i are parameters that are optimized during the training. The convolution layer is obtained from the convolution process between weight matrices $m \in R^{d\times m}$ with an activation matrix on the previous layer. So we get the matrix for wide convolution with dimension $d \times (s + m - 1)$ [11].

2.5.2. k-Max Pooling

k-max pooling is applied after the last convolution layer. K-max pooling is used to reduce the sample length of different vectors to the same length as the fully connected layer. Therefore k-max pooling is applied after the top convolution layer. K-max pooling is taken from k max in sequence instead of single max value.

2.5.3. Dynamic k-Max Pooling

The dynamic k-max pooling operation is a k-max pooling operation where k is a function of the length of the sentence and the depth of the network [11]. The pooling parameter can be seen in Eq. (5).

$$k_l = max\left(k_{top}, \left\lceil\frac{L-l}{L} s\right\rceil\right) \tag{5}$$

Where l is the number of convolution layers in progress where pooling is applied. L is the total number of convolution layers in the network, k_top is a fixed pooling parameter for the topmost convolution layer and s is the length of the sentence.

2.5.4. Folding

Folding is applied after the last convolution layer and before the k-max pooling layer by summing every two lines on the feature map [11]. For line map d, fold back the line map $d/2$, thereby reducing half the size of the representation.

3. RESULTS AND DISCUSSION

After all the designs have been completed and the system has been built, the next step is to use the system to generate probability models of training data and validation data, test the probability model's accuracy with data testing and experiment using the same data with other applications and compare the results accuracy with the created application.

3.1 Testing epoch count

Testing begins with determining the number of the epoch. Due to resource constraints, not all values are tested so that only three values are chosen, 10, 20 and 30. For this epoch test, other parameters are randomly assigned such as filter width, filter count, and k-top. Accuracy is calculated using the average of k-fold cross-validation. Table 1 shows the results of the epoch test.

Table 1. Epoch Testing

Epoch	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
10	3,45	78,88	76,00	78,00	77,00
20	6,91	78,14	78,00	78,00	77,00
30	10,22	78,88	77,00	79,00	77,00

From Table 1 above it can be seen that at the time of epoch 20 the accuracy decreased from 78.88% to 78.14% but by the time the epoch was changed to 30, the result of accuracy went up and the value was equal to epoch 10 that is 78,88% but time needed for a longer process when compared to epoch 10. The more epoch the longer the time it takes to process it.

3.2 Testing filter width

For testing the width of the filter is done by the number of epoch 10. The epoch value is selected because it results in a higher accuracy value in the previous epoch test. Accuracy is calculated using the average of k-fold cross-validation. The results of the filter width test can be seen in Table 2.

Table 2 Testing Filter Width

Filter width	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
5,3	3,09	80,99	81,00	81,00	81,00
7,5	3,45	78,88	76,00	78,00	77,00
9,7	3,83	79,75	81,00	79,00	78,00

From Table 2 it can be seen that the 5.3 filter width has higher accuracy results than other filter widths. The suitability of the filter width between the first and second layers can affect the results of the accuracy.

3.3 *Testing the number of filters*

Testing the number of filters is to determine the number of filters that will be used on the system. The test is done by the number of epoch 10, filter width of 5.3 and k-top 4. Accuracy is calculated using the average of k-fold cross-validation. For the test results, the number of filters can be seen in Table 3.

Table 3 Testing Number of Filters

Number of filters	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
5,10	2,81	79,50	80,00	79,00	78,00
6,14	3,09	80,99	81,00	81,00	79,00
8,16	3,38	79,50	78,00	79,00	78,00

Due to resource constraints, not all values are tested on the number of filters. From the results of Table 3 can be seen when the value of 6.14 is inserted the results of increased accuracy and when the value of 8.16 accuracy results down and the value is equal to the number of filters 5.10. Choosing the right number of filters will produce good accuracy.

3.4 *K-top test*

The k-top test determines the number of convolution layers used. The test was conducted with the number of epoch 10, the filter width 5,3 and the number of filter 6,14. These values are derived from values that have the highest accuracy results from the previous test. Accuracy is calculated using the average of k-fold cross-validation. For the results of k-top testing can be seen in Table 4.

Table 4 Testing k-top

k-top	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
4	3,09	80,99	81,00	81,00	79,00
5	3,14	80,13	81,00	80,00	78,00
6	3,09	79,87	80,00	79,00	78,00

From Table 4 above shows the best accuracy results when the k-top is worth 4. The more complex the architecture must be changed also other parameters such as learning rate, the number of the epoch and L2 regularization in order to avoid overfitting.

The results of epoch parameter testing, filter width, filter number and k-top can be seen in Table 5.

Table 5 Hyperparameter Testing Parameters

Epoch	10
Filter Width	5,3
Number of Filters	6,14
k-top	4
Time (minutes)	3,09
Accuracy (%)	80,99
Precision (%)	81,00
Recall (%)	81,00
f-measure (%)	79,00

The individual accuracy for each fold of the hyperparameter test results selected from the previous test can be seen in Table 6.

Table 6 Accuracy Each Fold of Hyperparameter Testing

Fold	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
I	3,30	78,26	79,00	78,00	76,00
II	3,05	80,75	82,00	81,00	81,00
III	3,03	78,88	77,00	79,00	77,00
IV	3,03	83,85	85,00	84,00	83,00
V	3,04	83,23	84,00	83,00	82,00
Average	**3,09**	**80,99**	**81,00**	**81,00**	**79,00**

3.5 Testing L2 Regularization

The purpose of this test is to get a better accuracy value. Due to resource constraints, not all L2 regularization values are tested. For L2 regularization test results can be seen in Table 7.

Table 7 Changes of Accuracy on System to Value Changes λ

λ	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
0,001	3,13	79,88	79,00	80,00	78,00
0,0001	3,09	80,99	81,00	81,00	79,00
0,00001	3,00	79,50	80,00	79,00	78,00

From the table above can be seen that the change of value λ affect the result of accuracy and best accuracy produced is 80,99%.

3.6 Activation function test

This activation function test aims to find the best accuracy value. The activation function to be tested is the activation function of tanh and the activation function of reLU. For the test results can be seen in Table 8.

Table 8 Test Results of Activation Functions

k-fold	Tanh					reLU				
	ime (minutes)	Test_acc (%)	Precision (%)	Recall (%)	f-measure (%)	time (minutes)	Test_acc (%)	Precision (%)	Recall (%)	f-measure (%)
k1	3,30	78,26	79,00	78,00	76,00	3,21	81,37	77,00	81,00	78,00
k2	3,05	80,75	82,00	81,00	81,00	2,94	51,55	50,00	52,00	46,00
k3	3,03	78,88	77,00	79,00	77,00	2,97	71,43	66,00	71,00	69,00
k4	3,03	83,85	85,00	84,00	83,00	2,97	69,57	65,00	70,00	67,00
k5	3,04	83,23	84,00	83,00	82,00	2,94	77,02	72,00	77,00	74,00
avg	**3,09**	**80,99**	**81,00**	**81,00**	**79,00**	**2,99**	**70,19**	**66,00**	**70,00**	**67,00**

From the tests performed show that the activation function tanh produce better accuracy value when compared with the activation function reLU. This is probably because the data used is small.

3.7 The overall results of testing

After testing the parameters that have been determined that epoch, filter width, the number of filters, k-top, fold and L2 regularization it will be evaluated with test data and the overall test results can be seen in Table 9.

Table 9 Overall Test Results

Parameter	Value	Accuracy
Epoch	10	80,99%
Filter Width	5,3	
Number of Filters	6,14	
k-top	4	
Dropout	0,5	
L2 *regularization*	0.0001	
Activation Function	*tanh*	

3.8 Testing with other methods

Testing with other methods in this research is by using the method of naive Bayes classifier. The data used as model and testing is the same as the data used in the DCNN method.

In this test In this test used about 1200 positive tweets, 1200 tweet negative and 1200 tweet neutral to train the classifier and stored in the file .txt so that the total data for the classifier is 3600.

This method is implemented using NLTK python and library. With the aim of being able to automatically classify tweets as positive, negative and neutral tweet sentiments. For the test results of each fold can be seen in Table 10.

Table 10 Naive Bayes Test Results

Fold	Time (seconds)	Accuracy (%)	Precision (%)	Recall (%)	f-measure (%)
I	8,24	73,91	81	74	76
II	9,43	82,93	82	83	82
III	9,66	81,68	87	82	79
IV	9,86	74,53	83	75	72
V	9,94	68,01	64	68	65
Average	**9,43**	**76,21**	**78,00**	**76,00**	**75,00**

3.9 Comparison of the results of opinion sentiment movies

This study aims to classify the positive sentiments, negative sentiments and neutral sentiments of Indonesian to the movie. The results of this sentiment by the system will be calculated accuracy.

After getting the accuracy, it can be compared with the non-learning method that is naive Bayes. Using test data and k-fold division equal to dynamic convolutional neural network algorithm then the comparison of accuracy can be seen in Table 11.

Table 11 Comparison of Accuracy Results

k-fold	**DCNN**					**Naive Bayes**				
	time (minutes)	Test_acc (%)	Precision (%)	Recall (%)	f-measure (%)	time (minutes)	Test_acc (%)	Precision (%)	Recall (%)	f-measure (%)
k1	3,30	78,26	79,00	78,00	76,00	8,24	73,91	81,00	74,00	76,00
k2	3,05	80,75	82,00	81,00	81,00	9,43	82,93	82,00	83,00	82,00
k3	3,03	78,88	77,00	79,00	77,00	9,66	81,68	83,00	82,00	79,00
k4	3,03	83,85	85,00	84,00	83,00	9,86	74,53	78,00	75,00	72,00
k5	3,04	83,23	84,00	83,00	82,00	9,94	68,01	64,00	68,00	65,00
avg	**3,09**	**80,99**	**81,00**	**81,00**	**79,00**	**9,43**	**76,21**	**78,00**	**76,00**	**75,00**

From the above comparative results, it can be concluded that the method of deep learning algorithm Dynamic Convolutional Neural Network is superior compared to Naive Bayes with Indonesian film opinion test data taken from tweet data.

4. CONCLUSION

In accordance with the purpose of this study is to test the accuracy of the dynamic convolutional neural network algorithm and naive Bayes method has been achieved. From the results of testing and analysis that has been done can be concluded some thing that is classification analysis of opinion sentiment film in Indonesian language can be done with deep learning method of dynamic convolutional neural network architecture, dynamic convolutional neural network algorithm better than naive bayes method based on accuracy obtained, where the highest accuracy value obtained dynamic convolutional neural network algorithm is 80.99% while the highest accuracy obtained naive bayes method is 76.21%, the best accuracy obtained with the value of each parameter that is the number of epoch 10, filter width 5,3, number of filter 6,14, k.top 4, probability dropout 0,5 and parameter of learning rate equal to 0,1 with value of accuracy 80,99%, preprocessing process used in this research only two process that is case folding and normalization of feature and data the inserted test must be in the dictionary so that the test data can be scanned ice. If none then unrecognized words in the dictionary will be deleted.

In this study, there are still some shortcomings that can be improved again in subsequent research. Some suggestions for further research are to try to test different parameter values, by changing the parameter values so that it is possible to get a better accuracy value, try to use other learning algorithms like Adadelta, RMSprop etc., improve the data preparation process where the quality of the data to be processed into better so that processing in the mining process becomes more optimal. Suppose an addition to a preprocessing process such as a slang word list, a list of words that have the same meaning etc., try to use other methods to handle overfitting like L1 regularization etc. and apply semantic sentences in subsequent research.

ACKNOWLEDGMENT

The authors would like to thank the parents who have given financial support to this research.

REFERENCES

[1] B. Liu, "*Sentiment Analysis and Opinion Mining.*" Morgan & Claypool Publishers, May 2012.

[2] K. Daniells, Infographic: *Social Media Statistics For 2012*, 2012, [Online]. Available: http://www.digitalbuzzblog.com/social-media-statistics-stats-2012-infographic/ , [accesed: 20-Jan-2017].

[3] H. Cui, V. Mittal, and M. Datar, "*Comparative Experiments on Sentiment Classification for Online Product Reviews.*" Department of Computer Science, National University of Singapore. Singapore, 2008.

[4] J. Pustejovsky, and A. Stubbs, "*Natural Language Annotation for Machine Learning.*" Cambridge University Press, 2012.

[5] B. Pang, L. Lee, and S. Vaithyanathan, "*Thumbs up Sentiment classification using machine learning techniques.*" Computer Science Department, Cornell University, New York, USA: Cambridge University Press, 2002.

[6] C.R. Fink, D.S. Chou, J.J. Kopecky, A.J. Llorens, "*Coarse and Fine-Grained Sentiment Analysis of Social Media Text*." Johns Hopkins Apl Technical Digest, Volume 30, Number 1, 2011.

[7] R. Feldman and J. Sanger, "*The Text Mining Handbook Advances Approaches in Analyzing Unstructured Data*." Cambridge University Press, New York, 2007.

[8] LISA lab, " *Deep Learning* Tutorial." Release 0.1. University of Montreal, 2015.

[9] A. Weibel, T. Hanazawa, G. Hinton, K. Shikano, and K.J. Lang, "*Readings in Speech Recognition*." *chapter Phoneme Recognition Using Time-delay Neural Networks*, pages 393–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

[10] R. Collobert, and J. Weston, "*A unified architecture for natural language processing: Deep neural network with multitasking learning*." In *International Conference on Machine Learning, ICML*, 2008.

[11] N. Kalchbrenner, E. Grefenstette, P. Blunsom, "*A Convolutional Neural Network for Modelling Sentence*." Department of Computer Science University of Oxford, 2014.