InComTech: Jurnal Telekomunikasi dan Komputer
vol.13, no.3, Desember 2023, 167-176
http://publikasi.mercubuana.ac.id/index.php/Incomtech
P-ISSN: 2085-4811 E-ISSN: 2579-6089

Multi-classification Sentiment Analysis using Convolution Neural Network and Long-Short Term Memory with Attention Model

Yohanes Christianto[1*], Suharjito Suharjito[1]

[1]*Computer Science, Binus University*
Jl. K H. Syahdan No. 9, Kelurahan Kemanggisan, Kecamatan Palmerah, Jakarta Bara
*Email Penulis Koresponden: yohanes.christianto@binus.ac.id

Abstract:
Multi-classification Sentiment Analysis from sentences in Bahasa is a challenging process due to problems in slang, local language combined with many English words. Current state-of-art methods rely on feature extraction using unsupervised treatment. A research to solve this problem was conducted using LSTM and CNN that are capable of learning complex features from the lower level. The objective of this study was to investigate the results of the sentiment analysis based on the extraction of aspects that were carried out with attention models and several deep learning methods. Research data was collected from Zomato comments in Bahasa for any Indonesian restaurants. The data was annotated manually based on four subjects namely place, taste, location, and service. The result of this study showed that Bi-LSTM with attention model and CNN without attention model had the best performance compared to other methods, while CNN without attention model for sentiment analysis using deep learning showed the best accuracy.

***Key Words**:*
Long-Short Term Memory; Convolution Neural Network; Attention Model; Natural Language Processing.

Riwayat Artikel:
Received November 21, 2021
Revised November 10, 2023
Accepted November 10, 2023

DOI:
10.22441/incomtech.v13i3.13812

1. INTRODUCTION

A better understanding of opinions in natural language is one of many problems in sentiment analysis. To understand opinion, knowledge must be combined with methods of reasoning so that machine can analyze at context/intention level. Analysis at the context/intention level ensures that the relevance of opinions is collected. The extraction of opinions will be specific to the preferences and needs of each user or group of users. Opinions also grant users information regarding a product to help them make decisions prior to purchase.

Deep Learning methods such CNN and LSTM are able to classify and extract complex features [1]. However, CNN and LSTM have some problems for example when faced with missing information. Slang words and local language in Bahasa are two prominent problems in sentiment analysis [2] as well as the lack of research regarding the current state of art in Indonesia and insufficient Bahasa dataset to solve sentiment analysis problems presently.

Sentiment analysis or opinion mining in itself is a processing of human language and the extraction of information that gives the author's feeling in positive and negative comments, questions and requests, when analyzing a large number of documents of feeling [3]. An example of feeling analysis is to process several movie reviews to discover the success of the movie.

Deep learning is a subtopic of machine learning that produces great strength and flexibility when learning the real world as a multilevel concept, each concept is associated with a younger concept and represents abstract computation in less abstract terms [4].

Features become more complicated because of the combination of slang, micro-text and combination between Bahasa and English. To improve model stability, attention model proposed to keep context of sentences. This work contributed to creating a model that is capable of handling complex features. The purpose of this study was to investigate the results of the multi-classification sentiment analysis that were carried out with attention models and several deep learning methods.

Attention model is an improvement of Encode-Decoder model. Attention aims to capture value as humans do. E.g., if there is a job that requires manual translation of a long sentence from one language to another, at any given point the model will be focusing more on the specific word or phrase for translating, no matter where it lies in the input sentence. Attention recreates this mechanism for neural networks.

Previous research [5], identified aspects of aspects, extraction of expressions of expression and identification of polarity to perform feelings analysis based on aspects using the Deep Neural Convolutional Network (CNN) model. This research combines vector aspects of each word by applying convolution. The results showed that the system worked well in the Yelp review. However, the lack of this research has not been able to identify neutral cases. Therefore, exploration is still needed to improve the convolutional neuronal network (CNN) system.

In addition, the research [6] focuses on Twitter as the largest and most popular social network for predicting and analyzing feelings. The SNS (Social Network Service) on Twitter has very large data with user posts. In this study the methods used are, deep learning and automatic learning. The results obtained indicate that the machine learning method that uses neural networks of deep advance with many hidden layers produces 75 percent.

In the same year [7] he used the recursive neural network (RNN) algorithm. CNN and RNN have differences, that is, RNN backpropagation, the purpose of verifying errors when the process occurs, but CNN does not backpropagation. The consequence of backward propagation in itself is to aggravate the process of classification of feelings. The RNN method used in Wang's research is DT-RNN (Dependency-Tree RNN), RNCRF (Recursive Neural Conditional Random Fields).

In the study [8] they modeled the participation approach in the SINAI research group by conducting a sentiment analysis based on aspects for SemEval Shop 2015. This study proposes a syntax approach to identify words that modify each aspect, with the objective of classifying the feelings expressed to each entity attribute. In this study, an increase is still needed to determine the words that modify the Opinion Objective Expression (OTE) with a more complete denial.

The research [9] extends long-short term memory structured by chains to explicit structures with the S-LSTM model for language structures or image analysis. This study uses the model to study the representation of distributed sentiment for the text and shows that the model exceeds the advanced recursive model by replacing the reinforced composition layer with tensors improved by the S-LSTM memory block. The results show that the structure of the information is useful to help S-LSTM achieve sophisticated performance.

Research [10] conducted a comparison of the ME, NB and SVM aspects of the extraction models with case studies of Chinese hotels based on machine learning methods. For the extraction of aspects of the reviews of Chinese hotels, ME is the best automatic learning method compared to other methods of machine learning with all the methods of representation of characteristics. The results also show that noisy data can result in a decrease in accuracy. If the text includes the amount of noisy data, the accuracy will decrease as the dimensions of the function increase.

In addition, the research [11] introduces a successful deep learning approach to carry out the analysis of feelings that involves the learning of word insertion, the classification of feelings, the extraction of opinions and the learning of lexicons of feelings. To build a lexicon of feelings, in this study a deep learning approach was carried out that could not deduce the polarity of the feelings of the existing sentences.

In the study [12], aspects were extracted based on three methods of RNN, MV-RNN and RNTN with JMAS (joint feeling model of multiple aspects). The evaluation is carried out using a 10-fold cross validation in which the three methods are compared with CFACTS, SVM (TF-IDF) and NB (TF-IDF) using baselines for a single aspect and multiple aspects.

2. METHODS

The data set was taken from Zomato for five hours through scrapping. The scrapping stage was commenced with obtaining a link to Zomato by accessing https://www.zomato.com/jakarta/restaurants?page={} during the specified time; in the link, a list of restaurants that can take comments and the data was stored in the text extension for processing. Then the file was accessed to retrieve the rating and reviews on the html elements contained in the web page that was stored in the form of JSON (JavaScript Object Notation). Scraped data was eliminated with three criteria; data row has different polarity and sentiment meaning, data row has one word, data row is in English.

Annotation step adopts [13] method, scraping data from Zomato and was annotated by 3 reviewers and constructed with three classes such as positive, neutral, and negative with 14740 rows. All of labelled aspect data was aggregated with voting. If voting percentage resulted more than 66,67% then the data is used for experiment.

Data annotation was done manually with four subject annotations as follows: price, food, place, and service. Data was also labelled by three polarity such as: positive, neutral, negative.

Proposed method started with vectorizing sentence using attention model. The result of attention model was used for LSTM and CNN method as shown in Fig. 4. CNN layer used method as [4] proposed with five hyperparameter. This research related to [8] uses the convolutional neural network (CNN) when comparing the linguistic pattern (LP) method and the combination of CNN with LP, which has 7 layers (1 input layer, 2 convolution layers, 2 max grouping, 1 fully connected, 1 output layer). After this step performance and accuracy was evaluated with F-measure.

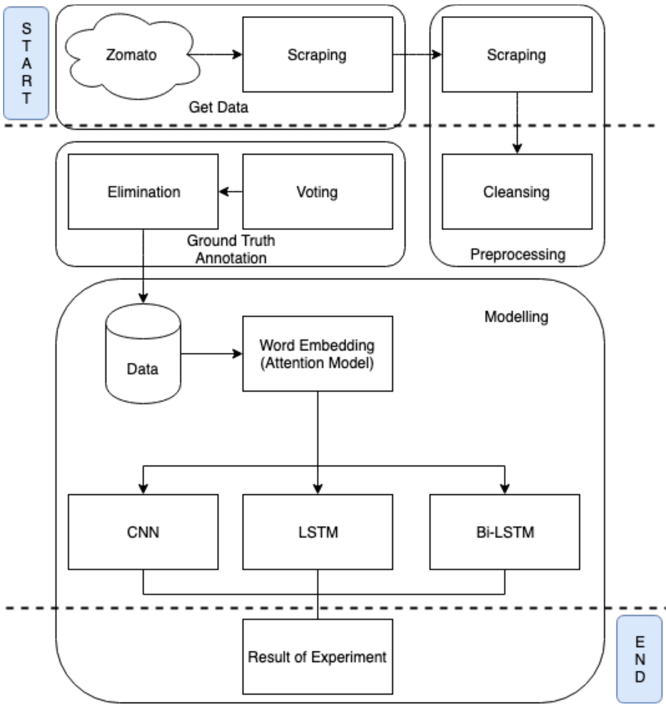

Figure 1. Research Workflow.

In machine learning modelling phase, this experiment stores every information weight of text using word embedding that processed by attention model, and then test to proposed methods such as CNN, LSTM, Bi-LSTM, compared to type of deep learning model combined with attention model.

Features inputted using 60-word inputs, method evaluation aimed for accuracy. Accuracy of all methods used F-measure training from epoch 1 until 200 with 70% training data and 30% testing data. CNN had 8 layers with one input layer, two convolutional layers, two max-pooling layers, one flatten-layer, one fully connected and one output layer. LSTM and Bi-LSTM had 128 units. These methods used ReLu activation function.

3. RESULTS AND DISCUSSION

The data was taken from Zomato by scraping the rating and review columns. In addition to the rating column and the revision, the "id" column is formed to facilitate processing, but the id starts from 0 as shown in Table 1.

Table 1. Scraped data from Zomato

id	rating	Review
0	4.0	pas masuk selera saya sate kambing nya lembut dan gurih. sop buntut nya empuk (piring nya kecil bener), bakso segar dan mantabnya banyak kerupuk (buat saya yg fanatik kerupuk sangat luar biasa) sashimi nya mantab walaupun pecking duck nya kurang pas buat saya tapi dessert-nya paling pas.
1	5.0	Makanan ini sangat enak, tempatnya bagus, sajiannya juga luar biasa.

Based on the data, a voting system was carried out on the data. It was found that each selected data had the same number, although a sentence could have more than one aspect. The amount of data that has been tagged in the voting system, among others, Food is 71371 lines, Price is 13777 lines, Place is 19424 lines and Service is 5837 lines.

Once the voting completed, data elimination is done if the sentence has more than one aspect, for example, "This meal is very good, the place is good, the dish is also extraordinary," which has aspects of food and place. Data elimination was also conducted if the aspects do not have the same minimum of two people. During elimination, the data also determined sentence polarity which divided into three categories namely negative, neutral, and positive. Elimination resulted in data reduction and the sentence polarity assignment can be seen in Table 2.

Table 2. Aspect Labelled Percent Agreement

Aspect	**Number of Rows**	**Negative**	**Neutral**	**Positive**
Food	56224	4315	10099	41810
Price	5644	496	985	4163
Place	12160	920	2201	9039
Service	3113	530	584	1999

As seen in Figure 2, the dataset is not balanced. In order to make the dataset more balanced, before continuing into model training, dataset will be balanced by SMOTE method.

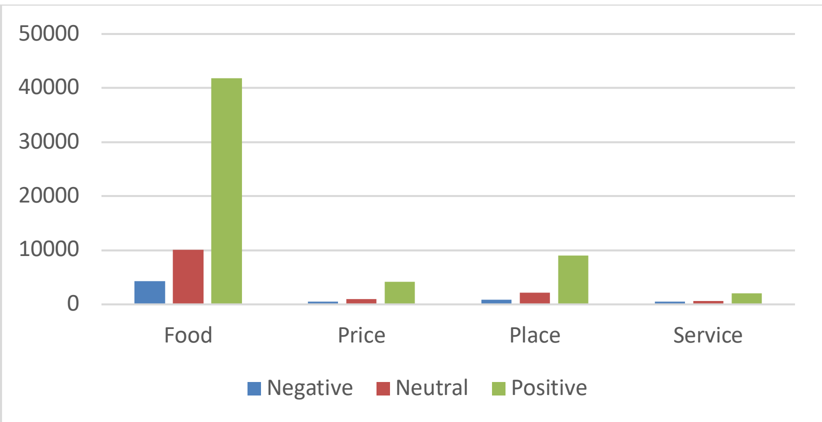

Figure 2. The comparation total of processed data

Graphs in figure 3 show fluctuation in results between the usage of CNN, LSTM, and Bi-LSTM with and without attention model. The highest increase of 0.05% can be seen for LSTM with attention model. Despite the increase with LSTM, the Bi-LSTM method experienced a slight decrease by 1%. However, the best method without using attention model is observed to be Bi-LSTM has gap 2% compared to LSTM. LSTM also proved to be the best method with attention model, having a significant average accuracy difference compared to LSTM by 88%.

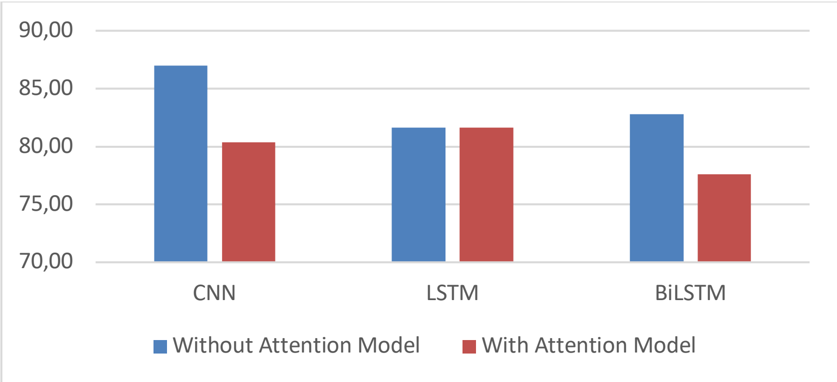

Figure 3. Accuracy of training models

As seen in Figure 4, it was found that the average loss increased and decreased between the attention model than before, with the highest increase that occurred in CNN from 1.0779 to 2.8542. The largest decline occurred in LSTM, decreasing from 1.0942 to 0.875655, while major loss didn't occur for Bi-LSTM that showed decrease only from 0.8415 to 0.8288. Figure 7 also illustrates that the method with the lowest loss before using the attention model was the LSTM showing value of 0.8415, while the highest loss occurred in LSTM by 1.0942. After the use of attention model, the lowest loss remained in Bi-LSTM with the final result of 0.8288, while CNN held the highest loss with 2.8542.

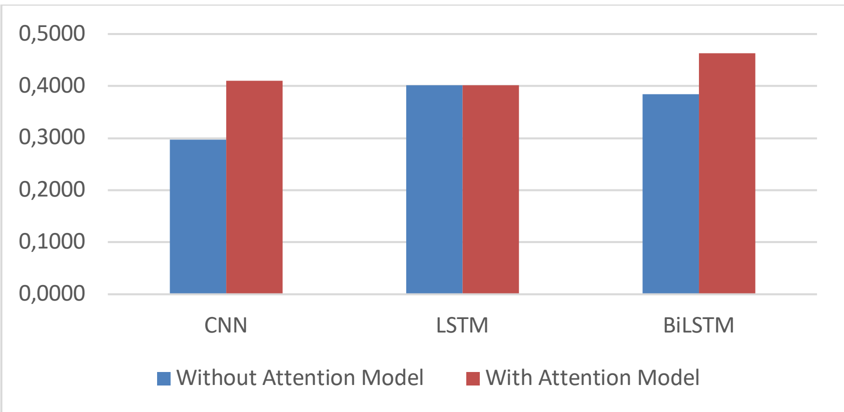

Figure 4. Average validation loss of training models

From Table 3, the best evaluation model is found in the LSTM with an accuracy value of 75.9%, followed by Bi-LSTM with attention model with accuracy value of 79.35%. But the smallest loss was obtained by CNN with attention with loss value of 0.58 followed by Bi-LSTM with attention model of 0.60. Convergent between machine learning model has slightly different between CNN with attention model than others.

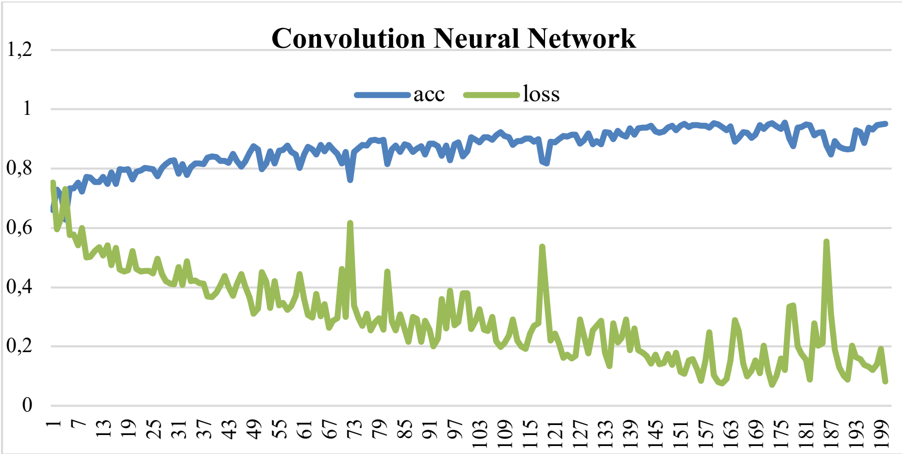

Figure 5. CNN without attention model training accuracy and loss

Meanwhile CNN method as can be seen in figure 5, has no convergent from start training until epoch iteration finished. Compared to figure 6, with attention model CNN method can get convergence at 183 epochs. This case not be found on LSTM and Bi-LSTM with or without attention model.

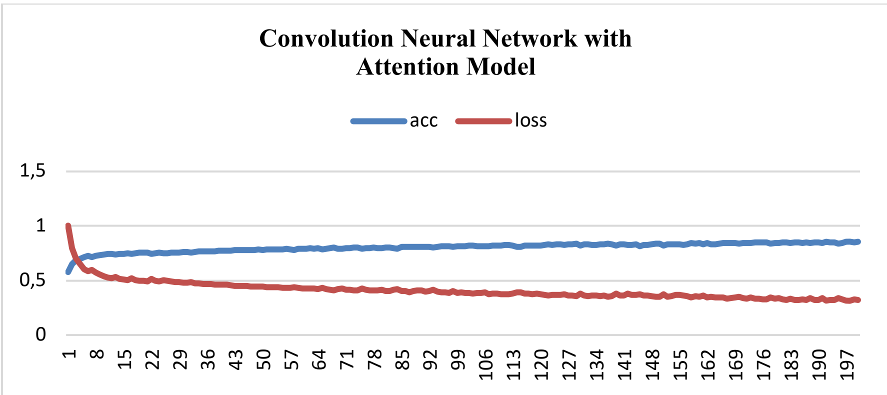

Figure 6. CNN without attention model training accuracy and loss

Table 3. Accuracy and Loss Evaluation Models

Model	**Loss**	**Accuracy (%)**	***Convergent***
CNN	1.12	70.00	-
CNN Attention Model	0.58	77.48	183
LSTM	0.77	79.35	186
LSTM Attention Model	2.03	77.19	186
Bi-LSTM	0.63	71.64	186
Bi-LSTM Attention Model	0.60	77.51	186

Evaluation models that can be seen in Figure 7 showed increases and decreases in accuracy. The decrease in accuracy occurred in CNN by 5.1%, while the highest increase in accuracy of 4.9% can be seen in Bi-LSTM. Before using attention model, the best method was shown to be CNN while Bi-LSTM was displayed to be the method with worst accuracy. After attention model was implemented, the method with the best precision turned out to be Bi-LSTM with CNN being the least accurate method.

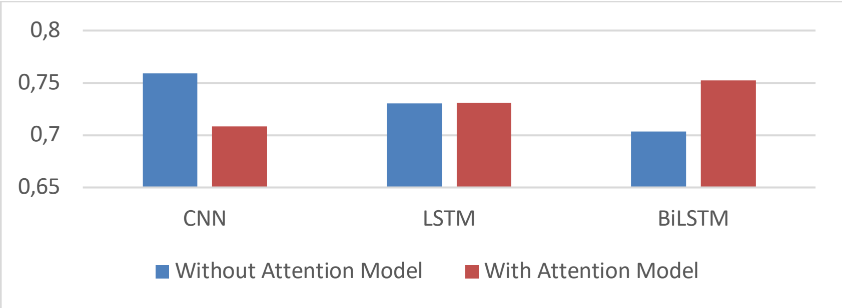

Figure 7. Accuracy of evaluation models

Based on the results of the loss value evaluation model that can be seen in Figure 8, it can be found that there was an increase and a decrease between the use of attention models than before. The increase in loss occurred in CNN from 1.1 to 2.5, while the largest decrease occurred in Bi-LSTM from prior value of 1.12 to 0.60. The method with the least loss before using the attention model was CNN, while the method with the highest loss was Bi-LSTM. After implementing attention model, the method with the least loss was Bi-LSTM, while the highest loss result can be seen in CNN method.

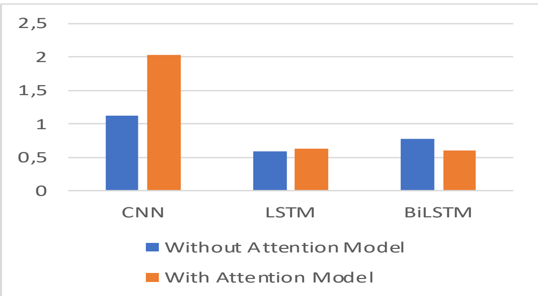

Figure 8. Validation loss of evaluation models

In Figure 9 of the averages of existing steps, it was found that there was a slowdown with the use of attention model compared to without attention model. The lowest deceleration occurred in CNN method where the difference was 900.7 µs, while the highest deceleration occurred in LSTM of 1100 µs. Data observation after using attention model showed that the fastest method is CNN, while the slowest method is Bi-LSTM with an average time of 2000 µs/step. The fastest method after using the attention model is CNN with 998,857,143 µs/step, while the slowest method was Bi-LSTM with attention model with an average time of 300 µs/step.

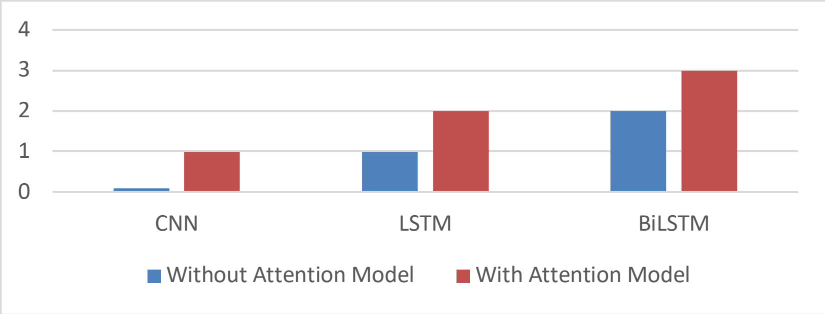

Figure 9. Average Step over Training with 3 Models

Figure 10 showed that there was a slowdown in average processing time after the use of attention model compared to without attention model. The method with the best deceleration was LSTM with a time of 122.18 seconds compared to BiLSTM with a time of 180.75 seconds. Without attention model, the fastest method was CNN with the average time of 15.6 s, while the slowest method was BiLSTM with average time 311.55 s. The fastest method with attention model was CNN with average time of 157.52 s, while the slowest method was BiLSTM with attention model, consuming average processing time of 492.6 s.

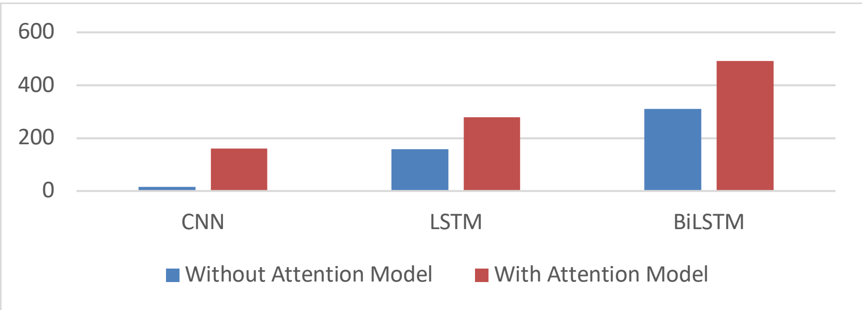

Figure 10. Average Processed over Training with 3 Models

4. CONCLUSION

Based on this study, it can be concluded that attention models relative improved stability model and has trade-off for accuracy because of noise data filtered by it. It was also found that the best accuracy performance after adapting attention model occurred in LSTM method. Although attention model can improve accuracy and loss, but it resulted in deceleration of processing time. Model can adapt with typo and microtext for classification. In the future, there must be comparative method between transformer methods with neural network methods to get insights which model more stable and accurate to do multi-classification sentiment analysis.

REFERENSI

[1] J. Barnes, R. Klinger, and S. S. im Walde, "Assessing State-of-the-Art Sentiment Models on State-of-the-Art Sentiment Datasets," pp. 2–12, 2017.

[2] L. Xu, J. Lin, L. Wang, C. Yin, and J. Wang, "Deep Convolutional Neural Network based Approach for Aspect-based Sentiment Analysis," Adv. Sci. Technol. Lett., vol. 143, pp. 199–204, 2017.

[3] G. Vinodhini and R. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey,"

Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 2, no. 6, pp. 282–292, 2012.
[4] I. Goodfellow, "Deep Learning."
[5] S. Poria, E. Cambria, and A. Gelbukh, "Aspect Extraction for Opinion Miningwith a Deep Convolutional Neural Network," Knowledge-Based Syst., vol. 108, pp. 42–49, 2016.
[6] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in 2017 7th International Annual Engineering Seminar (InAES), 2017, pp. 1–4.
[7] S. M. Jiménez-Zafra, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ur Na-López, "SINAI: Syntactic approach for Aspect Based Sentiment Analysis," Semeval 2015, no. SemEval, pp. 730–735, 2015.
[8] X. Zhu, P. Sobhani, and H. Guo, "Long Short-Term Memory Over Tree Structures," Int. Conf. Mach. Learn., no. Icml, Mar. 2015.
[9] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 23, no. 12, pp. 2111–2124, 2015.
[10] D. Tang, F. Wei, B. Qin, M. Zhou, and T. Liu, "Building Large-Scale Twitter-Specific Sentiment Lexicon: a Representation Learning Approach," Proc. 25th Int. Conf. Comput. Linguist. (COLING 2014), pp. 172–182, 2014.
[11] H. Lakkaraju, R. Socher, and C. D. Manning, "Aspect Specific Sentiment Analysis using Hierarchical Deep Learning," NIPS WS Deep neural networks Represent. Learn., pp. 1–9, 2014.
[12] L. Zhang and B. Liu, Data Mining and Knowledge Discovery for Big Data, vol. 1. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.
[13] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis," Proc. 2016 Conf. Empir. Methods Nat. Lang. Process., pp. 616–626, 2016.
[14] Bobicev, V., & Sokolova, M. (2017). Inter-Annotator Agreement in Sentiment Analysis: Machine Learning Perspective. Proceedings of Recent Advances in Natural Language Processing, 97–102. https://doi.org/10.26615/978-954-452-049-6_015
[15] Cambria, E., Gelbukh, A., & Thelwall, M. (2017). Affective Computing and Sentiment Analysis. IEEE INTELLIGENT SYSTEMS, 32(6), 74–80. https://doi.org/10.1109/MIS.2017.4531228
[16] Covington, M. A. (1994). Natural Language Processing for Prolog Programmers. New Jersey: Prentice Hall.
[17] Kao, A., & Poteet, S. R. (2007). Natural Language Processing and Text Mining. USA: Springer.
[18] Kemp, S. (2017, August 10). Three Billion People Now Use Social Media. Retrieved from https://wearesocial.com/blog/2017/08/three-billion-people-now-use-social-media.
[19] Lau, J. H., & Baldwin, T. (2016). An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In Proceedings of the 1st Workshop on Representation Learning for NLP (pp. 78–86). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-1609
[20] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011
[21] Mitchell, T. M. (1997). Machine Learning. New York: McGraw Hill.
[22] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.
[23] Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, 42–49. https://doi.org/10.14257/astl.2017.143.41.