

Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding

Putri Rizki Amalia*¹, Edi Winarko²

¹Bachelor Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia

²Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *¹putririzki99@mail.ugm.ac.id, ²ewinarko@ugm.ac.id

Abstrak

Opini seseorang terhadap suatu produk atau layanan yang dituangkan melalui sebuah ulasan merupakan suatu hal yang cukup penting bagi pemilik ataupun calon pelanggan. Namun, banyaknya jumlah ulasan menyulitkan mereka untuk menganalisis informasi yang terdapat pada ulasan-ulasan tersebut. Analisis sentimen berdasarkan aspek merupakan suatu cara untuk mengetahui polaritas sentimen suatu kalimat opini berdasarkan aspek yang telah ditentukan.

Penelitian ini bertujuan untuk menganalisis suatu ulasan restoran berbahasa Indonesia dengan menggunakan kombinasi model Convolutional Neural Network dan Contextualized Word Embedding. Lalu akan dibandingkan dengan kombinasi model Convolutional Neural Network dan Traditional Word Embedding. Pengujian klasifikasi aspek terhadap tiga model; BERT-CNN, ELMo-CNN, dan Word2vec-CNN, memberikan hasil terbaik pada model ELMo-CNN dengan nilai nilai micro-average precision sebesar 0.88, micro-average recall sebesar 0.84, dan micro-average f1-score sebesar 0.86. Sedangkan untuk klasifikasi sentimen memberikan hasil terbaik pada model BERT-CNN dengan nilai precision sebesar 0.89, recall sebesar 0.89, dan f1-score sebesar 0.91. Klasifikasi yang dilakukan dengan data tanpa stemming memiliki hasil yang hampir mirip bahkan lebih baik dibandingkan menggunakan data dengan stemming.

Kata kunci— *analisis sentimen berdasarkan aspek, ulasan restoran, CNN, BERT, ELMo, Word2vec*

Abstract

Someone's opinion on a product or service that is poured through a review is something that is quite important for the owner or potential customer. However, the large number of reviews makes it difficult for them to analyze the information contained in the reviews. Aspect-based sentiment analysis is the process of determining the sentiment polarity of a sentence based on predetermined aspects.

This study aims to analyze an Indonesian restaurant review using a combination of Convolutional Neural Network and Contextualized Word Embedding models. Then it will be compared with a combination of Convolutional Neural Network and Traditional Word Embedding models. The result of aspect-classification on three models; BERT-CNN, ELMo-CNN, and Word2vec-CNN give the best results on the ELMo-CNN model with micro-average precision of 0.88, micro-average recall of 0.84, and micro-average f1-score of 0.86. Meanwhile, the sentiment-classification gives the best results on the BERT-CNN model with a precision value of 0.89, a recall

of 0.89, and an f1-score of 0.91. Classification using data without stemming have almost similar results, even better than using data with stemming.

Keywords— *aspect-based sentiment analysis, restaurant review, CNN, BERT, ELMo, Word2vec*

1. INTRODUCTION

In this digital era, it is easier to express our opinion on something in the form of status, tweets, reviews, etc. As a result, there are also platforms that can accommodate these opinions in text form such as Zomato, Tripadvisor, Traveloka, Qraved that contain various places, especially restaurants, in various regions in Indonesia and users can express their opinions about the restaurant in a text called a review. Reviews to a restaurant is one of the important things that is very useful for potential customers and restaurant owners. The information contained in a review can be used by customers to identify more about the products or services available at the restaurant and for the owner it can be used as an evaluation to meet the needs of customers [14].

Positive reviews from customers can increase the impression of potential customers to a restaurant so they will be interested in visiting the restaurant after seeing positive reviews from other customers. On the other hand, negative reviews can be taken into consideration for potential customers when visiting the restaurant itself. However, with the large number of reviews available, potential customers sometimes only read them briefly which usually appear at the top, making it difficult to make conclusions and make decisions. Therefore, a solution is needed to deal with this, namely by building a model that can classify a review into positive or negative classes. Sentiment analysis is one method that can solve this problem.

Aspect-based sentiment analysis (ABSA) is a solution to the problem in sentiment analysis that is not able to classify a review into positive or negative classes based on aspect categories. ABSA plays a role in classifying a review into positive or negative classes based on the aspect category (Liu, 2015). At this level, a model will be able to classify a text document into the aspect category and the polarity of the sentiment. For example, in the review sentence "The food is delicious but the price is quite expensive", ABSA model is able to classify the sentence "the food is delicious" into food aspects and positive classes, then the sentence "the price is quite expensive" into price aspects and negative classes.

Research on the aspect-based sentiment analysis has been conducted by [7]. In this study, they are using Conditional Random Fields (CRF) and Maximum Entropy (MaxEnt) models and give an average F1-score of 0.642. [1] continued the research of [7] using the deep learning Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) methods resulting in increased performance with an average F1-score of 0.764. The research conducted by [2] using a combination of the CNN-GRU and word2vec models resulting in an F1-score of 0.67 for aspect classification and 0.66 for sentiment classification. However, [5] and [13] show a better results using a combination of contextualized word embedding for the model.

Based on the background description that has been explained, this study conducted an aspect-based sentiment analysis on Indonesian restaurant review using a combination of Convolutional Neural Network and Contextualized Word Embedding approach namely BERT and ELMo. Then, they will be compared with the model using Traditional Word Embedding approach namely Word2vec. This study will also test the effect of using stemming in the preprocessing on the three models of each classification.

2. METHODS

2.1 Data Collection

Data used in this research are reviews of Indonesian restaurant from [2]. There are four categories of aspects that will be used; food, service, price, and place.

2.2 Data Labelling

Data is manually labelled into aspects mentioned in 2.1.

- 1) Food: food tastes great, have a variety of menus, delicious dishes, etc.
- 2) Service: great service, staff are extremely knowledgeable, staff is really friendly, etc.
- 3) Price: price is reasonable, quite expensive for the price, etc.
- 4) Place: cozy restaurant, the ambience is welcoming and charming, such a lovely place, etc.

2.3 Data Splitting

Data split ratio used in this research is 80% as the train set, 15% of the train set as the validation set, and 20% as the test set. So, 2400 reviews used, 1632 reviews became the train set, 288 reviews became the validation set, and 480 reviews became the test set.

2.4 Preprocessing

The preprocessing steps in this research are; 1) case folding, 2) non-alphanumeric removal, 3) normalization, 4) stemming. At this stage, a review changed into a lower case, non-alphanumeric removal used to delete characters that non-alphanumeric, normalization is used to normalize words that has non-standards word, abbreviated word, and typo. Meanwhile, stemming is used to changing a word into its root word.

2.5 Feature Extraction

We use the pre-trained BERT embedding in Bahasa Indonesia: IndoBERT [6]. The pre-trained embedding is used as is, and not fine-tuned in downstream task, as it was trained in the same corpus domain as our review data. For comparison, we also used the pre-trained ELMo embedding ELMoForManyLangs [18] in Bahasa Indonesia, and Word2vec embedding uses corpus of our data so that the resulting word insertion can cover all the words contained in the data review.

2.6 Classification Model

Three models experimented; BERT-CNN, ELMo-CNN, and Word2vec-CNN for each of aspect classification and sentiment classification has four layers; convolutional layer, pooling layer, dropout layer, and fully-connected layer. Hyperparameter tuning using Keras-Tuner Hyperband model [16] is use to get the model with the best combination of parameters. This process is carried out on each model of aspect classification and sentiment classification to provide fair results. There are four parameters tested; convolutional filters, kernel size, pool size, and dense units

2.6.1 Aspect Classification

The model architecture for aspect classification can be seen in Figure 1. The preprocessed data is converted into vector form using contextualized word embedding and used as an input to the CNN model. The CNN model used has several layers:

- 1) Input layer: input in the form of a word vector resulting from feature extraction from the embedding layer which has a size of $x \times y \times z$ (x = number of reviews, y = number of hidden layers, z = number of tokens). Padding is done at the end of each review.
- 2) ConvPool layer: convolution and max-pool processes will be performed on this layer. In the convolution layer, we will try to use several combinations of filters, 128 and 256 filters and use the ReLU activation function. In the pooling layer, we will try to use kernel sizes (3×3) and (5×5) as well as pool sizes (2×2) , (3×3) , and (4×4) .
- 3) Dropout layer: regularization is performed on the ConvPool layer output to prevent overfitting and with a rate of 0.5.
- 4) Fully-connected layer: the output from the dropout layer will be flattened and then put into a fully-connected layer consisting of two layers. The first layer uses the ReLU activation function with 128 and 256 units of experimentation. The second layer uses four filters representing each label with a sigmoid activation function so as to get a binary representation.

2.6.2 Sentiment Classification

The model architecture for sentiment classification can be seen in Figure 2. The process is almost similar to the aspect classification, one of the things that distinguishes it is the data used as an input in the form of sentiment data for each aspect that has been separated based on the aspect category. Also, for the second layer of fully-connected uses a sigmoid activation function with one output neuron that produces output in the form of real numbers between 0-1 which can be subject to rules: if the result of the sigmoid function is more than or equal to 0.5 it will output 1, which means the polarity of the sentiment is positive and if the result of the function is sigmoid is less than 0.5 it will output 0, which means the polarity of the sentiment is negative.

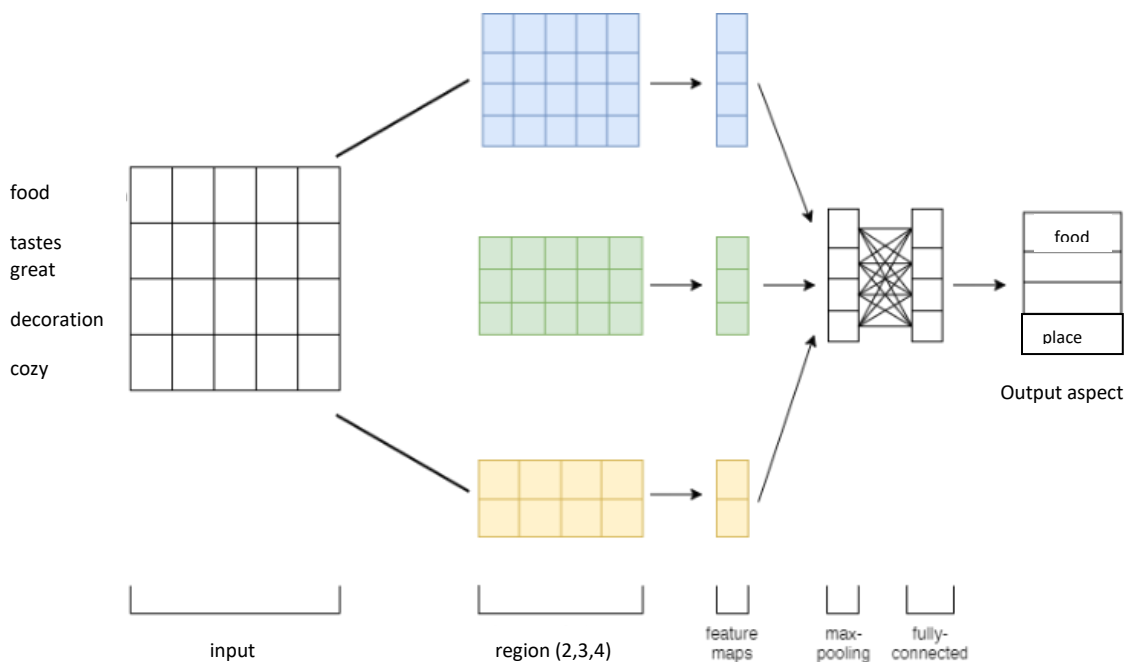


Figure 1. CNN architecture for aspect classification

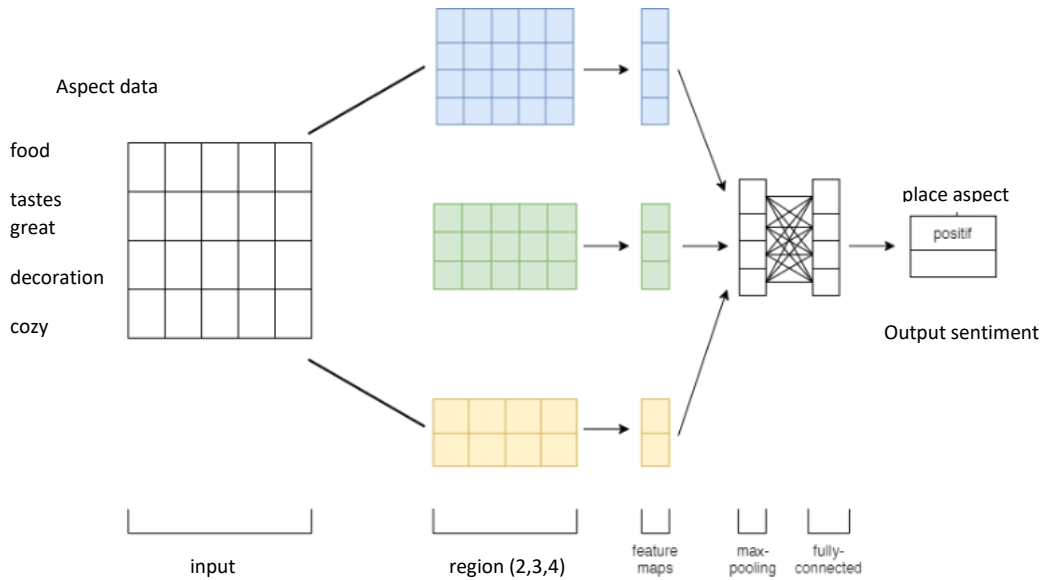


Figure 2. CNN architecture for sentiment classification

2.7 Model Evaluation

To evaluate the model we will use a label-based evaluation. Label-based evaluation is based on confusion matrix that can be seen in Figure 3, where TP is true positive which is data labelled True that is predicted True, FP is false positive which is data labelled False that is predicted True, FN is true negative which is data labelled True that is predicted False, and TN is true negative which is data labelled False that is predicted False.

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

Figure 3. Confusion matrix

The evaluation metrics that we will use are:

- 1) Micro-Averaged Precision

$$Precision_{Micro} = \frac{\sum TP}{\sum TP + \sum FP} \tag{1}$$

- 2) Micro-Averaged Recall

$$Recall_{Micro} = \frac{\Sigma TP}{\Sigma TP + \Sigma FN} \quad (2)$$

3) Micro-Averaged F-measure

$$F1_{-Micro} = 2 \times \frac{Precision_{Micro} \times Recall_{Micro}}{Precision_{Micro} + Recall_{Micro}} \quad (3)$$

3. RESULTS AND DISCUSSION

3.1 Data Exploration Results

Exploring into the collected and labelled data, we found the label distribution for the aspect data that can be seen in Table 1 and for the sentiment data that can be seen in Table 2.

Table 1 Aspect data per label

Label	Food	Price	Place	Service
Yes	1472	978	1139	960
No	928	1422	1261	1440

Table 2 Sentiment data per label

Label	Food	Price	Place	Service
Positif	978	478	316	604
Negatif	494	500	823	356

3.2 Aspect Classification Results

The aspect classification results can be seen in Table 3. It can be seen that the model using a combination of contextualized word embedding have better results compared to the model using a combination of traditional word embedding. What's interesting is ELMo-CNN model gave better results of micro-averaged F1-score than the BERT-CNN model.

The ELMo-CNN model which gives the best results in micro-averaged f1-score is using hyperparameters as follows: convolutional filters of 128, kernel size of 4, pool size of 3, dense units of 256.

Table 3 Aspect classification results

	BERT-CNN	ELMo-CNN	Word2vec-CNN
<i>Micro-averaged Precision</i>	0.80	0.88	0.55
<i>Micro-averaged Recall</i>	0.90	0.84	0.70
<i>Micro-averaged F1-score</i>	0.85	0.86	0.61

Meanwhile, the aspect classification using stemming data can be seen in Table 4. It can be seen compared to the model that is not using stemming the results are lower. This is because the stemming process can cause the text to lose context due to the deletion and simplification of some words.

Table 4 Aspect classification results using stemming data

	BERT-CNN	ELMo-CNN	Word2vec-CNN
<i>Micro-averaged Precision</i>	0.79	0.82	0.60
<i>Micro-averaged Recall</i>	0.82	0.91	0.53
<i>Micro-averaged F1-score</i>	0.80	0.86	0.56

3.3 Sentiment Classification Results

The sentiment classification results can be seen in Table 5. It can be seen that the model using a combination of contextualized word embedding have better results on each aspect compared to the model using a combination of traditional word embedding.

The BERT-CNN model which gives the best results in micro-averaged f1-score is using hyperparameters for each label as follows:

- 1) Food: convolutional filters of 128, kernel size of 4, pool size of 2, dense units of 128.
- 2) Price: convolutional filters of 256, kernel size of 5, pool size of 2, dense units of 128.
- 3) Place: convolutional filters of 256, kernel size of 4, pool size of 2, dense units of 256.
- 4) Service: convolutional filters of 256, kernel size of 3, pool size of 2, dense units of 128.

Meanwhile, the aspect classification using stemming data can be seen in Table 6. It can be seen that sentiment classification for models with a combination of contextualized word embedding with data without stemming or with stemming does not have a significant effect, while for the models with a combination of traditional word embedding, the use of data with stemming is having a significant effect, that can be seen from the difference in the average f1-score. This is due to the use of word2vec as a feature extraction between words that have affixes and basic words that have a fairly distant meaning.

Table 5 Sentiment classification results

		BERT-CNN	ELMo-CNN	Word2vec-CNN
<i>Precision</i>	Food	0.89	0.80	0.70
	Price	0.80	0.77	0.71
	Place	0.92	0.92	0.87
	Service	0.94	0.78	0.77
	Average	0.89	0.82	0.76
<i>Recall</i>	Food	0.90	0.85	0.58
	Price	0.84	0.70	0.71
	Place	0.93	0.83	0.85
	Service	0.90	0.86	0.78
	Average	0.89	0.81	0.73
<i>F1-score</i>	Food	0.94	0.89	0.83
	Price	0.82	0.79	0.61
	Place	0.94	0.91	0.89
	Service	0.94	0.82	0.82
	Average	0.91	0.85	0.79

Table 6 Sentiment classification results using stemming data

		BERT-CNN	ELMo-CNN	Word2vec-CNN
<i>Precision</i>	Food	0.90	0.83	0.65
	Price	0.79	0.77	0.56
	Place	0.91	0.93	0.72
	Service	0.89	0.78	0.57
	Average	0.87	0.83	0.63
<i>Recall</i>	Food	0.84	0.76	0.75
	Price	0.81	0.81	0.56
	Place	0.92	0.93	0.69
	Service	0.91	0.85	0.58
	Average	0.87	0.84	0.65
<i>F1-score</i>	Food	0.93	0.89	0.68
	Price	0.82	0.79	0.60
	Place	0.94	0.95	0.79
	Service	0.93	0.82	0.68
	Average	0.91	0.86	0.69

CONCLUSION

The use of a combination of Convolutional Neural Network and Contextualized Word Embedding gave better results compared to a model with a combination of Convolutional Neural Network and Traditional Word Embedding. For aspect classification, ELMo-CNN model gave the best results with a micro-average precision of 0.88, a micro-average recall of 0.84, and a micro-average f1-score of 0.86. For sentiment classification, BERT-CNN model gave the best results with a precision of 0.89, a recall of 0.89, and an f1-score of 0.91.

The use of data without stemming in the BERT-CNN, ELMo-CNN, and Word2vec-CNN models gives almost similar results or even better than the use of data with stemming. However, a significant difference was shown in the Word2vec-CNN model.

However, in this study, there are still shortcomings that can be added such as a better data labelling method can be employed. Other than that, using more diverse and balance data given the use of contextualized word embedding is able to work better for large data.

REFERENCES

- [1] A. Cahyadi, and M. Khodra, "Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory," 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), pg.124-129, 2018.
- [2] A. Ilmania, Abdurrahman, S. Cahyawijaya, and A. Purwarianti, "Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis," 2018 International Conference on Asian Language Processing (IALP), pg.62-67, 2018

- [3] A. Rahmah, "Analisis sentimen berdasarkan aspek pada ulasan restoran berbahasa indonesia menggunakan convolutional neural network (CNN) dan gated recurrent unit (GRU)," Skripsi. Program Studi Ilmu Komputer, FMIPA UGM, Yogyakarta, 2020.
- [4] B. Liu, "Sentiment Analysis and Opinion Mining," Cambridge: Cambridge University Press, 2015.
- [5] B. Talafha, M. Al-Ayyoub, A. Abuammar, and Y. Jararweh, "Outperforming State-of-the-Art Systems for Aspect-Based Sentiment Analysis," In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), pg. 1-5, 2019.
- [6] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti. 2020. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing.
- [7] D. Ekawati, and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," In 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), pg. 1-6, 2017.
- [8] H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," Expert Systems with Applications, 118, pg.272-299, 2019.
- [9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
- [10] J. Han, J. Pei, and M. Kamber, "Data mining: concepts and techniques," Elsevier, 2011.
- [11] M. Bangsa, 2019. "Analisis sentimen berbasis aspek pada ulasan toko online menggunakan convolutional neural network," Tesis. Program Studi Ilmu Komputer, FMIPA UGM, Yogyakarta, 2019.
- [12] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," arXiv preprint arXiv:1802.05365, 2018.
- [13] M. Trușcă, D. Wassenberg, F. Frasincar, F. and R. Dekker, "A Hybrid Approach for Aspect-Based Sentiment Analysis Using Deep Contextual Word Embeddings and Hierarchical Attention," 2020 International Conference on Web Engineering, pg.365-380, 2020
- [14] N. Jihan, Y. Senarath, and S. Ranathunga, "Aspect Extraction from Customer Reviews Using Convolutional Neural Networks," 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), pg.215-220, 2018.
- [15] R. Horev, "BERT Explained: State of the art language model for NLP," <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>, 2018. [accessed: 3-Dec-2020]
- [16] S. Wadekar, "Hyperparameter Tuning in Keras: TensorFlow 2: With Keras Tuner: RandomSearch, Hyperband, BayesianOptimization," <https://medium.com/swlh/hyperparameter-tuning-in-keras-tensorflow-2-with-keras-tuner-randomsearch-hyperband-3e212647778f>, 2021. [accessed 14-Apr-2021]
- [17] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, et al. "Keras Tuner," <https://github.com/keras-team/keras-tuner>, 2019.
- [18] W. Che, Y. Liu, Y. Wang, B. Zheng, and T. Liu, "Towards Better UD Parsing: Deep Contextualized Word Embeddings, Ensemble, and Treebank Concatenation," In Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pg. 55–64. 2018.

- [19] X. Ouyang, P. Zhou, C. Li, and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pg.2359-2364, 2015.
- [20] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.