Aspect-based Opinion Mining for Code-Mixed Restaurant Reviews in Indonesia

Andi Suciati
Faculty of Computer Science
Universitas Indonesia
Indonesia
andi.suciati@ui.ac.id

Indra Budi
Faculty of Computer Science
Universitas Indonesia
Indonesia
indra@cs.ui.ac.id

***Abstract*—The goal of opinion mining is to extract the sentiment, emotions, or judgement of reviews and classified it. These reviews are very important because they can affect the decision-making from a person. In this paper, we conducted an aspect-based opinion mining research using customer reviews of restaurants in Indonesia and we focused into analyzing the code-mixed dataset. The evaluation conducted by making four scenarios namely removing stopwords without stemming, without removing stopwords but with stemming, without removing stopwords and stemming, and preprocessing with removing stopwords and stemming. We compared five algorithms which are Random Forest (RF), Multinomial Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), and Extra Tree classifier (ET). The models were evaluated by using 10 folds cross validation, and the results show that all aspects achieved highest scores with different algorithms. LR achieved highest score for *food* (81.76%) and *ambience* (77.29%) aspects while the highest score for *price* (78.71%) and *service* (85.07%) aspects were obtained by DT.**

Keywords-opinion mining; restaurant; code-mixed; stemming; stopwords

I. INTRODUCTION

With the rapid development of technology in recent years, the massive amount of data can be found on internet. In past, it is difficult to know what people think about something, for instance a product. However, at this moment, people can express and write their reviews easily. These reviews are very important because they can affect the decision-making from a person. To illustrate, when people are going to buy products, they will check the reviews or the ratings of those products in online websites, such as Amazon[1] and eBay[2] , before buying them. People will choose the product based on the positive sentiments in its reviews. To learn whether the product has many positive, negative, or neutral reviews, the website needs the system that can classified the sentiment of those reviews.

In NLP, there is a task called opinion mining or sentiment analysis. Opinion mining is task for detecting public sentiment towards entities. The goal of opinion mining is to extract the sentiment, emotions, or judgement of reviews and classified it. Usually, a review contains sentiments that can be classified into different polarities. Initial research conducted by Turney [15] in 2002, classified reviews into positive and negative polarities. In SemEval-2015 Task 12, Pontiki et al. [13], divided the polarities into positive, negative, and neutral, while Ganu et al. [9] divided them into positive, negative, conflict, and neutral. Besides, he also classified the aspect of the restaurant into *food*, *service*, *price*, *ambience*, *anecdotes*, and *miscellaneous.*

There are several works that have been done about sentiment analysis or opinion mining task in various domains. In 2018, Zvarevashe and Olugbara [12] investigated four classification algorithms for hotel domain. Amrania et al. [19] proposed hybrid approach which are Random Forest and Support Vector Machine to identify sentiment of product reviews by Amazon. For movie domain, Yasen and Tedmori [14] used dataset movie reviews from IMBD and compared eight classifiers for classifying the sentiment of the reviews. Next, opinion mining task also important in Indonesia, as a country that has several internet companies that develop in it, according to Mustafa and Budi [1]. There are also several studies about opinion mining that have been conducted in Indonesia. For instances, Alfina et al. [10] conducted an experiment of sentiment analysis in political domain, Putra et al. [18] observed the task in government domain, Jaya et al. [16] analyzed two different domains (president election and online store), and Fiarni et al. [4] proposed rule-based and Naïve Bayes algorithm to classify the sentiment of online transportation *service* reviews.

In this paper, we conducted an aspect-based opinion mining research using customer reviews of restaurants in Indonesia. The aspect we used are *food*, *price*, *service*, and *ambience* and we divided the polarities into positive, negative and neutral. In Indonesia, there are several websites that can be accessed to see reviews from customers, for instance TripAdvisor [3] and Zomato [4] . However, in this research, we used data that obtained from Indonesian culinary review website namely PergiKuliner[5]. In addition, the dataset is using code-mixed language (Indonesian and English).

The rest of this paper is arranged as follows: In section 2, we review the related works with our study. We describe the research steps that applied in this work and the result from that we obtained in section 3. In section 4, we conclude our results.

II. RELATED WORK

Sentiment analysis or opinion mining using code mixing reviews is also gaining attention as a research study. Shalini et. al [11] conducted a study to analyze the

[1] https://www.amazon.com/
[2] https://www.ebay.com/
[3] https://www.tripadvisor.com/
[4] https://www.zomato.com
[5] https://pergikuliner.com/

sentiment of code-mixed Kannada-English, Bengali-English, and Hindi-English. After that, they annotated their data manually into three polarities which are positive, negative, neutral and compared the performances of Doc2Vec with SVM, Fasttext, Bi-LSTM, and CNN. The study shows that CNN gained accuracy 71.50% for Kannada (India)-English dataset while Bi-LSTM achieved 60.20% and 72.20% for Bengali (India)-English and Hindi-English respectively. In 2017, Pravalika et al. [3] did a research about sentiment analysis in movie domain using code-mixed Hindi-English and then implemented two approaches which are lexicon-based and machine learning approaches. The lexicon-based approach gained 86% accuracy while machine learning attained about 72%. However, both of studies were not aspect-based research.

In Indonesia, there are few studies that performed opinion mining task for restaurant domain. Sasmita et al. [7] proposed unsupervised aspect-based sentiment analysis method which was divided by two main tasks (aspect extraction and aspect sentiment orientation classification). The method achieved 88.40 for F1-measure score. Ekawati and Khodra [6] implemented the method from best research in SemEval 2016 for Indonesian restaurant reviews and achieved 0.793, 0.823, and 0.642 F1-measure scores for combination of feature in aspect extraction, aspect categorization, and sentiment classification respectively. To improve the study conducted by Ekawati and Khodra [6], Cahyadi and Khodra [2] performed aspect-based sentiment analysis study for restaurant domain by using Bidirectional Long-Short Term Memory, Conditional Random Field, and Convolutional Neural Network. The scores they obtained are 87.0%, 76.4%, and 78.7% respectively. Those studies for Indonesian reviews were getting promising scores, but all the datasets are in Indonesian. Besides, they divided the polarities into positive and negative only while the sentiment of review can be mildly positive and mildly negative which is classified to neutral according to [13].

III. RESEARCH STEPS

This section will describe the research steps that applied in this study. This work consists 5 important steps as shown in figure 1. The first is collecting the data from the website, the second step is applying few language preprocessing techniques, the third is feature extraction, the fourth is experiment with machine learning models, and the last step was evaluating the models.

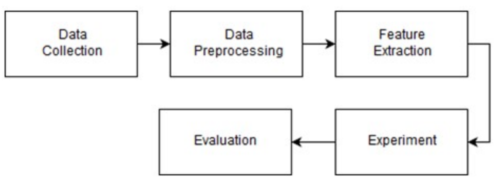

Figure 1. Research steps of opinion mining.

A. Data Collection

In this step, we collected 20000 reviews by scrapping PergiKuliner website. The data contain reviews using Indonesian, English, and mixed (Indonesian & English) languages. The example of the data that retrieved as follow:

- Indonesian review: *'Moen moen terkenal di kota solo dia buka cabang disini, harganya murah dan enak steaknya walaau kadang banyak tepungnya tapi ya pas lah'* (Moen moen that famous in Solo city opened branch in here, *price* was low and the steak was delicious even though the dough sometimes was thick but it's right)
- English review: *'Great place to drink and do hookah. Especially, all the workers are friendly. They know how to treat customer very well. It is good tho that they open from the afternoon now. Recommended place and worth it to pay!! FYI: Do not forget to order watermelon fruit with sisha is the best recipe in exhale!!'*
- Code-mixed review: *'Best waffle kalo menurut gue hehehe. Terakhir gue pesen waffle gelato yg nuttela banana dan cheese french fries, semuanya enaaaaak! Tempatnya juga asik bgt buat nongki dan untuk harga worth it'* (best waffle in my opinion hehehhe. Last time I ordered waffle gelato Nutella banana and cheese french fries, all were deliciousss! The place also was cozy for hanging and the *price* was worth it).

After retrieving our data, we annotated them manually. The aspects that we investigated are *food*, *price*, *service*, and *ambience* following [7]. Besides, we assumed that those aspects always appear in most of reviews. Then we divided the sentiment polarities based on [13] into positive, neutral, and negative. Aspects are classified as positive if the review mentioned positive terms such as '*enak*' (delicious), '*bersih'* (clean), '*murah'* (cheap), and '*sangat bagus'* (excellent). Negative aspects are classified if there were negative terms in review. For example, '*buruk'* (bad), '*mahal'* (expensive), '*kotor'* (dirty), and '*lambat'* (slow). Next, the aspects are classified as neutral aspect if the reviews mentioned term like '*standar'* (standard), '*biasa* saja' (nothing special), '*so so'*, and '*not bad but not great'*. In addition, we also classified the aspects that are not mentioned in the review as neutral aspects because we assumed that while people do not mention about the aspect, the polarity will be neither positive nor negative. The following example shows how our data were annotated:

Review: 'i ordered small chicken mushroom with drink and dumplings only 21K. the taste was so so but the *price* was worth.'

{'*food*': '*neutral'*; '*price'*: '*positive'*; '*service'*: '*neutral'*; '*ambience'*: '*neutral'*}

After checking the annotated data, we filtered the reviews and found there are 19201 reviews that can be used in this research.

B. Data Preprocessing

Before entering the feature extraction step, we performed several basic text processing techniques for cleaning our data as follows:

1) Emoticon Processing: In our first step for preprocessing, we normalized the emoticons that appeared in the text to string which in Indonesian. For example, we changed ':(' to '*negatif*' (negative) and ':)' to '*positif*' (positive).

2) Lowercasing: Second, we lowercased all the words for matching the structures of the words. For instance, ‘GREATTTT!!’ changed to ‘greatttt!!’.

3) Spelling Correction and Abbreviation (part 1): In this step, we corrected the spelling of the words into the formal words and expanding the abbreviation by building our dictionary. This dictionary was made by combining several informal Indonesian and English words. Specifically*, ‘tbh i don’t like the taste’ switched to ‘to be honest i do not like the taste’.*

4) Removing URL, Username, Numbers, and Punctuations: Next, we removed the urls, usernames, and punctuations that occurred in the data. such as, *‘www.instagram.com/Food’*, *‘@makanmakan’*, and ‘*segar banget*!!!’ (so fresh!!!) converted to *‘segar banget’* (so fresh).

5) Spelling Correction and Abbreviation (part 2): In this step, we checked the spelling and abbreviation after last step. This was performed to avoid the words that might still not corrected because of the non-alphabet characters still attached to the words. For example, *‘makanannya enk.’*, after removing the punctuation *‘makanannya enk’*, then changed to *‘makanannya enak’* (the *food* was delicious) in this step.

6) Removing Stopwords: For removing stopwords, we built our own dictionary by combining the stopwords that retrieved from NLTK[6] and from [8] as our base stopwords dictionary. After that, we removed words such as *‘not’* and *‘tidak’* which are have same meaning. The reason is we wanted to avoid the missing information about the negation of positive words because the words *‘not’* and *‘tidak’* are stopwords. For illustration, *‘the price was not cheap’*. If we used the default stopword from library, the sentence will change into ‘*price cheap’* instead of *‘price not cheap’*. So, by using our own stopwords dictionary, *‘the price was not cheap’* can be changed to *‘price not cheap’*.

7) Removing duplicate character including whitespace: Next, we checked whether there were strings that contain a character that appear more than once. If we found it, we eliminated the extra character. For example, *‘deliiiciousss’* is changed to *‘delicious’*. Besides, we also checked whether the review was containing extra space or not.

8) Stemming: In this last step of our preprocessing, we applied stemming function from libraries. We used Snowball Stemmer from NLTK library for English and Sastrawi [7] library for Indonesian. To illustrate, *‘saya memakan kuenya, lumayan’ (I’m eating its cake, not bad), changed to ‘saya makan kue, lumayan’ (I eat cake, not bad).*

C. Feature Extraction

After cleaning our data, we extracted the feature that will be used in the models. The feature we used was bigram term and its vector were extracted by vectorizing the word representations in the reviews. Besides, we also used the combination of stemming and stopwords steps to see whether they can increase the scores of the models because we used own multi-languages stopwords dictionary (English - Indonesian), and two stemmers for English and Indonesian.

D. Experiment

In our experiments, we used four scenarios and applied five machine learning algorithms. After that, we measured and compared their performances using their F1-scores.

1) Experiment Scenarios: We made four scenarios for our experiments in this research. The goal is to see the stopwords and stemming can affect the performances of the models when they applied to our data. In first scenario, we built machine learning models by applying removing stopwords, but we did not use stemming methods. Then in second scenario, we used stemming, but we did not apply the removing stopwords step. After that, we used all preprocessing steps but did not applied the stemming and removing stopwords, and in last scenario, we applied all the preprocessing steps including removing the stopwords and stemming steps.

TABLE I. EXPERIMENT SCENARIOS

Scenarios	Removing Stopwords	Stemming
Scenario 1	✓	×
Scenario 2	×	✓
Scenario 3	×	×
Scenario 4	✓	✓

2) Data: We used all the data (19201 reviews) for all scenarios. From figure 2, we can see that neutral polarity has the highest number in every aspect of the reviews except the *food* aspect. Positive reviews for *food* aspect appeared almost in all reviews and there were only 2965 reviews that were not positive. In contrast, all other aspects had more than 11000 reviews that were neutral. Moreover, *service* aspect has the least reviews with positive polarity and *price* aspect has the highest number of negative reviews.

3) Classification Algorithms: For our classification experiments, we used five algorithms. The machine learning algorithms we used are Decision Tree, Random Forest, Logistic Regression, Extra Tree (Extremely Randomized Tree) Classifier, and Multinomial Naïve Bayes. Logistic Regression was selected based on its performance in [17] which is a study about multiclass classification problem and it got best F1-scores in almost all datasets and scenarios. We also chose Multinomial Naïve Bayes based on Naïve Bayes classification results in [5] because it has good performance while classifying the sentiment of reviews from online retail shop in Indonesia with three polarities (positive, neagative, neutral). For Decision Tree, Random Forest, and Extra Tree, we selected them based on algorithms

[6] https://www.nltk.org/

[7] https://github.com/har07/PySastrawi

recommendation from scikit-learn [8] for multilabel and multiclass problem. After we selected the machine learning algorithms, we compared the F1 scores they obtained, so, we can see which machine learning algorithm works best with our dataset.

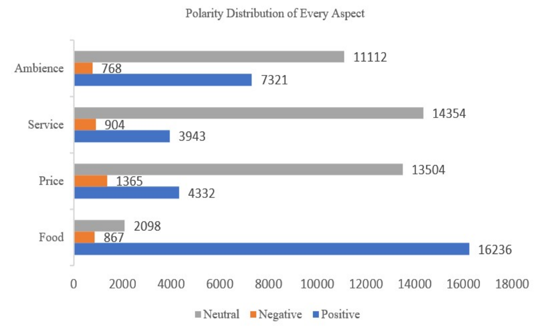

Figure 2. Polarity distribution in every aspect.

E. Evaluation

In these experiments, we used five classifiers: Random Forest (RF), Multinomial Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), and Extra Tree (ET); and cross validation as the validation technique and compared the F1 scores of machine learning models in all scenarios. Number of folds we used for the cross validation are 10 folds. This section shows the model performances in every aspect.

TABLE II. RESULT OF FIRST SCENARIO

Model	*Food*	*Price*	*Service*	*Ambience*
RF	78.52	75.32	78.15	73.03
NB	74.60	62.85	70.15	72.13
LR	**80.57**	74.16	80.33	**75.42**
DT	77.39	**78.53**	**85.00**	73.29
ET	78.33	78.35	83.06	74.83

The Table 2 shows the F1 scores of the models for the first scenario which is applying removing stopwords step without using stemming before entering the classification. From the table, we can see that LR has the highest F1 scores for *food* and *ambience* aspects which were 80.57% and 75.42% respectively. For *price* aspect, it was led by DT by obtaining 78.53%. Besides, it also gained highest score for *service* aspect with 85%.

TABLE III. RESULT OF SECOND SCENARIO

Model	*Food*	*Price*	*Service*	*Ambience*
RF	78.18	73.78	77.04	73.37
NB	77.64	65.04	72.32	74.57
LR	**81.76**	74.98	80.83	**77.29**
DT	78.19	76.64	**84.78**	73.24
ET	78.25	**77.43**	82.51	76.58

From Table 3, we can see the result of applying stemming technique but the stopwords were not removed. While comparing to Table 2, the scores that RF achieved were lower from that table except for the *ambience* aspect which was increased around 0.34%. In contrast, the scores of NB and LR in every aspect were increased. However, for DT and ET, the scores they obtained were lower in *price* and *service* aspects compared to Table 2. Despite of that fact, the highest score for *price* aspect was obtained by ET while DT led score for *service* aspect.

TABLE IV. RESULT OF THIRD SCENARIO

Model	*Food*	*Price*	*Service*	*Ambience*
RF	78.30	73.34	77.33	73.14
NB	77.61	64.71	72.35	74.52
LR	**81.72**	74.09	80.30	**76.69**
DT	77.96	76.44	**85.07**	73.31
ET	78.39	**76.62**	81.85	76.18

The Table 4 shows the result of all algorithms in third scenario which is preprocessing without applying stemming and removing stopwords steps. In comparison with Table 3, DT has better scores in *service* and *ambience* aspects while ET achieved increasing score only in *food* aspect. RF obtained increasing scores for *food* and *service* aspects while NB scores only increased in *service* aspect. LR got lower scores in all aspects, but while compared to Table 2, the scores for *food* and *ambience* aspects were increased whereas the *price* score 0.07% lower and *service* aspect got 80.30%.

For last scenario, RF achieved highest scores compared to its scores in all previous scores, but *ambience* score lower than second and third scenarios. For NB, the scores were higher only when compared to first scenario. LR got highest score for *price* aspect but lower than its scores in second scenario for other aspects.

TABLE V. RESULT OF FORTH SCENARIO

Model	*Food*	*Price*	*Service*	*Ambience*
RF	78.54	75.91	79.10	73.09
NB	75.52	63.89	71.10	72.49
LR	**80.91**	75.69	80.48	**76.12**
DT	77.39	**78.71**	**85.02**	73.59
ET	78.21	78.49	82.98	75.65

DT attained highest scores for *price* and *service* aspects compared to all algorithms, but its *service* score still lower 0.05% than the score it achieved in previous scenario. For ET, it accomplished the second highest for *price* aspect, 0.22% behind DT.

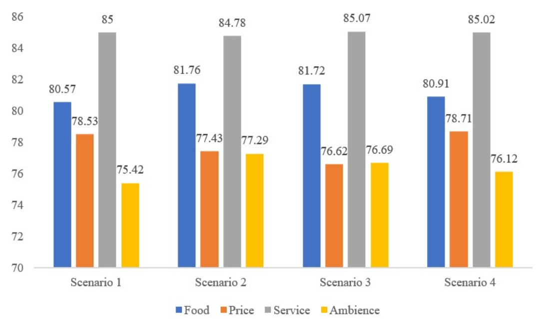

Figure 3. Comparison of highest scores in every aspect of all scenarios.

[8] https://scikit-learn.org/stable/modules/multiclass.html

In summary, by seeing the Figure 3, combination of stemming and stopwords can affect the performance from models. The *food* aspect obtained highest score 81.76% by LR in scenario 2 which is using stemming without removing the stopwords. It seems that stopwords affect the *food* aspect more than stemming because its scores were higher every time the stopwords were not removed. However, its score was better when applied stemming and removing stopwords than only removing stopwords. In contrast, the *price* aspect got highest scores if the removing stopwords technique was used. Even, its best score obtained when using stemming and removing stopwords methods which is 78.71% attained by DT. For *service* aspect, DT achieved highest score for that aspect when removing stopwords and stemming were not applied, but its score also high when using both methods which is only 0.05% lower than without applying both. Besides, its score was better while the stopwords were removed than only using stemming without remove them. Moreover, for *ambience* aspect, same with *food* aspect, its highest score also obtained by LR in second scenario. Similar to *food* aspect, LR classified it better if stopwords were not removed.

Furthermore, the difference of scores that obtained by all classifiers for *food* and *ambience* were not far. This could be caused by the number of reviews from combination of positive and negative in *food* aspect that was higher than its neutral polarity. Consider the neutral polarity was combination of neutral and not mentioned aspect. *Ambience* aspect also has large number of positive and negative reviews compared to *service* and *price* aspects. So, by seeing the number of both positive and negative polarity, the variety of its samples was many. For *ambience*, even though the combination number of the positive and negative reviews was higher that *service* and *price*, but the score of the latter aspects were better than *ambience* for RF, DT, and ET. This might due to reviews rarely mention "*ambience*" or "*suasana*" directly, and people describe it more variative. For instance, "*ruangannya sepi jadi nyaman banget* (the room was dark, so it felt comfortable)" or "*dekorasinya unik jadi bagus banget buat foto tapi lantainya kotor* (the decoration was unique, so, it was cool for taking picture, but the floor was dirty". It can make the algorithm classified the review wrongly if in training data there are "Unique decoration", "cool for taking picture" and "the floor was bit dirty and dusty" in separated reviews. However, *service* aspect has higher scores than *price* and *ambience* in almost all algorithms except for NB. This might highly cause by the reviews that contain *service* aspect usually straight forwarded about it and mention *service* aspect directly. For example, "*pelayannya sangat ramah* (the waitress was really kind)" or "*pelayanannya buruk* (the *service* was bad)". Same goes with reviews mentioning *price* aspect which were straight forwarded like *service*, so the variation of the samples was not many. Besides, even though the data were small, models can classify it better.

IV. Conclusion and Future Work

In this work, we have examined the performances of five machine learning algorithms which are Random Forest, Naïve, Bayes, Logistic, Regression, Decision Tree, and Extra Tree, to classify the opinion of the aspects that mentioned in code-mixed reviews. The aspects are *food*, *price*, *service*, and *ambience*. The evaluation conducted by making four scenarios namely removing stopwords without stemming, without removing stopwords but with stemming, without removing stopwords and stemming, and preprocessing with removing stopwords and stemming. The model performances were measured by using 10 folds cross validation, and the results show Logistic Regression achieved highest score for *food* (81.76%) and *ambience* (77.29%) aspects in second scenario. The highest score for *price* (78.71%) aspect was obtained by Decision Tree in last scenario, and *service* (85.07%) in third scenario. By seeing the results, it can be concluded that removing stopwords and stemming can affect the algorithm performances, specifically for small number of reviews.

In this work, we still using dataset with mixing languages. In the future, we will try to use one language in the same dataset by translating the dataset to either English or Indonesian to see how the language can affect the model performance. In addition, consider we still applied traditional machine learning algorithms in this research, in next study we will use Deep Learning and compare the result with the traditional machine learning models.

Acknowledgment

The authors acknowledge the PIT 9 research grant NKB-0010/UN2.R3.1/HKP.05.00/2019 from Directorate Research and Community *Service*s, Universitas Indonesia.

References

[1] A. A. Mustafa and I. Budi, "Recommendation System Based on Item and User Similarity on Restaurants Directory Online," 2018 6th International Conference on Information and Communication Technology (ICoICT), Bandung, 2018, pp. 70-74, doi: 10.1109/ICoICT.2018.8528775.

[2] A. Cahyadi and M. L. 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), Krabi, 2018, pp. 124-129, doi: 10.1109/ICAICTA.2018.8541300.

[3] A. Pravalika, V. Oza, N. P. Meghana and S. S. Kamath, "Domain-specific sentiment analysis approaches for code-mixed social network data," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8204074

[4] C. Fiarni, H. Maharani and E. Irawan, "Implementing Rule-based and Naive Bayes Algorithm on Incremental Sentiment Analysis System for Indonesian Online Transportation *Service*s Review," 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Kuta, 2018, pp. 597-602, doi: 10.1109/ICITEED.2018.8534912.

[5] C. Fiarni, H. Maharani and R. Pratama, "Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique," *2016 4th International Conference on Information and Communication Technology (ICoICT)*, Bandung, 2016, pp. 1-6. doi: 10.1109/ICoICT.2016.7571912

[6] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Denpasar, 2017, pp. 1-6, doi: 10.1109/ICAICTA.2017.8090963.

[7] D. H. Sasmita, A. F. Wicaksono, S. Louvan and M. Adriani, "Unsupervised aspect-based sentiment analysis on Indonesian restaurant reviews," 2017 International Conference on Asian Language Processing (IALP), Singapore, 2017, pp. 383-386, doi: 10.1109/IALP.2017.8300623.

[8] F.Z. Tala, "A Study of Stemming Effects on Information Retrieval in Indonesian Indonesia". M.Sc. Thesis, Appendix D, Indonesia, 2003, pp. 39–46.

[9] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: Improving rating predictions using review text content," in WebDB, 2009.

[10] I. Alfina, D. Sigmawaty, F. Nurhidayati, and A. N. Hidayanto, "Utilizing Hashtags for Sentiment Analysis of Tweets in The Political Domain," Proceedings of the 9th International Conference on Machine Learning and Computing (ICMLC), Singapore ,2017, pp. 43-47, doi: 10.1145/3055635.3056631.

[11] K. Shalini, H. B. Ganesh, M. A. Kumar and K. P. Soman, "Sentiment Analysis for Code-Mixed Indian Social Media Text With Distributed Representation," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1126-1131. doi: 10.1109/ICACCI.2018.8554835.

[12] K. Zvarevashe and O. O. Olugbara, "A framework for sentiment analysis with opinion mining of hotel reviews," 2018 Conference on Information Communications Technology and Society (ICTAS), Durban, 2018, pp. 1-4. doi: 10.1109/ICTAS.2018.8368746.

[13] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis," Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval), Colorado, 2015, pp. 486–495. doi: 10.18653/v1/S15-2082.

[14] M. Yasen and S. Tedmori, "Movies Reviews Sentiment Analysis and Classification," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Jordan, 2019, pp. 860-865. doi: 10.1109/JEEIT.2019.8717422.

[15] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL), Pennsylvania, 2002, pp. 417-424. doi: 10.3115/1073083.1073153.

[16] O. S. Jaya, H. Murfi, and S. Nurrohmah, "Analysis of Convolution Neural Network for Transfer Learning of Sentiment Analysis in Indonesian Tweets," Proceedings of the 2018 International Conference on Data Science and Information Technology (DSIT), Singapore, 2018, pp. 18-22. doi:10.1145/3239283.3239299.

[17] S. Chawla and M. Mehrotra, "An Ensemble-Classifier Based Approach for Multiclass Emotion Classification of Short Text," *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2018, pp. 768-774. doi: 10.1109/ICRITO.2018.8748757

[18] S. J. Putra, I. Khalil, M. N. Gunawan, R. Amin, and T. Sutabri, "A Hybrid Model for Social Media Sentiment Analysis for Indonesian Text," Proceedings of the 20th International Conference on Information Integration and Web-based Applications & *Service*s (iiWAS2018), Indonesia, 2018, pp. 297-301. doi:10.1145/3282373.3282850.

[19] Y. A. Amrania, M. Lazaar, and K. E. E. Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," Procedia Computer Science, vol 127, 2018, pp. 511-520. doi: https://doi.org/10.1016/j.procs.2018.01.150.