Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 3, December 2023, pp. 1495~1502
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i3.pp1495-1502

Classifying product review quality based on semantic and structural features

Ilham Akhyar Firdaus[1], Dwi Rolliawati[1], Anang Kunaefi[1], Firdaus[2]
[1]Department of Information System, Faculty of Science and Technology, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
[2]Department of Management, Faculty of Economics and Business, Dr. Soetomo University, Surabaya, Indonesia

Article Info

Article history:

Received Feb 7, 2023
Revised Aug 2, 2023
Accepted Aug 30, 2023

Keywords:

Product reviews
Review helpfulness
Semantic feature
Structural feature
Support vector machine
Text classification

ABSTRACT

Product reviews are written opinions submitted by consumers in assessing a product. The existence of product reviews is important because it can help consumers make better product purchasing decisions. But product reviews can also be unimportant if the quality of the information from the reviews is not helpful. This can be minimized if a classification is carried out to find out which reviews are helpful or not. For this to be achieved, this research will apply a support vector machine model using semantic and structural features to be able to classify review texts based on their characteristics. By applying the appropriate preprocessing stages, the final results show that the semantic features produce the highest F1-score value of 0.825. Whereas the structural features produce the highest F1-score value of 0.823. From this, it can be concluded that semantic features can be used to identify the characteristics of a review text that are helpful or not properly. This success also shows outstanding performance in classifying reviews as helpful or not compared to previous studies.

Corresponding Author:

Ilham Akhyar Firdaus
Department of Information System, Faculty of Science and Technology, UIN Sunan Ampel Surabaya
Surabaya, East Java, Indonesia
Email: ilhamakhyar.f@gmail.com

1. INTRODUCTION

Buying and selling transactions using e-commerce has become a common lifestyle behaviour in today's society because people can buy products without having to physically visit a store which can reduce efforts to shop offline. The focus of e-commerce today is to maximize shopping efficiency with strategies such as product search convenience, one-click purchases, and virtual catalogues that have specifications and recommendations based on consumer shopping behaviour in the past [1]. This resulted in consumers from e-commerce experiencing an increase and the occurrence of electronic word of mouth. Electronic word of mouth (eWOM) refers to consumer feedback and points of view regarding products or services which can be in the form of votes, comments, ratings, reviews, or a post on a blog [2], [3]. Due to the existence of eWOM, consumers are more interested in products discussed online compared to products discussed traditionally (offline) in which eWOM is a richer source of objective information [4]. One type of eWOM is product reviews. Product reviews refer to textual reviews from consumers that describe characteristics such as the advantages or disadvantages of a product [5]. Product reviews are important because they have an effect on consumer decisions in purchasing a product based on the attributes, usage situations, and performance of the product by other consumers [6]. Even so, product reviews can also be unimportant if the quality of the information from the reviews is not helpful. This can be solved by classifying the reviews as helpful or not based on the features (characteristics) of the reviews.

Different researchers have proposed various methodologies to be able to classify reviews as helpful or not, most of which focus on the application of feature engineering. Like Krishnamoorthy [7] who proposed the use of linguistic features, review metadata, readability, and subjectivity. Where by carrying out the pre-processing stage, namely deleting duplicate reviews, deleting reviews that have a low total vote, and deleting reviews that do not contain content, obtaining an f-measure value of 0.614 for naive bayes, 0.753 for support vector machine (SVM), and 0.778 for random forests. Saumya *et al.* [8] proposed the novelty of using product descriptions and customer question-answers as feature extraction. By applying the synthetic minority over-sampling (SMOTE) technique, the naive bayes, SVM, and random forest models yield F1-score values of 0.565, 0.805, and 0.87, respectively. But unfortunately, in this study, the pre-processing stages were only explained briefly, such as removing unicode characters and images and reducing reviews that had high votes. Du *et al.* [9] proposed the use of semantics and sentiment features where at the pre-processing stage reviews that are not in English, duplicates, and have a total vote of less than 10 will be deleted. The review has also changed its form to lowercase and word tokens. The final results show that semantic features are superior with the highest accuracy value obtained at 0.81. Akbarabadi and Hosseini [10] proposes novelty by using reviewers and title characteristics as its features. Using the pre-processing stage, namely deleting reviews that have a total vote of less than 10 and paying attention to the proportion of the helpfulness ratio received by the review, these two features get the highest f-measure value of 0.96 for the random forest model and 0.88 for the decision tree model. All of these studies used Amazon's product review dataset in their tests. The research uses different datasets and approaches, such as Ma *et al.* [11] which uses text reviews and photos as its features, and Yelp and TripAdvisor as its datasets. Where by using the proposed deep learning model, the model can outperform other baseline models (decision tree, SVM, and logistic regression) with the highest F1-score value of 0.79. Then Luo and Xu [12] proposed a new approach, namely by conducting an exploratory analysis using semantic, sentiment, and latent dirichlet allocation which was tested on the Yelp dataset. The combination of SVM and fuzzy domain ontology produces the highest F1-score of 0.795 compared to other models.

Based on the previous research described above, these studies still do not explain in detail or completely how the preprocessing stages were carried out before to get a review of whether is helpful or not. As for Meng *et al.* [13] which also uses the Amazon product review dataset, in the use of structural features, it proposes not only to look at the character or word features of a review text but also to consider the relationship or meaning of the character or word itself. This work proposes the use of structural features namely features that focus on the structure of a word, and semantic features namely features that focus on the meaning of the word it self [14] as a process in its feature extraction. As for the preprocessing stage, each process will be explained in detail so that you can find out how the correct process is to get a clean review. Furthermore, for the classification model, the SVM model will be used based on its performance in previous studies [15]–[17]. As another form of contribution, in this study the two features will also be combined to find out whether the combination has an effect or not on the value obtained by the classification results in the model later.

2. METHOD

To answer the problems previously described, the proposed research stages are shown in Figure 1. Where in Figure 1 there are six main stages in conducting a classification to find out whether a review is helpful or not based on the features that have been proposed previously. The six main stages are data collection, data labelling, data pre-processing, feature extraction, modelling, and model evaluation.

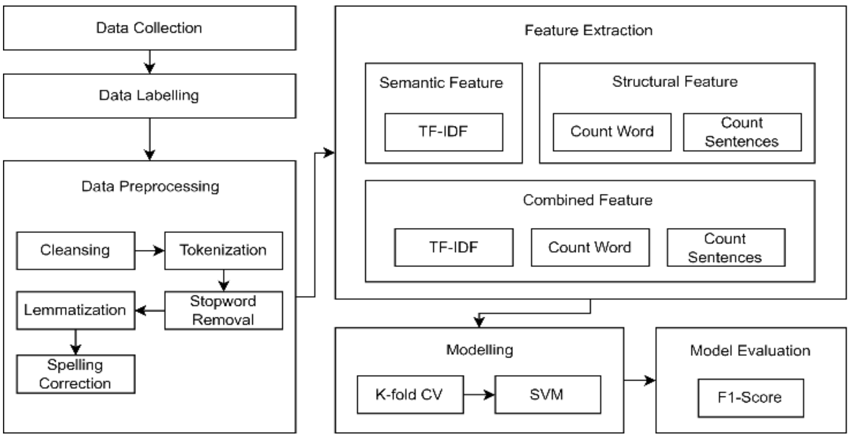

Figure 1. Proposed method

A summary of the six main stages is as follows. In the data collection, we collect data first by adjusting to the research that is being carried out. Because this study is related to a product review, the data needed is data that is also in the form of reviews. Then in data labelling, the data that has been obtained will be given a label to state the ground truth of the data. For classification problems, the ground truth of the data is important not only to help to train the model but also to assist in validating the results of the model. Next in data preprocessing, existing data will be processed to obtain clean data such as removing punctuation, symbol, or even expanding abbreviations. This is also necessary to do because a model tends to produce better classification performance if the data used is clean. After that, in the feature extraction, the clean data will be transformed in such a way that the machine learning model can understand. Then in the final stage (model evaluation), the model will be evaluated for its performance based on the classification results that have been obtained. Detailed explanations related to each stage will be explained further in each sub-chapter.

2.1. Data collection

In this study the data used is the 2015 Amazon product review dataset based on its use in previous studies. The dataset used is an open dataset obtained from the official website [18]. In the dataset there are 15 variables. However, this study was limited to only using six variables because only those variables will be used for the next stage such as data labelling, data preprocessing, and feature extraction. The six variables are customer id, product id, helpful votes, total votes, verified purchase and review body. The values and descriptions of the six variables are shown in Table 1.

Table 1. Research data variables

Variable name	Example value	Description
customer_id	302120, 445	Random value identifier that can be used to aggregate reviews written by one customer.
product_id	‘B00MUTIDKI’	Unique id of a product.
helpful_votes	12, 8, 24	Total helpful votes received by a review.
total_votes	8, 23, 46	Total votes, both helpful votes (likes) and unhelpful votes (dislikes) received by a review.
verified_purchase	‘Y’ or ‘N’	It has a ‘Y’ value if the customer has been verified to buy the product directly on Amazon without being given an excessive discount. While the value is ‘N’ if the customer buys the product indirectly through Amazon and does not pay the price available for most buyers on Amazon.
review_body	‘the product is so good’	The contents of the reviews given.

Because there are differences in assessing the quality of products based on their categories [19], this research will use two different types of product categories, namely search products and experience products. Where the search product is a product category that consumers can value before buying the product [20]. Whereas the experience product is a product category that consumers cannot value before buying the product, so they must try it first [21]. In the search product, the video games dataset will be used, while the experience product will use the beauty product dataset.

2.2. Data labelling

In data labelling, reviews will be given a label. The label indicates whether a review is helpful review or not. One indicator of a helpful review is by looking at the number of helpful votes in the review. The more the number of helpful votes, the more likely it is that the review is helpful [22]. Based on Ghose and Ipeirotis [23] which also uses the Amazon product review dataset, a review can be labelled as a helpful review if the value of the helpful ratio in the review is worth more than 0.6. Where the value of the helpful ratio is as in (1) [24].

$$\text{h}elpful\_ratio = \frac{helpful\_votes}{total\_votes} \tag{1}$$

This value of 0.6 is then used in many of the same studies, which use the Amazon product review dataset and classify it to find out whether the review is helpful review or not. Reviews must also have a minimum total vote value of more than 10. Because reviews with a small total vote value can be biased and unreliable [25].

2.3. Data preprocessing

Data preprocessing is needed to get clean reviews as well as reviews that have done restoration on its words. Cleaning and restoration of words in the review are important to do because they can provide a strong foundation for the next modelling stage [26]. Six processes will be carried out on data preprocessing, namely as follows:

- Data cleansing: In data cleansing, you will get reviews that are reliable and structurally clean. Where reviews will be cleaned from incomplete/blank data, consumers who don't buy products, and duplicate data. Furthermore, the review will be changed to normal form, removing uniform resource locator (URL) and hypertext markup language (HTML) tags, expanding abbreviations and slang words, and removing punctuation, symbols, and numbers.
- Tokenization: Tokenization is used to separate sentences into word tokens. Where the results will be used in the next process, namely stopword removal.
- Stopword removal: Stopword removal is the process of removing common words in the review language, namely English. These common words (stopwords) appear frequently in many parts of the text in a document but carry little information about which part of the text they belong to [27] it can affect the performance of the model [28] and need to be removed.
- Lemmatization: Lemmatization is needed to change the word to its basic form while still paying attention to the content and meaning of the word.
- Spelling correction: In spelling correction, a word is checked first, namely whether it is present or not in the English word dictionary (WordNet). Furthermore, if the word is not in WordNet, then the word has the possibility of having an error in its spelling which needs to be corrected.

2.4. Feature extraction

Feature extraction is a process of transforming existing features in text into quantitative data structures such as numbers to be entered as input in the model training process. In terms of semantic features, feature extraction is a process of extracting features whose role is to find out the meaning of words in a text. Where in this study, the term frequency-inverse document frequency (TF-IDF) [29] method will be used on the semantic features represented in (2) for term frequency (TF).

$$TF_{ij} = \frac{f_{ij}}{max_k f_{kj}} \tag{2}$$

As an example, given a collection of N documents, then f_{ij} in TF is the number of occurrences of a word i in document j. And $max_k f_{kj}$ is the maximum value k of the occurrence of any word in document j. Whereas in IDF the word i is denoted as in (3).

$$IDF_i = log\frac{N}{n_i} \tag{3}$$

where N is the number of documents in the corpus and N_i is the number of documents containing the word i in the N documents in the corpus. The value obtained from this calculation is between zero and one. Where the largest value is the value that approaches the value of one.

Next structural feature extraction is a process of extracting features whose role is to find out the structure and format of a text document. Where this study will be carried out by the calculation of the number of words and the number of sentences from the review. The value obtained from this calculation can be in units, tens, or even hundreds. The flow of the calculation can be seen in Figure 2. In Figure 2, the review's body serves as the primary variable (x) and is processed to determine the number of words ($x1$) and sentences ($x2$). To calculate the number of words, the review's body must first be cleaned of any punctuation, symbols, and numbers. Afterwards, it will be transformed into word tokens, allowing for the number of tokens to be counted. Next, to be able to produce the number of sentences, the review's body is inputted into the sent tokenize function from the natural language toolkit (NLTK) library. This library identifies a period as the end of a sentence while considering other factors that may impact sentence identification. The function then calculates the total number of sentences produced.

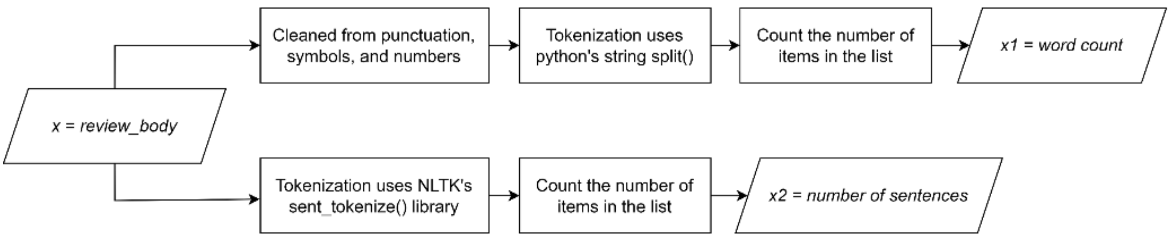

Figure 2. Structural feature extraction flow

As previously explained, this research will also combine these two features. In the merger, a standard scaler will be used to be able to equalize the existing values, because the results of the extraction of the two features have different value scales that vary. The formula for the standardization process using a standard scaler is as in (4) where μ is the mean and σ is the standard deviation.

$$z = \frac{x-\mu}{\sigma} \tag{4}$$

2.5. Modelling

At the modelling stage, there are three scenarios (in the feature extraction) to determine how well the model can predict which reviews are helpful or not. Namely, in the first scenario, the model will use semantic features, then in the second scenario the model will use structural features, and for the last scenario, the model will use a combination of these two features. Next, the k-fold cross validation method will be used with a value of k equal to 10 to be able to carry out the modelling process.

2.6. Model evaluation

To be able to evaluate the model, this study will use the F1-score assessment matrix. F1-score is a balancing value between precision and recall with a range of values from zero to one. As before, the F1-score formula itself is a combination of precision (5) and recall (6) as in (7).

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{5}$$

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \tag{6}$$

$$F1 - score = 2 \times \frac{precision * recall}{precision + recall} \tag{7}$$

3. RESULTS AND DISCUSSION

In this study, a dataset of Amazon product reviews was obtained for the video games and beauty category where each dataset has 1,048,576 rows and 15 existing variables. After getting the data to be used, the next step is to label each review. As previously explained, labelling is done by looking at the number of helpful votes and the total votes received by the review. The results obtained after data labelling and preprocessing in the review are shown in Table 2.

Table 2. Example of review labelling and preprocessing results

Helpul votes	Total votes	Review body	Helpful ratio	Label	Clean review
10	27	Very good, inexpensive brush! Bought 5 for my wife. On the last one and amp ready to re-order. Can't beat the price https://www.youtube/nVHP49g5IPQ.	0.37	unhelpful	good inexpensive brush buy wife last one ready order beat price
18	23	Who pays 4 dollars more for a $20 gift card? What store doesn't sell gift cards that the extra 4 dollars sounds like a good idea?	0.78	helpful	pay dollar gift card store sell gift card extra dollar sound like good idea
12	12	Fun game, fast delivery. No problems or complaints. Nice aqnd fast delivery. Game is in excellent condition. Brand new i believe. So, it is gr8	1	helpful	fun game fast delivery problem complaint nice and fast delivery game excellent condition brand new believe great

The results of this labelling are as explained in the research methodology, namely that reviews can be labelled as helpful reviews if the results of the division between helpful votes and total votes are more than 0.6 and the minimum value of the total votes is more than 10. Furthermore, the modelling results are as follows. The model was tested with three scenarios, namely the first using semantic features, the second using structural features, and the third using combination features (combining semantic and structural features). Because the problem of imbalance classes is not the focus of this study, the review data will also be class balanced by manually equating the number of positive (helpful) and negative (unhelpful) classes. By using 8,000 lines of review, the results of the modelling are shown in Table 3.

Table 3 shows the results of the modelling by using the SVM model and the features that have been proposed previously. The results obtained are that the two proposed features both have high F1-score values but with the difference that semantic features are superior using search products (video games), while structural features are superior using experience products (beauty). From the two types of datasets, another conclusion can be drawn, namely, the F1-score obtained by the video game product review dataset has a higher value when compared to the beauty product review dataset. Where this shows that reviews in the search product category are easier to classify than reviews in the experience product category. Malik [30] which says that review content indicators are more influential on search products in predicting which reviews are helpful or not when compared to experience products. Chua and Banerjee [31] also said that search products are easier to evaluate product quality even before the product is purchased compared to experience products. Furthermore, this study also combines semantic and structural features based on suggestions from previous studies. In the results of combined feature, the average value is lower when compared to each of the semantic and structural features, even though scaling has been done on the data. This can also be possible if the data generated from the merging process is not optimal, such as some data are not equal in shape.

Table 3. Modelling results

Classifier and feature name	Dataset category	F1-score
SVM+Semantic feature	Beauty	0.774 ± 0.01
	Video games	0.825 ± 0.01
SVM+Structural feature	Beauty	0.780 ± 0.01
	Video games	0.823 ± 0.01
SVM+Combined feature	Beauty	0.736 ± 0.01
	Video games	0.785 ± 0.01

4. CONCLUSION

In this study, as suggested by previous research, it has been proposed to use semantic and structural features as a feature extraction process to be able to classify in predicting which reviews are helpful or not. By using the SVM model and carrying out the right preprocessing stages, semantic features can produce the highest F1-score value of 0.825 and for structural features can produce the highest F1-score value of 0.823. Furthermore, in other forms of novelty, this research has also combined these two features. By scaling the data first, this combination feature can produce the highest F1-score value of 0.785. However, when compared to the two features, the combination feature has a lower F1-score. Even so, the two features that have been proposed previously proved to be able to predict well which reviews are considered helpful or not. For the SVM model itself, the model has also been proven that it can work well in text classification namely by using semantic, structural, or even combination features as feature extraction. Model is also proven to be able to overcome the problem of document vectors which are generally sparse (few non-zero values), which sparse data can increase the time and space complexity of the model. There are some limitations to our research. Firstly, the data labelling still uses an automatic method, namely by looking at the number of helpful votes and total votes received by reviews which are based on previous research which also used the same method. It is necessary to validate experts manually by taking samples from all data that has been previously labelled. Secondly, for scaling the data, namely to combine the two features, other methods can be used besides the standard scaler. Because the standard scaler does not work optimally on the data used, that is, it cannot calculate the mean of the data in the form of a compressed sparse matrix. Lastly, for the SVM model itself, because this research focuses more on applying the features that have been proposed previously, the model used is a standard SVM model where no modifications are made to the model. The model can be modified such as tuning parameters on the kernel (linear, poly, and gaussian) or the regularization value. Since this study only uses the default rbf kernel and the default value of the regularization parameter which is 1.0.

REFERENCES

[1] X. Li, X. Zhao, W. Xu, and W. Pu, "Measuring ease of use of mobile applications in e-commerce retailing from the perspective of consumer online shopping behaviour patterns," *Journal of Retailing and Consumer Services*, vol. 55, p. 102093, Jul. 2020, doi: 10.1016/j.jretconser.2020.102093.

[2] N. Donthu, S. Kumar, N. Pandey, N. Pandey, and A. Mishra, "Mapping the electronic word-of-mouth (eWOM) research: a systematic review and bibliometric analysis," *Journal of Business Research*, vol. 135, pp. 758–773, Oct. 2021, doi: 10.1016/j.jbusres.2021.07.015.

[3] A. Reyes-Menendez, J. R. Saura, and J. G. Martinez-Navalon, "The impact of e-WOM on hotels management reputation: exploring tripadvisor review credibility with the ELM model," *IEEE Access*, vol. 7, pp. 68868–68877, 2019, doi: 10.1109/ACCESS.2019.2919030.

[4] S. Verma and N. Yadav, "Past, present, and future of electronic word of mouth (EWOM)," *Journal of Interactive Marketing*, vol. 53, pp. 111–128, Feb. 2021, doi: 10.1016/j.intmar.2020.07.001.

[5] Y. Wang and X. Li, "Mining product reviews for needs-based product configurator design: a transfer learning-based approach," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6192–6199, Sep. 2021, doi: 10.1109/TII.2020.3043315.

[6] Y. Zhao, X. Xu, and M. Wang, "Predicting overall customer satisfaction: big data evidence from hotel online textual reviews," *International Journal of Hospitality Management*, vol. 76, pp. 111–121, Jan. 2019, doi: 10.1016/j.ijhm.2018.03.017.

[7] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3751–3759, May 2015, doi: 10.1016/j.eswa.2014.12.044.

[8] S. Saumya, J. P. Singh, A. M. Baabdullah, N. P. Rana, and Y. K. Dwivedi, "Ranking online consumer reviews," *Electronic Commerce Research and Applications*, vol. 29, pp. 78–89, May 2018, doi: 10.1016/j.elerap.2018.03.008.

[9] J. Du, J. Rong, S. Michalska, H. Wang, and Y. Zhang, "Feature selection for helpfulness prediction of online product reviews: an empirical study," *PLoS ONE*, vol. 14, no. 12, p. e0226902, Dec. 2019, doi: 10.1371/journal.pone.0226902.

[10] M. Akbarabadi and M. Hosseini, "Predicting the helpfulness of online customer reviews: the role of title features," *International Journal of Market Research*, vol. 62, no. 3, pp. 272–287, May 2020, doi: 10.1177/1470785318819979.

[11] Y. Ma, Z. Xiang, Q. Du, and W. Fan, "Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning," *International Journal of Hospitality Management*, vol. 71, pp. 120–131, Apr. 2018, doi: 10.1016/j.ijhm.2017.12.008.

[12] Y. Luo and X. Xu, "Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: a case study of yelp," *Sustainability (Switzerland)*, vol. 11, no. 19, p. 5254, Sep. 2019, doi: 10.3390/su11195254.

[13] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: an interpretation framework using xgboost and shap values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466–490, Nov. 2021, doi: 10.3390/jtaer16030029.

[14] C. Qiao and X. Hu, "Text classification for cognitive domains: a case using lexical, syntactic and semantic features," *Journal of Information Science*, vol. 45, no. 4, pp. 516–528, Aug. 2019, doi: 10.1177/0165551518802522.

[15] P. Wang *et al.*, "Classification of proactive personality: text mining based on weibo text and short-answer questions text," *IEEE Access*, vol. 8, pp. 97370–97382, 2020, doi: 10.1109/ACCESS.2020.2995905.

[16] X. Luo, "Efficient english text classification using selected machine learning techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3401–3409, Jun. 2021, doi: 10.1016/j.aej.2021.02.009.

[17] J. Balakumar and S. V. Mohan, "Artificial bee colony algorithm for feature selection and improved support vector machine for text classification," *Information Discovery and Delivery*, vol. 47, no. 3, pp. 154–170, Aug. 2019, doi: 10.1108/IDD-09-2018-0045.

[18] "Amazon product reviews dataset." github.com. https://github.com/akhyaruu/classifying-product-review-quality-based-on-semantic-and-structural-features-dataset (accessed Sept. 20, 2023).

[19] C. C. Yin, H. C. Chiu, and Y. C. Hsieh, "Aligning conflict brand stories with product types," *Journal of Consumer Marketing*, vol. 37, no. 7, pp. 909–919, Sep. 2020, doi: 10.1108/JCM-09-2019-3428.

[20] K. Li, Y. Chen, and L. Zhang, "Exploring the influence of online reviews and motivating factors on sales: a meta-analytic study and the moderating role of product category," *Journal of Retailing and Consumer Services*, vol. 55, p. 102107, Jul. 2020, doi: 10.1016/j.jretconser.2020.102107.

[21] X. Fan, Z. Chai, N. Deng, and X. Dong, "Adoption of augmented reality in online retailing and consumers' product attitude: a cognitive perspective," *Journal of Retailing and Consumer Services*, vol. 53, p. 101986, Mar. 2020, doi: 10.1016/j.jretconser.2019.101986.

[22] X. Sun, M. Han, and J. Feng, "Helpfulness of online reviews: examining review informativeness and classification thresholds by search products and experience products," *Decision Support Systems*, vol. 124, p. 113099, Sep. 2019, doi: 10.1016/j.dss.2019.113099.

[23] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498–1512, Oct. 2011, doi: 10.1109/TKDE.2010.188.

[24] J. N. Wang, J. Du, and Y. L. Chiu, "Can online user reviews be more helpful? Evaluating and improving ranking approaches," *Information and Management*, vol. 57, no. 8, p. 103281, Dec. 2020, doi: 10.1016/j.im.2020.103281.

[25] J. E. Fresneda and D. Gefen, "A semantic measure of online review helpfulness and the importance of message entropy," *Decision Support Systems*, vol. 125, p. 113117, Oct. 2019, doi: 10.1016/j.dss.2019.113117.

[26] A. Kunaefi and M. Aritsugi, "Extracting arguments based on user decisions in app reviews," *IEEE Access*, vol. 9, pp. 45078–45094, 2021, doi: 10.1109/ACCESS.2021.3067000.

[27] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS ONE*, vol. 16, no. 8 August, p. e0254937, Aug. 2021, doi: 10.1371/journal.pone.0254937.

[28] R. Rani and D. K. Lobiyal, "Performance evaluation of text-mining models with Hindi stopwords lists," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2771–2786, Jun. 2022, doi: 10.1016/j.jksuci.2020.03.003.

[29] S. W. Kim and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, p. 30, Dec. 2019, doi: 10.1186/s13673-019-0192-7.

[30] M. S. I. Malik, "Predicting users' review helpfulness: the role of significant review and reviewer characteristics," *Soft Computing*, vol. 24, no. 18, pp. 13913–13928, Sep. 2020, doi: 10.1007/s00500-020-04767-1.

[31] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Computers in Human Behavior*, vol. 54, pp. 547–554, Jan. 2016, doi: 10.1016/j.chb.2015.08.057.

BIOGRAPHIES OF AUTHORS

Ilham Akhyar Firdaus is a student at the Department of Information System, Faculty of Science and Technology, Sunan Ampel State Islamic University, Surabaya, Indonesia. Currently, the work he does is in the area of software engineering, especially in the field of web development. He has also won research-based competitions such as the Sunan Ampel State Islamic University young scientist competition and has also been a presenter at an international conference at Kumamoto University which was held online in 2022. His research interests include machine learning and natural language processing. He can be contacted at email: ilhamakhyar.f@gmail.com.

Dwi Rolliawati is a lecturer in the Department of Information System, at Sunan Ampel State Islamic University, Surabaya, Indonesia. Her research interests include computer science, modelling simulation, and machine learning. She can be contacted at email: dwi_roll@uinsby.ac.id.

Anang Kunaefi received the B.C. and M.C. degrees in informatics engineering from the Sepuluh Nopember Institute of Technology, Surabaya, Indonesia, in 2004 and 2013, respectively. After that, in 2021, he received a Ph.D. degree in computer science and electrical engineering from Kumamoto University, Japan. From 2003 to 2010, he was a Software Engineer and focused on the development of a web-based application system. Since 2014, he has been with the Department of Information System, Sunan Ampel State Islamic University, Surabaya, Indonesia. His research interests include software engineering, data-driven requirements, business process automation, semantic web services, and the implementation of information systems for educational institutions. He can be contacted at email: akunaefi@uinsby.ac.id.

Firdaus is a lecturer in the Management Study Program, Faculty of Economics and Business, Dr. Soetomo University, Surabaya, Indonesia. Obtained a bachelor's degree in economics from the Management Study Program at Jember State University and a Masters's degree in education management from Surabaya State University. Aside from being a lecturer, he is also a researcher. His research interests are in Quantitative Management (operation research) and Business feasibility studies. Books and articles have also been produced during their profession as lecturers and researchers. He can be contacted at email: firdaus@unitomo.ac.id.