Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

Sentiment Analysis on Beauty Product Review Using Modified Balanced Random Forest Method and Chi-Square

Antika Putri Permata Wardani*, Adiwijaya, Mahendra Dwifebri Purbolaksono

Fakultas Informatika, Program Studi Informatika, Telkom University, Bandung
Jl. Telekomunikasi. 1, Terusan Buahbatu - Bojongsoang, Telkom University, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, Indonesia
Email: [1,*]antikaputripew@student.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id, [3]mahendradp@telkomuniversity.ac.id
Email Penulis Korespondensi: antikaputripew@student.telkomuniversity.ac.id
Submitted: **05/08/2022**; Accepted: **02/09/2022**; Published: **31/10/2022**

Abstrak−Pengguna internet di Indonesia telah menggunakan layanan e-commerce dalam membeli berbagai produk. Salah satu website yang menyediakan layanan informasi seputar produk kecantikan wanita adalah Female Daily. Pada website tersebut terdapat ulasan dari produk kecantikan. Fitur ulasan merupakan salah satu fitur yang sangat membantu pengguna dalam menentukan produk kecantikan yang akan dibeli. Banyaknya ulasan akan membutuhkan waktu lama untuk membaca, hampir tidak mungkin pengguna akan membaca keseluruhan informasi yang ada. Oleh karena itu diperlukan penelitian untuk mempermudah pengguna dalam mempertimbangkan produk seperti analisis sentimen. Analisis sentimen bertujuan untuk mengelompokkan opini yaitu review pengguna menjadi opini positif, netral dan negatif. Pada penelitian ini, sentimen analisis menggunakan metode *Modified Balanced Random Forest*(MBRF) dan *Chi-square* sebagai seleksi fitur. Model terbaik dari penelitian ini menghasilkan nilai rata-rata *accuracy* dan *f1-score* masing-masing sebesar 81,75% dan 71,90%.

Kata Kunci: Ulasan; Produk Kecantikan; Analisis Sentimen; Chi-Square; MBRF

Abstract−Internet users in Indonesia have used e-commerce services to buy various products. For example, one website that provides information services about women's beauty products is Female Daily. On the website, there are reviews of beauty products. The review feature is one feature that helps users in determining which beauty products to buy. Unfortunately, many reviews will take a long time to read, and it is almost impossible for users to read all the information. Therefore, research is needed to make it easier for users to consider products such as sentiment analysis. Sentiment analysis aims to classify opinions, namely, user reviews, into positive, neutral, and negative opinions. In this study, sentiment analysis uses the *Modified Balanced Random Forest*(MBRF) and *Chi-square* method as feature selection. The best model from this study produces an average accuracy and an average f1-score of 81.75% and 71.90%, respectively.

Keywords: Review; Beauty Product; Sentiment Analysis; Chi-Square; MBRF

1. INTRODUCTION

In 2021, OJK found that 88.1% of Indonesian internet users use e-commerce services to buy some products. This figure is obtained from the results of the We Are Social survey. The survey shows that Indonesia ranks first in e-commerce services [1]. One website that provides information services about women is Female Daily. On the website https://reviews.femaledaily.com/, there are 300,000 reviews from more than 30,000 beauty products. The review feature is one feature that helps users determine the beauty products to be purchased for use in the user's daily beauty care [2]. Unfortunately, many reviews will take a long time to read, and it is almost impossible for users to read all the information. [3]. It requires a method to analyze the reviews. Sentiment analysis, or opinion mining, is a field of a study analyzing people's opinions, sentiments, evaluations, judgments, attitudes, and emotions [4]. Sentiment classification aims to overcome this problem by automatically classifying opinions in this research dataset, namely user reviews, into positive or negative opinions.

Several studies related to similar topics have been done before—sentiment analysis on reviews of Indonesian consumer products using *Naïve Bayes* and TF-IDF methods. The results of this study were able to produce an accuracy value of 82% [5]. Another research found that Random Forest is a good classifier for classifying multi-classes by producing an accuracy value of 72% [6]. In *Chi-Square* research, it is a feature selection used to determine the feature's dependency on a class. In research [7], the *Chi-Square* method assigns a value to the features. The selected features are used for classification. This study's results can improve the system performance of the built model. In previous research, a new approach is MBRF. The method improves not only accuracy but also time complexity [8]. The *Modified Balanced Random Forest* classification performs very well for the English movie review dataset. *Modified Balanced Random Forest* and *Mutual Information* as feature selection can produce 79% accuracy values and 75% f1-scores [9].

This study references previous studies on classification methods, feature selection, preprocessing, feature extraction and evaluation. In addition, many studies on sentiment analysis are similar such as product reviews, restaurants, movies, and others data.

In a previous study [8], researchers compared the results of the *Random Forest*(RF), *Balanced Random Forest*(BRF) and MBRF. *Random Forest* results in a very long running time in processing imbalanced data. Then testing with the BRF and MBRF methods takes time to run much faster. Of the three methods, MBRF has the best running time of the three methods and produces an accuracy of 93.51%. In research [9] by Firdausi Nuzulul

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

Zamzami et al., machine learning was carried out using *Modified Balanced Random Forest* and *Mutual Information*. The data used in this study are diverse and imbalanced. The study can produce the best value of performance accuracy and f1-score, namely 79% and 75%.

In a study [10] by Trian Basofi Rohman et al., researchers grouped restaurant reviews into a positive, negative, or neutral class. This study uses a preprocessing method of sentence solving, replacement, POS Tagging, Content Words, Case Folding, Removing Stopword, Stemming, and Convert Negation. Then, the feature extraction stage uses TF-IDF and the *Random Forest* classification method. This study resulted in 74% accuracy.

In research[7] by Deni Irvantoro et al., a study combined the *chi-square* and N-Gram models. The classification results using uni-gram using 75% and 25% features produce the same accuracy of 89%. In general, *chi-square* does not affect the order of the classification results. Still, it can provide improvised accuracy, precision, and recall results in comparing the three n-gram models.

In the study [11] by Dwiki Bayu Satmoko et al., researchers used the *Multiclassifier Ensemble Learning* and *Chi-Square*. The classification methods used are KNN, *Naïve Bayes*, and *Random Forest*. The prediction labels from the three methods will be combined using the majority vote. And then *Random Forest* produced the highest accuracy value of 99.43% compared to the other two methods.

In a similar study[12], researchers analyzed Indonesian-language reviews of online Sambat applications that accommodate online criticism, input and suggestions. The model built uses *chi-square* to select features and KNN as a method to classify data. The test results are 90% precision, 78% recall and 78% f1-measure. From this research, using the selection feature process can increase the f1-measure value.

In this study, the authors use the *Modified Balanced Random* and *Chi-Square* methods as feature selection. This research aims to analyze sentiment towards reviews of Indonesian-language beauty products on the Female Daily website. The authors use the *Modified Balanced Random*(MBRF) method for classifying imbalanced data, and this method has performed very well in previous studies [8][9]. The Chi-Square has a function to select the important and required features. It can improve system performance in previous studies [7].

2. RESEARCH METHODOLOGY

2.1 Research Flow

The system built in this study is a system that analyzes sentiment on reviews of beauty products in Indonesian on the Female Daily website.

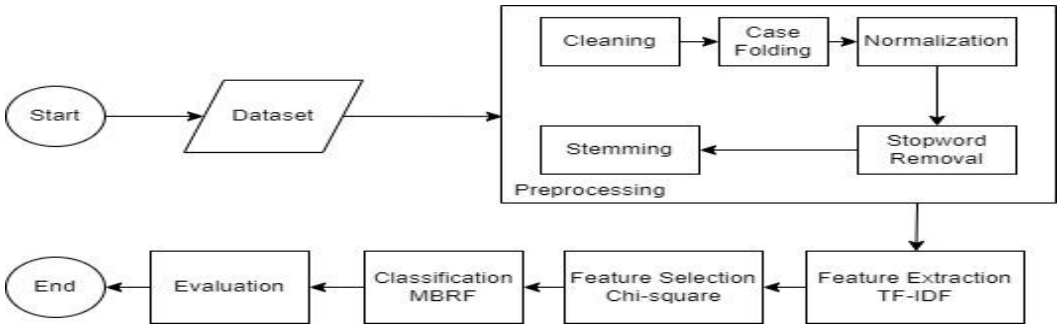

Figure 1. System Design

"Figure 1" shows the research flow. There are several stages in the flow of this research. First, The dataset has classes, namely positive, negative and neutral. Then, preprocessing aims to improve data quality. Then the dataset was extracted with TF-IDF word weighting and Chi-Square as feature selection. The last stage evaluates the system performance by looking at the accuracy value and f1-score.

2.2 Dataset

The dataset was obtained from the website https://femaledaily.com/. This dataset has 3959 rows and six columns(price, packaging, product, and aroma). In addition, four levels of aspects (price, packaging, product, and aroma) have been manually labelled. The data are classified into positive, negative, or neutral based on the criteria.

Table 1. Data Labeling

Review Text	Price	Packaging	Product	Aroma
Saya beli produk ini karena suka banget wanginya. Tapi karena harganya cukup mahal, saya masih mikir-mikir nih apakah akan beli lagi atau nggak ;)	-1	0	0	1

Based on the examples of beauty product reviews in "Table 1", the data has been labelled 1 for positive and 0 for neutral reviews, while negative reviews are labelled -1. For example, in this review, no words indicate the

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

packaging and product aspects, so the packaging and product aspects are labelled 0. Furthermore, in the review, there is the word "mahal", which includes the price aspect, where the word “expensive” is negative, so the review for the price aspect is labelled -1. In the review there is also the word "wangi" or "menang wangi", where the word indicates a positive word for the smell or aroma of the product so that the review for the aroma aspect is labelled

2.3 Preprocessing

Preprocessing is the stage carried out by text mining, where this stage also disappears words that are not important and are not needed. This stage aims to facilitate the system's performance in the classification stage.. The preprocessing stages carried out in this study were data cleaning, case folding, normalization, stopword removal, and stemming. Data Cleaning is the stage of cleaning data that contains punctuation marks, numbers, symbols and others [13]. Next, case folding is changing capital letters to lowercase letters contained in the review data so that there is no duplication of terms by the shape of the letters [10]. In the next stage, normalization replaces abnormal words such as writing errors, non-standard, etc . At this stage, it uses the normalization dictionary. Then, stopword removal eliminates words that do not contain meaning because the words in the stopword category are general words [10]. At this stage, there is also a process of eliminating meaningless words with a dictionary file. Finally, stemming is a process of cutting affixes (prefix, suffix, combination) [10]. “Table 2” shows an example of a review sentence that went through the processing stage.

Table 2. Data Preprocessing Stage

Stage	Input	Output
Data Cleaning	Saya beli produk ini karena suka banget wanginya**.** Tapi karena harganya cukup mahal**,** saya masih berpikir**2** nih apakah akan beli lagi atau nggak **;)**	Saya beli produk ini karena suka banget wanginya Tapi karena harganya cukup mahal saya masih berpikir nih apakah akan beli lagi atau nggak
Case Folding	Saya beli produk ini karena suka banget wanginya Tapi karena harganya cukup mahal saya masih berpikir nih apakah akan beli lagi atau nggak	saya beli produk ini karena suka banget wanginya tapi karena harganya cukup mahal saya masih berpikir nih apakah akan beli lagi atau nggak
Normalization	saya beli produk ini karena suka banget wanginya tapi karena harganya cukup mahal saya masih berpikir nih apakah akan beli lagi atau **nggak**	saya beli produk ini karena suka banget wanginya tapi karena harganya cukup mahal saya masih berpikir nih apakah akan beli lagi atau **tidak**
Stopword Removal	**saya** beli produk **ini karena** suka banget wanginya **tapi karena** harganya cukup mahal **saya masih** berpikir **nih apakah akan** beli **lagi atau** tidak	beli produk suka banget wanginya harganya mahal berpikir beli tidak
Stemming	beli produk suka banget wangi**nya** harga**nya** mahal berpikir beli tidak	beli produk suka banget wangi harga mahal pikir beli tidak

2.4 TF-IDF

After going through the preprocessing stage, the next stage is the TF-IDF weighting process. Weighting is used to convert words into numeric data. For example, TF-IDF (Term Frequency – Inverse Document Frequency) weighting transforms data from textual into numeric data to weight each word or feature[14]. The step that needs to be done is to calculate the Term Frequency (TF) and Document Frequency (DF), then calculate the number of inverse frequencies that can be seen from the following equation:

$$w(t,d) = tf(t.d) \times idf(t) = tf(t,d) \times \log\frac{N}{df(t)} \quad (1)$$

Description :
w(t,d)= TF-IDF weight or term *(t)* weight in document *(d)*
tf(t,d) = number of occurrences of term *(t)* in document *(d)*
idf(t) = number of document frequency inverses per word
df(t) = number of document frequencies per word
N = total number of documents

The result of the TF-IDF is the multiplication of the TF and IDF values, which will result in a greater weight if the word appears infrequently and vice versa [14].

2.5 *Chi-Square*

The next stage is the feature selection stage. The *Chi-Square* method can improve the performance value in several previous studies [7][11][12][15]. *Chi-Square* aims to test the independence of a term with its category and to remove confounding features before classification. In feature selection, the *chi-square* function tests the

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

independence and estimate, which aims to calculate the dependence of a class on the feature. The *Chi-square* has three tests: observe frequency, expected frequency, and test [7]. The first step is to do the calculation with the formula in the following equation:

$$e_{ij} = \frac{o_i . o_j}{N} \quad (2)$$

Description :
e_{ij} = expected frequency
o_i = frequency of the marginal column
o_j = frequency of the marginal row
N = the number of samples

$$x^2 = \sum_{i=1}^{n} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \quad (3)$$

Description :
x^2 = *Chi-square*
o_{ij} = Observed Frequency
e_{ij} = Expected Frequency

In this study, the data used has 8294 features, and the features taken are the best 5000 features from the results of the *chi-square* value. Then the data is classified.

2.6 Data Splitting

The data in this study are divided into train data and test data. Therefore, the percentage of data sharing is 80% train data and 20% test data. Data sharing also uses the k-fold cross-validation method. K-Fold cross-validation is a testing process aiming to assess the built model's performance. This stage works by dividing the data randomly and dividing the data as many as k-partitions. The number of k used in this study is k=5.

2.7 *Modified Balanced Random Forest(MBRF)*

The method used in this study is the MBRF method. MBRF is modified by the *Balanced Random Forest* (BRF) method. BRF handles imbalanced data by applying undersampling for each decision tree formation from *Random Forest*(RF). The MBRF method improves accuracy and reduces time complexity. This method modifies the process of the BRF algorithm, which discards most of the data[8]. This method goes through the undersampling stage because the dataset is imbalanced.

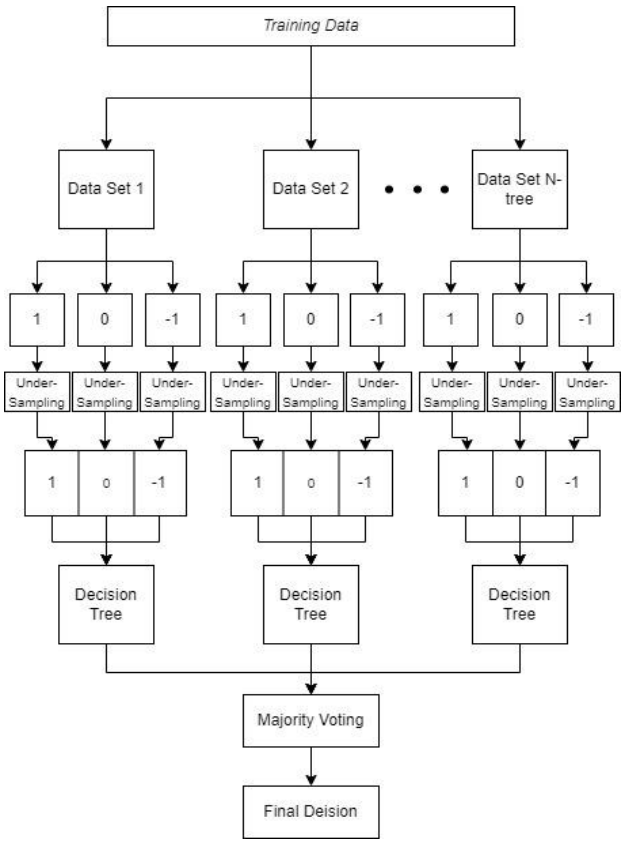

Figure 2. *Modified Balanced Random Forest* Model[8]

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

"Figure 2" shows *Modified Random Balanced Forest* Model. First step is to read the training data. And then, sampling determines the data that becomes a tree of N data. At each tree formation, undersampling is done by selecting the majority class and randomly deleted so that the amount of dataset is balanced with the minority class. In this study, there are four aspects. Each aspect has a different class distribution. So, undersampling in each aspect is different. For example, "1" and "0" are the majority class in the price aspect. While "-1" and "0" are the majority class on product aspects. Then, the dataset is re-labelled to go through the decision tree process. Processing will continue until as many as N-trees. After getting the decision tree results, a majority vote process is carried out to get the output. Then an evaluation is carried out to measure the performance of this method.

2.8 Evaluation

The evaluation aims to measure the performance of the system model with a confusion matrix. The *confusion matrix* displays information from the classification results in the form of actual and prediction classes—the evaluation of the performance of the classification system in general uses metric data [16]. The following is a table illustration of the confusion matrix.

Table 3. *Confusion Matrix* [16]

Confusion Matrix		**Actual Values**	
		Positive	**Negative**
Predicted Values	**Positive**	TP	FP
	Negative	FN	TN

.

To measure the performance of the system model in calculating predictive results from test data by calculating accuracy, recall, precision, and f1-score[17]. Here are the formulas for calculating evaluations:

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

$$recall = \frac{TP}{TP+FN} \quad (5)$$

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$f1 - score = 2\ x\ \frac{Recall\ x\ Precision}{Recall+Precision} \quad (7)$$

3. RESULT AND DISCUSSION

At this evaluation stage, it is a system test to see the performance results of the model built by carrying out several test scenarios. The first scenario is at the preprocessing stage. This test aims to determine the effect of using the process with and without the stemming stage. Next, the second scenario is at the feature selection stage. This second test aims to determine the effect of the process with and without the feature selection stage. Finally, in the third scenario, the testing phase is carried out with and without k-fold cross-validation.

3.1 The Effect of Preprocessing

Scenario one is testing at the preprocessing stage, where this test uses two datasets, namely data that goes through the stemming stage and without the stemming stage. The next step is the feature selection stage and classification stage. Next, the researchers compared the performance of these test results.

Table 4. The Effect of Preprocessing

Dataset	Accuracy	F1-Score
With Stemming	**81,60%**	**71,52%**
Without Stemming	80,01%	64,70%

Based on the test results using scenario 1 in "Table 4", the data that went through the stemming stage at the preprocessing stage got better performance results than those without the stemming stage. The difference in the average performance of the accuracy value is smaller than the average performance value of the f1-score. The f1-score value has the farthest difference, which is 6.82%. Meanwhile, the difference in the average accuracy value is 1.59%. The stemming stage is the process of processing data by cutting affixes such as prefixes, suffixes or combinations. At this stage, one or more words will go through a process of eliminating their affixes into one root word. Examples of cutting affixes in the dataset, such as the words "dipakai" or "memakai" into "pakai", by cutting these words, the system makes it easier to process words so that it causes an increase in the average performance of the built model.

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

3.2 The Effect of Feature Selection

In scenario 2, testing is carried out at the feature selection stage. This test uses data that goes through the feature selection stage and without feature selection. In scenario 1, the best results were obtained, namely by using a dataset with a stemming step, so that in the second scenario, using a dataset that goes through the stemming stage at the preprocessing stage. Next, the data goes through the classification and evaluation stages to determine the system's performance.

Table 5. The Effect of Feature Selection

Dataset	Accuracy	F1-Score
With Feature Selection	**81,60%**	**71,52%**
Without Feature Selection	79,74%	69,60%

Based on the test results using scenario 2 in "Table 5", the data that went through the stemming stage and the feature selection stage produced a better performance value than the data that went through the stemming stage without the feature selection stage. The difference in the average f1-score value reaches 1.92%. Meanwhile, the difference in the average value of accuracy is 1.86%. Feature reduction aims to improve performance accuracy by taking important and needed features with calculations using *chi-square*. The dataset has 8294 original features and selected the best 5000 features from the *chi-square* value. This stage causes the system work to increase with an average accuracy of 81.60% and an average f1- score 71.52% higher than the system without going through the feature selection stage.

3.3 The Effect of cross-validation

In scenario 3, testing is carried out using the best dataset from the results of the previous scenario. This test uses a dataset that goes through the stemming and feature selection stages. The next stage is classification with MBRF. This test aims to compare the performance results with and without *k-fold cross-validation.*

Table 6. The Effect of Cross-validation

Dataset	Accuracy	F1-Score
With K-Fold	**81,75%**	**71,90%**
Without K-Fold	81,60%	71,52%

Based on scenario 3 in "Table 6", it can be seen that the system using 5-fold cross-validation gets better performance results than the test without k-fold cross-validation. K-fold has a way of working by dividing the data as many as k-partitions randomly. Then, the data goes through the classification process. The system can produce an average value of accuracy and f1-score that is 81.75% and 71.90%. Therefore, in this study, the data use of k-fold resulted in better performance. However, the performance results with k-fold did not improve significantly.

3.4 Best Model

After doing three scenarios, so in this study to get the best model. The best model is built by going through the preprocessing stage with the stemming stage, through the feature selection stage, and using k-fold cross-validation.

Table 7. Best Model

Dataset	Accuracy	F1-Score
Price	93,80%	92,80%
Packaging	88,40%	68,60%
Product	59,80%	56,80%
Aroma	85%	69,40%
Average	**81,75%**	**71,90%**

average accuracy value of 81.75% and an average f1 score of 71.90%. The dataset has four aspect levels: price, packaging, product and aroma. And also, the dataset has three classes: negative, positive, and neutral. "Table 7" shows that the value of accuracy and f1-score on the price aspect is higher than the other aspect levels. The difference in the performance value is due to an imbalance in data distribution. Causes of class division imbalance include inconsistent and inappropriate labelling processes. The price aspect is more balanced than other aspects of distribution classes. There is a relatively flat distribution of positive, negative and neutral classes in the price aspect. In comparison, the product aspect has an imbalanced class distribution with more positive classes than negative and neutral classes. Because there are less specific features in the product aspect, the labelling is inconsistent. Furthermore, for the distribution of classes on the packaging and aroma aspects, the distribution of neutral classes is more than the positive and negative classes.

Journal of Information System Research (JOSH)
Volume 4, No. 1, Oktober 2022, pp 1−7
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v4i1.2047

4. CONCLUSION

Based on the research that has been done, researchers have built a system to analyze the sentiment of reviewing Indonesian-language beauty products on the female daily website using the *Modified Balanced Random Forest* (MBRF) classifier and *chi-square* as feature selection. At the testing stage, there are three test scenarios. Scenario one compares system performance results through the stemming stage and without going through the stemming stage. In addition, this test aims to determine the effect of the preprocessing stage. Then the second scenario compares the system's performance through the feature selection stage and without the feature selection stage. Furthermore, in the third stage, determine the effect of using system testing with and without k-fold cross-validation The researcher concludes that the three tests, the selection of the preprocessing stage, feature selection stage, and k-fold cross-validation, affect the performance of sentiment analysis results in beauty product reviews. And the classification method used helps to deal with imbalanced datasets in this study. The best result is using the dataset through the preprocessing stage using stemming, through the feature selection stage and performance test with k-fold with the best average accuracy value of 81.75% and the f1-score average value of 71.95%. There are several suggestions for further research. The first is to do more specific and accurate labelling. The second suggestion is to check the word dictionary list on stopwords and normalization to avoid words that affect performance. Then use other feature extraction and selection methods. And then try to modify the classification method to get a more efficient system and better performance.

REFERENCES

[1] H. Rika, "88,1 Persen Pengguna Internet Belanja dengan E-Commerce," *https://www.cnnindonesia.com/*, 2021. https://www.cnnindonesia.com/ekonomi/20211111123945-78-719672/881-persen-pengguna-internet-belanja-dengan-e-commerce.
[2] Female Daily, "Female Daily," 2022. https://femaledaily.com/.
[3] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, 2011, doi: 10.1016/j.eswa.2010.12.147.
[4] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.
[5] H. Ardian and S. Kosasi, "Analisis Sentimen Pada Review Produk Kosmetik Bahasa Indonesia Dengan Metode Naive Bayes," *J. ENTER*, vol. 2, no. 1, pp. 306–320, 2019.
[6] Y. Hegde and S. K. Padma, "Sentiment analysis using random forest ensemble for mobile product reviews in kannada," in *Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017*, Jul. 2017, pp. 777–782, doi: 10.1109/IACC.2017.0160.
[7] D. Irvantoro, "Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Review Produk Elektronik," *PhD Thesis. Univ. Muhammadiyah Jember*, no. 1410651199, 2019.
[8] Z. P. Agusta and Adiwijaya, "Modified balanced random forest for improving imbalanced data prediction," *Int. J. Adv. Intell. Informatics*, vol. 5, no. 1, pp. 58–65, 2019, doi: 10.26555/ijain.v5i1.255.
[9] F. N. Zamzami and M. D. P, "Analisis Sentimen Terhadap Review Film Menggunakan Metode Modified Balanced Random Forest dan Mutual Information," vol. 5, no. April, pp. 415–421, 2021, doi: 10.30865/mib.v5i2.2835.
[10] T. B. Rohman, D. D. Purwanto, and J. Santoso, "Sentiment Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest," *Pros. Semin. Nas. Teknol. Inf. Apl.*, pp. 7–11, 2018.
[11] D. B. Satmoko, P. Sukarno, and E. M. Jadied, "Peningkatan Akurasi Pendeteksian Serangan DDoS Menggunakan Multiclassifier Ensemble Learning dan Chi-Square Pendahuluan Studi Terkait," vol. 5, no. 3, pp. 7977–7985, 2018.
[12] C. F. Suharno, M. A. Fauzi, and R. S. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square," *Syst. Inf. Syst. Informatics J.*, vol. 3, no. 1, pp. 25–32, 2017, doi: 10.29080/systemic.v3i1.191.
[13] M. A. A. Jihad, Adiwijaya, and W. Astuti, "Analisis sentimen terhadap ulasan film menggunakan algoritma random forest," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10153–10165, 2021.
[14] J. A. Septian, T. M. Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," *J. Intell. Syst. Comput.*, no. September, pp. 43–49, 2019.
[15] N. D. Pratama, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2982–2988, 2018.
[16] A. K. Santra and C. J. Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering," *Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 322–328, 2012.
[17] M. Awad and R. Khanna, *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, no. April. 2015.