OPEN ACCESS
J DATA SCI APPL, VOL. 4, NO. 1, PP.031-042, JANUARY 2021

E-ISSN 2614-7408
DOI: 10.34818/JDSA.2021.4.71

JOURNAL OF DATA SCIENCE AND ITS APPLICATIONS

Sentiment Analysis of Beauty Product Reviews Using the K-Nearest Neighbor (KNN) and TF-IDF Methods with Chi-Square Feature Selection

Yusrifa Deta Kirana[1], Said Al Faraby[2],

School of Computing, Telkom University
Bandung, Indonesia

saidalfaraby@telkomuniversity.ac.id

Received on 05-08-2021, revised on 09-10-2021, accepted on 15-10-2021

Abstract

The rise of beauty products in recent times can make consumers hesitate to choose a beauty product, especially for women. Beauty product reviews have become a very valuable source of information for consumers in making decisions to purchase a product in improving their products and marketing strategies. The process of sentiment analysis on negative and positive beauty product reviews will be classified one by one. Therefore, in this study, sentiment analysis was applied to the beauty product review data using the K-Nearest Neighbor (KNN) method to find the best k in the case of this study. The dataset used will be pre-processed with case folding, noise removal, tokenization, stemming, stopword removal, and slang words, for feature extraction using Term Frequency Inverse Document Frequency (TF-IDF) to calculate the weight of a word in the document, and The feature selection method uses Chi-Square which aims to select the features needed to increase the accuracy value. In this study, the best accuracy value was 71% of the data classified using KNN with a k value of 50 and the model on feature selection with 76 features.

Keywords: K-Nearest Neighbors (KNN), Term Frequency Inverse Document Frequency (TFIDF), Chi-Square, Sentiment Analysis, Product Reviews

I. INTRODUCTION

Beauty products or commonly referred to as cosmetics are treatments used to treat improve the appearance or smell of the human body [1]. The rise of beauty products in recent times can make consumers hesitate to choose a beauty product, especially for women. Every consumer must-see product reviews before buying. Product Reviews are ratings given by users on a product that is used. The product review itself can be a review that is positive, negative, or neutral, this can be used as an evaluation material to buy a product. The importance of the role of beauty product reviews on women is the author's reason for choosing this problem. In some E-Commerce such as femaledaily.com provides a review column to help consumers make decisions from information on reviews of a beauty product.

In this journal discusses the classification of beauty product reviews that will be tested using the KNN method and added feature selection with chi-square. In this journal, the author will test datasets taken from the femaledaily.com website which are translated into Indonesian, then extracted using the TF-IDF method and feature selection using the chi-square method and classified using the KNN method.

II. LITERATURE REVIEW

Sentiment analysis is a type of text classification that is often referred to as opinion mining which involves analyzing people's opinions on entities such as products, services, organizations, individuals, problems, events, topics, and their attributes from a piece of text.[2] Research on sentiment analysis has been carried out by several people, one of which is in research [3] comparing the classification process using the Naive Bayes method, KNN, SVM, and Random Forest, the results of this study KNN is a method that produces accuracy by 96.8%.

The study [16] used feature selection and K-NN in the categorization of exam questions, where feature selection here is proven to increase accuracy. In this study, 600 training data were comparing the classification using KNN without feature selection, which was much lower, while KNN using the Chi-Square method had a difference of 16.81 and achieved an accuracy of 79.36%.

Research [19] states that the use of chi-square feature selection and classifying review text in tweet reviews can also increase the accuracy of the F1-Score. Where in this study compares the effect of the value of k on KNN with the use of chi-square feature selection. Results of research[19] by 78% with the selection of k on the KNN which is 15.

Based on previous research, the author uses sentiment analysis on this beauty product review, it is necessary to preprocess the data at an early stage to manage the data. Then proceed with data processing with feature extraction using TFIDF, feature selection using the Chi-Square method, classification using the K-NN method, and measure the performance of all features using the confusion matrix.

In this study, the author will build a model for sentiment analysis on beauty product reviews. This research focuses on the preprocessing process, feature selection, and feature extraction. In the processing process, the author will compare the effect on the stemming and stopword removal methods. The feature selection process uses the chi-square method, the feature extraction process uses TF-IDF. Then in the classification process using the KNN method. The dataset used is taken from the femaledaily.com website as many as 3960 data using Indonesian and English reviews, labeling the dataset based on price, packaging, product, and aroma with positive, neutral, and negative sentiment classes.

Based on the background and limitations described above, the objectives to be achieved are: This study focuses on applying sentiment analysis to beauty product reviews on a dataset taken from the femaledaily.com website using the KNN classification method with the right k selection and also analyzing the performance of the chi-square feature selection against the KNN algorithm in conducting sentiment analysis on beauty product reviews. This study also examines how the effect of using stemming and stopword removal in the preprocessing process on the results of sentiment analysis on beauty product reviews.

Related Research is a [19] In writing this final project, the author uses several references regarding research related to sentiment analysis on beauty product reviews. There have been many studies on good sentiment analysis about a product, film, brand, even comments on social media. [20] In research [4] conducted by Zuhdiyyah Ulfah Siregar, Riki Ruli A. Siregar, Rakhmat Arianto with the title "Classification of Sentiment Analysis on Training Participants Comments Using the K-Nearest Neighbor Method" proposes that the more training data used, the higher the accuracy of the test. In this study, through the preprocessing stage with sentiment analysis, calculating word weighting with TFIDF and calculating the distance between test data and training data using cosine similarity which produces an accuracy rate of 94.23%. [21] Another study by Mary Sowjanya, K. Sridiva with the title [5] "Aspect Based Sentiment Analysis using POS Tagging and TFIDF" conducted an aspect-based sentiment analysis on a comment on social media. In this study, we compare model 1 using POS Tagging and model 2 using TFIDF and also compare the classification method using Naïve Bayes, KNN, Decision Tree, and SVM methods. In this study the performance of TFIDF is better than POS Tagging, the accuracy obtained on TFIDF and nave Bayes is 75.71%, TFIDF with KNN is 80.0%, TFIDF with Decision Tree is 87.30%, and TFIDF with SVM 94 %. The conclusion that can be drawn is that the difference between the classifications in the method is not too high. [22] Research on sentiment analysis on beauty product reviews has also been conducted by [23] [19] used the Naïve Bayes classification method which focused on identifying aspect-based beauty product reviews, and in research [20] researchers translated beauty product reviews into Indonesian. The study [20] obtained an F1-Score of 62.82%.

Product Reviews is a according to Mo, et al, in Saripa, 2019 argues that reviews contain images that reflect the actual quality of the goods, such as color problems, inconsistent specifications, usage problems or high quality and a good experience. According to Spink Nurul, product reviews are reviews given by customers on the product [13].

Sentimen Analysis is a according to Feldman, sentiment analysis is the task of finding an opinion from the author about a particular entity. According to Tang, sentiment analysis on reviews is the process of investigating product reviews on the internet to determine opinions or feelings about a product as a whole. According to Thelwall in [11], sentiment analysis is treated as a classification task that classifies oriental texts into positive or negative [12].

III. Research Method

A. System Description

The system built in this study is a system that can analyze the sentiment contained in beauty product reviews into positive, negative, and neutral sentiments using the K-Nearest Neighbor algorithm. The sentiment analysis process in this study was taken from a dataset based on a beauty product review from the femaledaily.com website which provides a feature to provide reviews. The dataset that has been taken is then preprocessed using the Chi-Square selection feature and the TF-IDF weight extraction feature. In this study, there are 4 models in the classification, namely based on the label aspects of price, packaging, product, and aroma. The system design built can be seen in Figure 1.

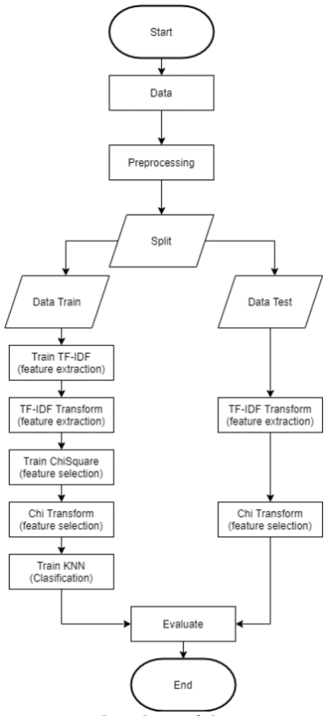

Fig. 1. Flowchart of the System

B. Dataset

In this study, the authors use a dataset derived from beauty product reviews on the femaledaily.com website page which uses Indonesian and English reviews, but the Indonesian language is more dominant in this beauty product review. In this dataset, there are 3960 data manually labeled into 4 aspects, namely price, packaging, product and aroma aspects. Label 1 is used for positive reviews, label 0 is used for neutral reviews, and label -1 is used for negative reviews. The price aspect contains negative reviews (-1) with 716 reviews (18%), neutral (0) with 2187 reviews (55%), and positive with 1057 reviews (26%). In the packaging aspect, it contains negative reviews (-1) with a total of 189 reviews (4.7%), neutral (0) with a total of 3324 (83%), and positive (1) with a total of 447 reviews. (11.3%). In the aspect of the product containing negative reviews (-1) with a total of 688 reviews (17%), neutral (0) with a total of 659 reviews (16%), and positive reviews (1) with a total of 2612 reviews (65%). The aroma aspect contains negative reviews (-1) with a total of 218 reviews (5.5%), neutral (0) with a total of 3077 reviews (77%), and positive reviews (1) with a total of 669 reviews (16%). The results of the dataset representation used in this study are in the table below.

TABLE I
DATASETS USED

Product Review Text	Price	Packaging	Product	Fragrance
Sunscreen termahal yang pernah gue beli ini kayanya. but it's worth it sih and will definitely buy again. sukanya sama suncreen ini: - high spf - nggak meninggalkan white cast. Perfectly blends into the skin - nggak membuat muka berminyak - very light - doesn't clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen :)	-1	0	1	0

C. *Preprocessing Data*

This stage is the initial stage to prepare text data so that it can be processed in the classification process. Sometimes in the dataset, there are various problems that can change the results of the dataset itself, such as missing values in inappropriate data. For this reason, in overcoming these problems, a preprocessing stage is needed on the dataset [17]. There are several stages carried out in data preprocessing,

1) *Noise Removal*

At this stage, it is done to remove punctuation marks, characters, links, and numbers. The following table results from noise removal:

TABLE II
PROCESS PREPROCESSING NOISE REMOVAL

Prior Data	*Noise Removal* Results
Sunscreen termahal yang pernah gue beli ini kayanya. but it's worth it sih and will definitely buy again. sukanya sama suncreen ini: - high spf - nggak meninggalkan white cast. Perfectly blends into the skin - nggak membuat muka berminyak - very light - doesn't clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen :)	Sunscreen termahal yang pernah gue beli ini kayanya but it s worth it sih and will definitely buy again sukanya sama suncreen ini high spf nggak meninggalkan white cast perfectly blends into the skin nggak membuat muka berminyak very light doesn t clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen

2) *Case Folding*

At this stage, a product review sentence is converted into all lower case letters. The following are the results of the case folding table:

TABLE III
PROCESS PREPROCESSING CASE FOLDING

Prior Data	*Case Folding* Results
Sunscreen termahal yang pernah gue beli ini kayanya but it s worth it sih and will definitely buy again sukanya sama suncreen ini high spf nggak meninggalkan white cast perfectly blends into the skin nggak membuat muka berminyak very light doesn t clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen	sunscreen termahal yang pernah gue beli ini kayanya but it s worth it sih and will definitely buy again sukanya sama suncreen ini high spf nggak meninggalkan white cast perfectly blends into the skin nggak membuat muka berminyak very light doesn t clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen

3) *Tokenization*

At this stage, a sentence is broken/beheaded based on each word that composes it. The following is the result of the Tokenization table:

TABLE IV
PROCESS PREPROCESSING TOKENIZATION

Prior Data	*Tokenization* Results
sunscreen termahal yang pernah gue beli ini kayanya but it s worth it sih and will definitely buy again sukanya sama suncreen ini high spf nggak meninggalkan white cast perfectly blends into the skin nggak membuat muka berminyak very light doesn t clog pores produk ini berhasil membuat gue jadi mau pakai sunscreen	['sunscreen', 'termahal', 'yang', 'pernah', 'gue', 'beli', 'ini', 'kayanya', 'but', 'it', '""s""', 'worth', 'it', 'sih', 'and', 'will', 'definitely', 'buy', 'again', 'sukanya', 'sama', 'suncreen', 'ini', 'high', 'spf', 'nggak', 'meninggalkan', 'white', 'cast', 'perfectly', 'blends', 'into', 'the', 'skin', 'nggak', 'membuat', 'muka', 'berminyak', 'very', 'light', 'does', '""n't""', 'clog', 'pores', 'produk', 'ini', 'berhasil', 'membuat', 'gue', 'jadi', 'mau', 'pakai', 'sunscreen']"

4) *Stopword Removal*

Stopword Removal is the process of taking important words from the previous tokenization results by removing words that have no meaning. What can be called stopwords (common words) in Indonesian and English, examples in Indonesian are: di, too, from, yang, and, or, this, that, and others, and in English such as my, the, is, it, but, and others. In English, the words are taken from the English-language reviews which are removed using the English-language stopword removal library. The following is an example of the stopword removal process. The following are the results of the stopword removal table:

TABLE V
PROCESS PREPROCESSING STOPWORD REMOVAL

Prior Data	S*topword Removal* Results
['sunscreen', 'termahal', 'yang', 'pernah', 'gue', 'beli', 'ini', 'kayanya', 'but', 'it', """s""", 'worth', 'it', 'sih', 'and', 'will', 'definitely', 'buy', 'again', 'sukanya', 'sama', 'suncreen', 'ini', 'high', 'spf', 'nggak', 'meninggalkan', 'white', 'cast', 'perfectly', 'blends', 'into', 'the', 'skin', 'nggak', 'membuat', 'muka', 'berminyak', 'very', 'light', 'does', """n't""", 'clog', 'pores', 'produk', 'ini', 'berhasil', 'membuat', 'gue', 'jadi', 'mau', 'pakai', 'sunscreen']"	['sunscreen', 'termahal', 'gue', 'beli', 'kayanya', """s""", 'worth', 'sih' 'will', 'definitely', 'buy', 'again', 'sukanya', 'suncreen', 'high', 'spf', 'nggak', 'meninggalkan', 'white', 'cast', 'perfectly', 'blends', 'into', 'the', 'skin', 'nggak', 'muka', 'berminyak', 'very', 'light', 'does', """n't""", 'clog', 'pores', 'produk', 'berhasil', 'gue', 'pakai', 'sunscreen']

5) *Stemming*

At this stage, it is done to find the root word of each word from the stopword removal in the previous stage, by removing all affixes, both at the beginning, insertion and at the end of the word. To carry out the stemming process, the author will use the stemming library with Sastrawi. Literary libraries are usually used to carry out the stemming process using Indonesian. The following are the results of the Stemming table:

TABLE VI
PROCESS PREPROCESSING STEMMING

Prior Data	S*topword Removal* Results
sunscreen termahal gue beli kayanya 's worth sih will definitely buy again sukanya suncreen high spf nggak meninggalkan white cast perfectly blends into skin nggak muka berminyak very light doesn't clog pores produk berhasil gue pakai sunscreen	sunscreen mahal gue beli kaya but s worth sih will definitely buy again suka suncreen high spf nggak tinggal white cast perfectly blends into skin nggak muka minyak very light does n't clog pores produk hasil gue pakai sunscreen

6) *Slang Word*

At this stage, it is done to change non-standard words into standard words and also change abbreviated words into long ones. To carry out this slang word process, the author enters a slang_word dictionary that can change non-standard words into standard words, and also change abbreviated words. The dictionary slang_word[23] is obtained through the GitHub platform which uses dictionaries to convert non-standard languages into standard ones. In this process, the author also translates reviews that use English into Indonesian manually. As in the table below the word "sunscreen" will be changed to the word "sunscreen". The following table results from slang words:

TABLE VII
PROCESS PREPROCESSING SLANG WORD

Prior Data	*Slang Word* Results
sunscreen mahal gue beli kaya s worth sih will definitely buy again suka suncreen high spf nggak tinggal white cast perfectly blends into skin nggak muka minyak very light does n t clog pores produk hasil gue pakai sunscreen	tabir surya mahal saya beli, kaya s bernilai pasti akan membeli lagi Saya suka tabir surya spf tinggi tidak ada gips putih yang tersisa menyatu sempurna dengan kulit tidak ada minyak wajah saya, sangat ringan tidak menyumbat pori-pori produk saya menggunakan tabir surya

D. Feature Extraction

Feature extraction is the process of taking features that can describe the required information. In this study, the feature extraction process uses TF-IDF. Term Frequency-Inverse Document Frequency (TF-IDF) is a method used to give weight to terms as a strategy to classify documents [7]. The weighting process on terms uses TF-IDF, consisting of calculating the value of TF(term frequency) and IDF(Inverse Document Frequency. TF(Term Frequency) which is calculating the frequency of occurrence of words which are usually transformed into log values to, and IDF(Inverse Document Frequency).) which is the calculation of a term that is widely distributed throughout the document, TF-IDF is useful for calculating the weight of each word resulting from preprocessing. Word weighting is used to extract words into a numeric format that can represent the data as a whole. The results of this process are in the form of a matrix consisting of rows and columns, data as rows, and features as columns[18]. The following is the weighting calculation process using the TF-IDF method:

$$TF\ (tk, dj) = \ f\ (tk, dj)$$

$(tk,$) is the number of occurrences of the term (k) in the document (j)

$$IDF(tk\ ) = \log \frac{N}{df\ (t)}$$

idf is the value of the inverse document frequency, while N is the total number of documents and idf is the number of documents containing the word i.

$$w(\text{t,d}) = t(f, d) \times idf(t)$$

wt,d is the weight of term (t) against the document (d), while tff,d is the number of occurrences of term(t) in the document (d) and idf$t$= the value of inverse document frequency.
For example, the calculations in the dataset used in this study are attached in the appendix.

E. Feature Selection

The next process is feature selection, after getting the data that has been extracted using TF-IDF, the feature selection will be carried out using the Chi-Square method. Chi-Square is a feature selection algorithm to measure the lack of flexibility between categories and terms[10] Chi-square is also one of the feature selection methods that is widely used in research. The feature selection method has been proven to increase accuracy in several previous studies [8][9]. Feature selection is carried out for each word that will calculate the Chi-Square value using the following equation:

$$X^2 \ = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$$

Information:
O_i = oneservation value to - i
E_i = i-th expectation value

Chi-Square is a feature selection that looks at the dependence of the term with its category, where several uhi conditions on the chi-square can be used, namely [22]:

1) There are no cells with an actual count value or reality frequency (F0) which is 0 (zero)
2) If the contingency table is in the form of 2x2, then there should be no cells with an expected count or expected frequency (Fn) of less than 5
3) If the table form is more than 2x2, for example, 2x3, then the number of cells with an expected frequency (Fn) of less than 5 cannot be more than 20%.[21]

F. Classification

The next process is classification, in this study the K-Nearest Neighbor (KNN) algorithm method is used as a model in the classification of sentiment analysis on beauty product reviews. K-Nearest Neighbor (KNN) is one of the popular methods for generating text classification (Sreemathy and Balamurungan, 2012). The K-Nearest Neighbor (KNN) algorithm is one method for classifying objects based on learning data, namely the object approach [6] which aims to group the closest K. This KNN algorithm belongs to the category of supervised algorithms. where the process is based on the comparison of the value of k nearest neighbors. The parameter k has a big influence on the prediction results. K-Nearest Neighbor (KNN) measures the similarity of data to measure distance. To calculate the distance between the data used in this KNN method using Euclidean distance with the following equation:

$$d = \sqrt{\sum_{i=1}^{p}(x_{2i} - x_{1i})^2}$$

Based on equation , d is the Euclidean distance, x2i is the value of the i-th test data, x1i is the value od the i-th test data and p is the number of attributes. Broadly speaking, the KNN determines the number of k in the KNN, calculates the distance to a data object than has been selected against all data trains, then sorts the distance result by value in ascending order, then selects several closest neighbor object as desired, then the last one is to determine a class with the highest frequency at the nearst k.

G. Evaluate

After the classification process is carried out, the last stage is the performance evaluation of the system used to measure the performance of the system to be built using the classification method. K-Nearest Neighbor (KNN). This performance can be measured using several methods, here the author uses one method, namely the confusion matrix. Confusion Matrix is one method to measure the performance of a classification system that will compare the data from the classification with the actual data. With the help of the results from the confusion matrix, it can also be done to calculate accuracy, precision, and recall [15]. A confusion matrix is a method for measuring the performance of a system by comparing the predicted data with the actual data. Accuracy is an evaluation parameter of the results of the system being built. Accuracy is referred to as the behavior of the closeness between the predicted results and the actual (actual) value. Precision is the level of accuracy between the requested information and the results provided by the system. The recall is the success rate of the system in retrieving information. And the F1 measure is a combination of the recall value with the precision value that has been calculated previously. The following is a table of the confusion matrix:

TABLE VIII
CONFUSION MATRIX

Predicted Value	Actual Value	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Based on the confusion matrix table above, TP is data that is predicted to be positive and data that is actually positive, FP is data that is predicted to be positive but data is actually negative, TN is the data that is predicted to be negative and the data that is negative, and FN is the data that is predicted to be negative but the data is positive. Furthermore, from the results of the confusion matrix, we can find the value of the f1-score, precision, and recall of the classification system built on each label. The author uses a macro average to get the average multilabel accuracy.

IV. RESULTS AND DISCUSSION

A. Evaluate

the data that is predicted to be negative and the data that is actually negative, and FN is the data that is predicted to be negative but the data is actually positive. Furthermore, from the results of the confusion matrix, we can find the value of the f1-score, precision, and recall of the classification system built on each label. The author uses a macro average to get the average multilabel accuracy. In this study, there were 3 kinds of test scenarios carried out. The test in the first scenario aims to determine the effect of the use of preprocessing on stemming and stopword removal. The second test was conducted to determine the effect of feature selection on the dataset. Then the third test was carried out to determine the effect of choosing k on KNN on the classification results. In this test, the data used is a dataset derived from beauty product reviews through the official website femaledaily.com which uses Indonesian and English reviews with a total of 3960 data. And the test was carried out with 4 models based on their aspects, namely price aspects, packaging aspects, product aspects, and aroma aspects with 3 different scenario trials.

1) *Testing the Effect on Stemming and Stopword Removal*

In the first scenario, this was carried out to determine the effect of the process on the use of preprocessing stemming and stopword removal in data sharing conditions, namely 20% test data and 80% train data and the K value on KNN was 50 by using processes during preprocessing data such as noise removal, casefolding, tokenization, and slangwords, then using the feature extraction process with TF-IDF, feature selection using the chi-square method with 1% features, i.e. 76 features. Test results can be seen as follows:

TABLE IX
TEST RESULTS SCENARIO 1

Stemming	Stopword Removal	Performance			
		Accuracy	Precission	Recall	F1-Score
Y	Y	71%	44,25%	35%	31,75%
Y	N	**72,50%**	51,25%	36,25%	**33%**
N	Y	71,50%	52,25%	36%	32%
N	N	71,50%	43,50%	35,25%	31,75%

Based on the table above, the use of stemming and stopword removal in the data preprocessing process is very influential, in the first experiment the author examined if the data processed during preprocessing used stemming and stopword removal which resulted in an accuracy value of 71%. In the second experiment, the author examined if the data was processed using stemming but did not use stopword removal which resulted in an accuracy value of 72.50%. In the third experiment, the author examined if the data processed did not use stemming but used stopword removal, the accuracy obtained was 71.50%. In the last experiment, the authors examined not using stemming and also not using stopword removal which resulted in an accuracy value of 71.50%. So that the greatest results are obtained when the system uses the stemming process but does not use stopword removal with an accuracy value of 72.50% and an F1-Score value of 33%. This is because during the stemming process each word is successfully converted into its root word, so as to minimize the use of features that have no effect on the next process. In the use of stopword removal, words that have important information and meanings are removed so that they can change the meaning of different sentences. In one of the reviews there is the word "no" which can change the meaning of the labeling, as shown in the table below. In this beauty product review, it should have a positive value because the review is "not sticky", but during the stopword removal process the word "no" is removed so that the meaning of the label which was originally positive becomes negative.

TABLE X
AFTER PROCESS STEMMING AND STOPWORD

Dataset	Review	Price	Product
Original	cukup mahal memang, tapi worth it. nge blend di wajah. **tidak lengket** meskipun teksturnya kental (krim).	-1	1
Preprocessing Complete	mahal bernilai nge blend wajah **lengket** tekstur kental krim	-1	-1
without stopword	cukup mahal memang tapi bernilai it nge blend di wajah **tidak lengket** meski tekstur kental krim	-1	1

2) *Testing the Effect on Feature Selection*

In the second scenario, this was carried out to determine the performance of the selection features used in this study under conditions using all preprocessing processes from noise removal, casefolding, tokenization, stemming, stopword removal and slang words, then using the feature extraction process with TF-IDF and value K on KNN is 50. Test results can be seen as follows:

TABLE XI
TEST RESULTS SCENARIO 2

Ratio	Performance			
	Accuracy	Precission	Recall	F1-Score
Without selection	70,75%	41,75%	34,75%	31,35%
75%	71%	26,25%	33%	27,50%
50%	71%	26,25%	33%	27,50%
30%	71%	26,25%	33%	27,50%
5%	71%	34,50%	33,25%	27,75%
1%	**71%**	44,25%	35%	**31,75%**

In the first experiment, the author did not include the chi-square feature selection, the accuracy value was 70.75%. In the second experiment, the authors enter a 75% model with the selected data, namely 5633 features, the accuracy results obtained are 71%. In the third experiment, the author entered a 50% model with the selected data, namely 3755 features, the results obtained were 71% accuracy. In the fourth experiment, the author entered a 30% model with the selected data, namely 2253 features, the accuracy obtained was 71%. In the fifth experiment, the author entered a 5% model with the selected data, namely 376 features, the accuracy obtained was 71%. In the last experiment, the author chose a 1% model with 76 features selected data, the results obtained were 71% accuracy. Of the six experiments that have been carried out, the sixth experiment with the selection of a 1% model with 76 features that produces the highest accuracy value and F1-Score value is 31.75%, this is because the number of features used will affect the F1-Score and F1-Score values. The effect of feature selection is that the fewer features you take, the greater the F1-Score value you get, this also happens because of unselected product reviews.

3) Testing the Effect of the value of K on KNN

In the third test scenario, this is done by comparing the K value in KNN using data sharing as much as 20% test data and 80% train data, all processes during data preprocessing such as noise removal, case folding, tokenization, stopword removal, stemming and slang words, then use the process feature extraction with TF-IDF, chi-square selection of features that use 1% data with 76 features, and text classification which is applied in 4 aspects, namely price, packaging, product, and aroma, then the evaluation results are obtained using the confusion matrix to measure performance with the average of a system by comparing the predicted data with the actual data as follows:

TABLE XII
THE RESULTS SCENARIO 3

K Value on KNN	Performance			
	Accuracy	Precission	Recall	F1-Score
2	63%	42%	39,75%	38%
5	67,25%	48,75%	38,50%	38,50%
7	69%	53,25%	38,25%	37,75%
9	69,25%	51,25%	37,75%	36,50%
13	68,50%	48,25%	35,50%	33,50%
20	70%	42,25%	35,50%	33,25%
50	**71%**	44,25%	35%	31,75%

In testing this third scenario, the author conducted several experiments to find the best K on KNN, in this scenario the author conducted seven experiments, the first experiment was tried with k = 2 which resulted in an accuracy value of 63%. In the second experiment, it was tried with k = 5 which resulted in an accuracy value of 67.25%, then in the third experiment with k = 7 which resulted in an accuracy value of 69%, then in the fourth experiment with k = 9 which resulted in an accuracy value of 69.25%, then in the fifth experiment with k = 13 which resulted in an F-1 score of 68.50%, in the sixth experiment with k = 20 which produced an accuracy value of 70%, and in the last experiment with a value of k = 50, it resulted in an F1-Score of 71%. Based on the results of the above test as a whole, it can be seen that the k value in KNN greatly affects the accuracy value, the closer the distance k is to the KNN value, the more visible the sentiment classification will be, in the data above, the k value which greatly affects the F1-Score value is k=50 with an F1-Score of 71%. The comparison can be seen in the graph below:

V. Conclusion

Based on the results of several test scenarios that have been carried out to analyze sentiment on a beauty product, it can be concluded that the best system performance is produced when test data is 20% and train data is 80% using a combination of preprocessing, namely noise removal, case folding, tokenization, stemming. , stopword removal, and slang words, then added feature extraction using TF-IDF, using feature selection with chi-square and k selection on KNN which is 50 which produces an accuracy of 71%.

In the preprocessing process, the stemming technique has a fairly good effect, because in the stemming process the system has succeeded in removing affixes into their basic form which has meaning, thereby minimizing errors in words that have no effect. Then the stopword removal system is less successful in removing words that have no meaning because there are words that have meaning in this beauty product review such as the word "non-sticky" which should have a positive value, with the stopword removal process now the word "no" is removed so that the review becomes "sticky" which has a negative value. so that the results of the dataset from the preprocessing process that does not use stopword removal get the highest accuracy value of 72.50% because the data is balanced so that there are some unbalanced data in each class, as explained in the dataset section. The performance of the Chi-Square selection feature here is good enough to increase the accuracy of the F1-Score value because the selected data is only 1% or 76 features from 11919 features in this beauty product review data. After all, the accuracy value obtained is greater using the selection feature. which is 71% compared to not using the selection feature of 70.75%. The KNN method in conducting sentiment analysis on beauty product reviews can provide different values, depending on how the scenario is used and the selection of k. In this research, the writer found the best k for KNN performance, namely k=50, with an accuracy value of 71%.

Suggestions that can be applied for further research are the need to add a large number of datasets so that the variation in data increases, the amount of data with each aspect is different and can be applied properly. If it is not possible to overcome unbalanced data, it is also possible to add a feature selection method such as Multi-Information (MI) or Information Gain to compare the performance of chi-square feature selection and can overcome these problems.

ACKNOWLEDGMENT

The author would like to thank God Almighty, the family who always supports the author in completing this research. The authors thank Mr. Said and Mr. Mahendra as my mentors and thank you also to colleagues who always help the author in difficulties in doing this research.

REFERENCES

[1] Winter, Ruth (2005). A Consumer's Dictionary of Cosmetic Ingredients: Complete Information About the Harmful and Desirable Ingredients in Cosmetics (Paperback). US: Beginning Press. ISBN 1-877988-30-8.

[2] B. Liu (2012). Sentiment Analysis and Opinion Mining Sentiment Analysis and Opinion Mining.

[3] Adiwijaya (2020). Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor

[4] Zuhdiyyah Ulfah Siregar, Riki Ruli A. Siregar, Rakhmat Arianto."klasifikasi sentiment analysis pada komentar peserta diklat menggunakan metode k-nearest neighbor" Vol. 8, No. 1, April 2019,

[5] Kotagiri. Srividya, A. Mary Sowjanya. "Aspect Based Sentiment Analysis using POS Tagging and TFIDF" Volume-8 Issue-6, August, 2019

[6] Achmad Bayhaqy, Kaman Nainggolan, Sfenrianto, Emil R.Kaburuan. "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes " Oktober, 2018

[7] D. Zhu and J. Xiao, "R-tfidf, a Variety of tf-idf Term Weighting Strategy in Document Categorization," Seventh Int. Conf. Semant. Knowl. Grids R-tfidf, pp. 83–90, 2011.

[8] D. A. Abduljabbar and N. Omar, "Exam Questions Classification Based On Bloom's Taxonomy Cognitive Level Using Classifiers Combination," J. Theor. Appl. Inf. Technol., vol. 78, no. 3, pp. 447–455, 2015

[9] J. Ling and T. B. Oka, "Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square," vol. 3, no. 3, pp. 92–99, 2014

[10] Dhuha Abdulhadi Abduljabbar, Nazlia Omar. "Exam Questions Classification Based on Bloom's Taxonomy Cognitive Level using Classifiers combination", vol.78. no.3. 2015

[11] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis The Role of Text Pre-processing in Sentiment Analysis," Procedia Comput. Sci., vol. 17, no. December 2014, pp. 26–32, 2013.

[12] D. A. Muthia, "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Mengunakan Algoritma Naive Bayes," Jurnalilmu Pengetah. Dan Teknol. Komput., vol. 2, no. 2, pp. 39–45, 2017.

[13] Khafidatul Ilmiyah, Indra Krishernawan, "Pengaruh ulasan produk, kemudahan, kepercayaan, dan harga terhadap keputusan pembelian pada marketplace shopee di Mojokerto",vol.6, no. 1, Juni 2020

[14] A. A. Farisi, Y. Sibaroni, S. Si, S. Al Faraby, and M. Sc, "Analisis Sentimen pada Ulasan Hotel Menggunakan Multinomial Naïve Bayes Classifier."

[15] Gaol, S. L., Herdiani, A., Asror, I. (2019). Analisis Sentimen Komentar Masyarakat Terhadap Calon Presiden 2019 Pada Media Sosial Youtube Menggunakan Metode SVM dan Lexicon Based. Jurnal. S1 Informatika.Universitas Telkom, Bandung

[16] Indah Listiowarni1, Eka Rahayu Setyaningsih2, "Feature Selection Chi-Square dan K-NN pada Pengkategorian Soal Ujian Berdasarkan Cognitive Domain Taksonomi Bloom", Vol. 4, No. 1, Mei 2018,

[17] Sacra, S. A., S. Al Faraby., & D Triantiro M. KLASIFIKASI ANJURAN, LARANGAN, DAN INFORMASI PADA HADITS SHAHIH BUKHARI MENGGUNAKAN NAÏVE BAYES CLASSIFIER. e-Proceeding of Engineering, Vol. 4, No. 3, pp. 4794-4802, Des. 2017.

[18] L. G. Irham, A., Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," J. Media Inform. Budidarma, vol. 3, no. 4, pp. 284–292, 2019.

[19] C. Fresta, M. Ali Fauzi, R. Setya Perdana, "Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online menggunakan Metode K-Nearst Neighbor (K-NN) dan Chi-Square". Vol. 1, No. 10, Oktober 2017, hlm. 1000-1007

[20] C. Hasya Yutika, Adiwijaya, Said Al Faraby, " Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes", Jurnal Media Informatika, Vol. 5, No. 2, April 2021, Page 422-430

[21] A. Nisa, E. Darwiyanto, I. Asror, "Analisis Sentimen Menggunakan Naïve Bayes Clasifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi", Vol. 6, No. 2, Agustus 2019, Page 8650

[22] I. C. Negara and A. Prabowo, "Penggunaan Uji Chi-Square untuk Mengeahui Pengaruh Tingkat Pendidikan dan Umur terhadap Pengetahuan Penasun Mengenai HIV-AIDS di Provinsi DKI Jakarta," in Prosiding Seminar Nasional Matematika dan Terapannya 2018, Purwokerto, 2018.

[23] Liang Wu, Fred Morstatter, Huan Liu, "Building and Using a Sentiment Dictionary of Slang Words for Short-Text Sentiment Classification", Arizona State University, 17 Agustus 2016.