Sentiment Analysis of Consumers for Determining the Packaging Features of Eucalyptus Oil Products

Wilma Latuny[1a◆], Victor Oryon Lawalata[1b], Daniel Bunga Pailin[1c], Rahman Ohoirenan[1d]

Abstract. *This study aims to accurately predict eucalyptus oil packaging features and extract the most features to be improved for redesigning eucalyptus oil packaging. This research begins with taking consumer comments using a power query and then processing it using the data mining method and processed using WEKA to find sentiment analysis and accuracy of consumer comments regarding eucalyptus oil products. This study obtained the tendency of comments on each attribute with an assessment of the accuracy for all classes of 83% and each positive sentiment 3% of comments and 57% of comments for negative courses. The sentiment that shows the packaging tends to be normal at 20%, which is interpreted as neutral. This research can provide a suggestion to redesign the packaging based on the commentary sentiment of eucalyptus oil.*

Keywords: Sentiment Analysis, Eucalyptus Oil Packaging, Accuracy of Prediction

I. INTRODUCTION

One of the industrial sectors that contributes to the development process in Maluku is the small industry of eucalyptus oil refining. The existence of these small industries has contributed to strengthening the industrial structure in Indonesia, including in Maluku (Smith and Idrus, 2016). Eucalyptus oil is a mainstay because of its properties to provide warmth to the body, relieve itching from insect bites, relieves stomach aches, flatulence, colds, and its fragrant and aromatic aroma for health. These products are generally produced by small and medium industries (IKM) in Maluku.

Competition for eucalyptus oil products in the market is quite tight, making SMIs need innovation to have an advantage over their competitors. One of them is creating attractive product packaging designs to invite consumers to buy marketed products.

The appeal of a product cannot be separated from the packaging. Packaging is a "trigger" because it is directly dealing with consumers. Therefore, the packaging has to influence consumers to provide positive feedback (Dwiningsih, 2012).

The challenge for IKM relates to communication with consumers to find out the wants, needs, or attractiveness inherent in eucalyptus oil product packaging. Limited access to information and the ability to operate supporting technology has made the industry more reliant on consumer experience and "mouth to mouth" promotion to get responses from consumers regarding the quality of its products. At this time, people depend more on social media such as Facebook, Twitter, Instagram, and WhatsApp for access to information, communication, and trade. In a global online purchasing survey including Indonesia, as many as 71% of consumers review a product before buying the product. As many as 43% agree that social media is a tool to meet knowledge needs in product reviews and forum reviews to help make purchasing decisions (Manalu, 2014). Product reviews and forum reviews are submitted through comments on social media that contain complaints, praise, or views on products or services from an online store. The comments describe the different responses from each customer. Comments in the form of text can be collected and processed with sentiment analysis. This approach analyzes

[1] Department of Industrial Engineering, Faculty of Engineering, University of Pattimura, Jl. Ir. M. Putuhena, Kampus Poka, Ambon, Maluku, 97233, Indonesia.

[a] email: wlatuny@gmail.com
[b] email: victor.lawalata@fatek.unpatti.ac.id
[c] email: dani.ti.fatek@gmail.com
[d] email: rahmanohoirenan45@gmail.com
◆ corresponding author

opinions, sentiments, evaluations, judgments, attitudes, and public emotions towards entities such as products, services, organizations, individuals, problems, events, topics, and their attributes (Liu, 2012). Sentiment analysis studies opinions expressing positive or negative views (Ibrahim and Othman, 2012). The argument needed to conduct the analysis comes from the comments of the online shop page on Facebook. Facebook was chosen because its users interact massively with each other, where the total number of Facebook users is 1.44 billion, with daily users of 936 million (Hapsari et al., 2018). Since the increasing use of social media and the limited social activities of the community, opinion surveys will be more effective and efficient in using social media to improve the packaging of eucalyptus oil products.

Related study of sentiment analysis of the product packaging was conducted severely in the last decade. For instance, the research method using big data is possible to understand the customer's emotions and feelings (Kang et al., 2018). The product review has an eligible result by applying the sentiment analysis technique to gather insightful information from a consumer (Jebaseeli and Kirubakaran, 2012). In line with that, the findings of the product review feature using sentiment analysis and joint sentiment topic (JST) model have an accuracy prediction (Lee and Lyu, 2020). The extraction of customer sentiment comments of product reviews will be well predicted to obtain high accuracy using a support vector machine (SVM) compared to a naïve Bayes classifier (Shivaprassad and Shetty, 2017).

On the other hand, sentiment analysis is rarely found to predict the packaging product. For example, sentiment analysis was applied to gather feature for designing process of the product by employing machine-generated data to identify a series of customer need. The result showed many incorporating logical, computational methods into the highly subjective and creative design process (Ireland and Liu, 2018). The study to identify the to-be-improved product features with iPhone 6 showed that the survey data online combined with traditional survey may be a comprehensive method of evaluating customers to identify product development improvement (Xang et al., 2018). However, there is a limitation study related to applying sentiment analysis to redesign the essential oil package. Therefore, this study aims to predict eucalyptus oil packaging features accurately and extract the most features to be improved for redesigning eucalyptus oil packaging.

II. RESEARCH METHOD

Text mining is a cross-disciplinary that refers to information retrieval, data mining, machine learning, statistics, and computational linguistics (Han, 2012). Text mining, also known as text data mining or searching for knowledge in textual databases, is a semi-automatic process of extracting data from patterns (Turban, 2005).

The types of text mining work include categorization, text clustering, concept/entity
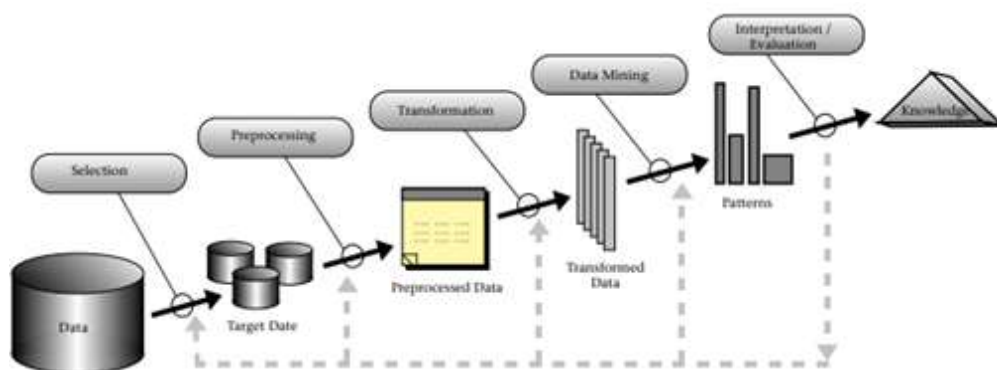
Figure 1. Process Text Mining (Andini, 2013)

extraction, sentiment analysis, document summarization, and entity-relation modeling (i.e., learning relationships between entities) (Han, 2012). Data sources used in text mining are collections of text that have to unformat, unstructured, or at least semi-structured. The purpose of text mining is to get helpful information from a set of documents.

Text mining is a variation of data mining that attempts to find interesting patterns from large amounts of textual data. The difference lies in the practices used. The text-mining pattern is taken from a set of unstructured natural languages, while in data mining, the pattern is taken from a structured database (Han & Kamber, 2006). There are several main process stages in text mining: text processing, text transformation or feature generation, feature selection, and pattern discovery or data mining.

Sentiment Analysis

The sentiment means the opinions or views that are based on exaggerated feelings towards something. Meanwhile, according to Merriam-Webster's Online Dictionary, sentiment shows a fixed opinion (continuously) that reflects a person's feelings.

Opinions and related concepts such as sentiment, evaluation, behavior, and emotions are studied from sentiment analysis and opinion mining (Liu, 2012). Opinion or opinion is central to almost all human activities and becomes the primary influence of behavior. Perception of reality to evaluate surrounding objects.

The definition of opinion in the journal Opinion Mining and Sentiment Analysis: A Survey in 2012 (Ibrahim & Othman, 2012) are: (1) Views or judgments that are formed about something, are not always based on facts or knowledge, and (2) The beliefs or views of a large number or majority of people on a particular subject. In general, the opinion refers to what people think about something. In other words, opinions are subjective beliefs and are the result of emotions or interpretation of facts.

Sentiment analysis, or opinion mining, is a domain of science that analyzes public opinion, sentiment, evaluation, judgment, attitudes, and emotions towards entities such as products, services, organizations, individuals, problems, events, topics, and their attributes (Liu, 2012). Sentiment analysis focuses on opinions that express or express positive or negative sentiments.

In general, the sentiment analysis studied has three levels: (1) Document-level: classifies whether all opinion documents reveal positive or negative sentiments. The analysis assumes that each document expresses an objective opinion about a single entity (for example, a single product), (2) Sentence level: refers to the expression of the sentence that contains a positive, negative, or neutral opinion; and (3) Entity and aspect level: find sentiments on the entity or its aspects. For example, the phrase "iPhone call quality is good, but battery life is short." There are two evaluation aspects, call quality and battery life, of the iPhone (entity). Sentiment on iPhone call quality is positive, but sentiment on life battery is negative. Call quality and iPhone battery life are the targets of opinion.

Sentiment analysis is a branch of text mining research (Sudalma, 2015). Sentiment analysis exists to deal with the explosion of unstructured textual information. Putten (2002) predicted in his research that the state of the information explosion more made the data mining process difficult because the data was unstructured, and the number was huge.

Naïve Bayes Classifier

The naïve Bayes classifier makes very strong (naive) assumptions about each labeled event class (Han, 2012). The Naive Bayes classifier is used to classify sentiment from the data obtained. Naive Bayes is also used by Hidayatullah (2014) to determine sentiment towards public shops that are conveyed via tweets in Indonesian.

Naïve Bayes is a machine learning method that has a model for shaping probabilities and opportunities. Therefore, Naïve Bayes will calculate the probability of features that present comments based on positive or negative classes.

K-Nearest Neighbor

Nearest-neighbor classification is based on

learning by analogy, namely, comparing testing with similar training data (Han, 2012). K-Nearest Neighbor is a machine learning method that classifies objects based on learning data closest to the object (Dwi, 2014).

Decision Tree

A decision tree is a method that summarizes statements from a testing data label on training data to find problems that lead to making a decision (Kamber., 2006). A decision tree is a prediction model in a miniature structure where internal nodes (not leaves) describe the attributes. Each branch describes the results of the details being tested, and each leaf represents the class.

Testing or measuring accuracy uses:

a. K-Fold Cross-Validation. Cross-validation is used as a method of evaluating classification results. Tests are carried out to predict the error rate. The training data is divided into K random subsets of the same size. One of these random subsets is used as testing data. After that, iterates K times and calculates the error rate for each subset. Based on the results for the error rate of each subset, the average is estimated to get the error rate overall.

b. Confusion Matrix. One of the evaluation methods used for the classification of naïve Bayes is the confusion matrix. Confusion matrices are critical tools in the visualization method used on machine learning which usually contains two or more categories (Horn & Horn, 2010). As much as half or two-thirds of the overall data is used for the training process, while the rest is used for testing purposes (Kantardzic, 2003)

III. RESULT AND DISCUSSION

Data Processing used in Ms. Excel Power query

Data sets are taken from pages in Facebook researchers who have been in post for 6 months and every 1 month do reposting to get the number of data sets in want researchers to analyze the sentiments of public comment.

Comments obtained later in the Crawling text to Ms. Excel Power Query are then carried out in the text preprocessing stage, where the data

totals 100 comments from different people.

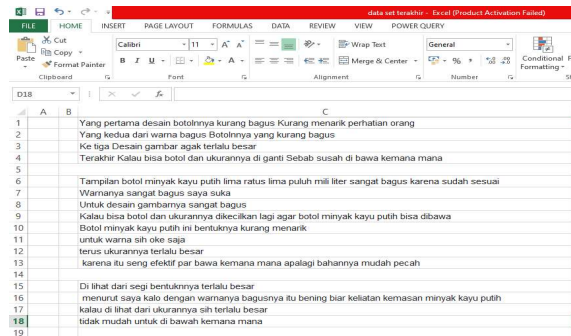This dataset is 100 sets of data that is cleaned as well- data set to use in training and
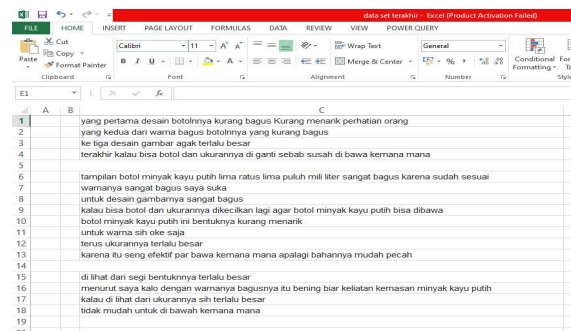

Figure 2. Tokenizing

Figure 3. Case Folding

testing tasks. Cleaning consists of processing the beginning of the text (text preprocessing), Transforming text (text transformation) or (Feature Generation), Selection feature (feature selection), pattern discovery or data mining (pattern discovery), Interpretation / Evaluation.

Text Preprocessing

Text preprocessing here is done using Ms. Excel by paying attention to data that has been separated, such as suggestions in comments that are not related to processing for classification methods for comment sentiment with attributes that have been determined, order the optimal classification stage in its calculations. The preprocessing step in this research includes:

1. Tokenizing. At this stage, delimiters will be numeric characters and symbol characters except for letter characters by listing the required character codes.
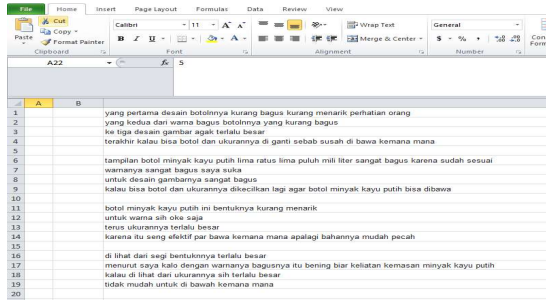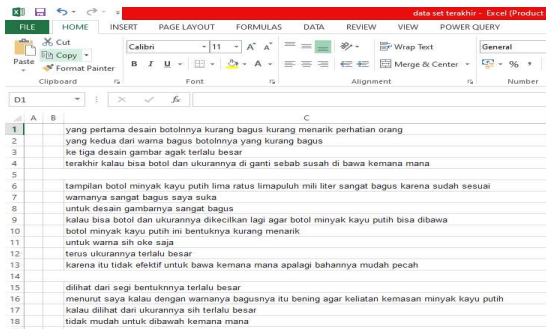
Figure 4. Feature Generation

Figure 5. Example of Filtering

2. Case Folding. At this stage, the data in comments is changed to All lowercase letters, and other emoticon characters will be removed.

3. Text transformation or (Feature Generation). At this stage, the researcher divides and selects each comment corresponding to 4 attributes for the classification method to be processed. The data that is split consists of the shape, size, color, and efficiency data.

Features Selection

The feature selection stage is an advanced

Table 1. Stratified Cross-Validation

Stratified cross-validation		
Correctly Classified Instances	62	64,58%
Incorrectly Classified Instances	34	35,42%

Table 2. Confusion Matrix

Confusion Matrix						
a	b	c	<--		classified as	
62	0	0	|	A	=	Negatif
7	0	0	|	B	=	Positives
27	0	0	|	C	=	Neutral

stage of reducing dimensions in the text transformation process. At the feature selection stage, it is divided into:

1. Filtering. At this stage, will be removed the words that often appear and are general.

2. Stemming. At this stage, the words in the comments will be returned to their primary form by eliminating the affixes to each word.

3. Convert Negation. At this stage, the words that are negated will be converted, such as "no," "nga," "zinc," etc.

Pattern discovery

Data Mining Classifiers and Data Testing

Data Mining Classification and Data Testing For data classification that has been preprocessed above, the researcher use WEKA software to classify data wherein WEKA there are Naïve Bayes, k-NN and Decision Tree.

The histogram image above shows that the blue block is the total number of harmful data in the data set obtained by the number 62. In contrast, the red block is the total number of positive data obtained by the number of as many as 7 data and for the light blue blocks are the total neutral data totaling 27 data. The datasets are trained with Zero-R, Naïve Bayes, and K-NN classifier. The result of each classifier is represented below.

a. Zero-R Classifier

Classifier Zero-R in this data classification is a baseline to see accurate data to compare the three classification methods. This classification is used to assess the success rate of data accuracy for the three methods later tested.

In the Zero-R classification, it can be seen that the data obtained in the classifier output are Run information, Stratified cross-validation, Detailed Accuracy by Class, and Confusion Matrix. From this result, it is known that the traditional value that will be the baseline for the 3 methods used for Correctly Classified Instances is 64.58% and for the Confusion Matrix, namely 62 data instances for negative classes, 7 instances for types positive, and 27 instances for neutral classes.

b. Naïve Bayes Classifier

For the test results using this method, it can be seen that the output classifier is Run information, Classifier model (complete training set), Time taken to build the model: 0.01 seconds,

Table 3. Stratified cross-validation

Stratified cross-validation		
Correctly Classified Instances	78	81,25%
Incorrectly Classified Instances	18	18,75%

Table 4. Confusion Matrix

Confusion Matrix						
a	b	c	<--		classified as	
58	1	3		a	=	Negatif
0	0	7		b	=	Positive
6	1	20		c	=	Neutral

Stratified cross-validation, Detailed Accuracy By Class, and Confusion Matrix. From that, it is seen that Correctly Classified Instances exceed the baseline standard by 81.25%. The Confusion Matrix is classified as 62 instances for negative, seven examples for positive, and 27 instances for neutral.

c. k-NN Classifier

In testing using the k-NN method, it still uses the same data set as the previous method with the Output Classifier, including Run information, Stratified cross-validation, Detailed Accuracy by Class, and Confusion Matrix. The results show that Correctly Classified Instances exceed the baseline

Table 5. Stratified cross-validation

Stratified cross-validation		
Correctly Classified Instances	62	64,58%
Incorrectly Classified Instances	34	35,42%

Table 6. Confusion Matrix

Confusion Matrix						
a	b	c	<--		classified as	
62	0	0		a	=	Negatif
7	0	0		b	=	Positive
27	0	0		c	=	Neutral

standard by 83.33%. The Confusion Matrix is classified as 62 instances for negative, 7 instances for positive, 27 instances for neutral.

d. Decision Tree.

Processing this data, it is known that the Output Classifier that is obtained is Run information, Classifier model (complete training set), Time taken to build the model: 0.26 seconds, Stratified cross-validation, Detailed Accuracy By Class, and Confusion Matrix. from the results seen for Correctly Classified Instances are at the baseline standard with a total of 64.58% and a total classified Confusion Matrix of 62 instances for negative, 7 instances for positive and 27 instances for neutral.

Table 7. Stratified cross-validation

Stratified cross-validation		
Correctly Classified Instances	80	83,33%
Incorrectly Classified Instances	16	16,67%

Table 8. Confusion Matrix

Confusion Matrix						
a	b	C	<--		classified as	
57	0	5		a	=	Negatif
1	3	3		b	=	Positive
6	1	20		c	=	Neutral

Interpretation of the classifier result

In the calculation using the 3 methods above, it can be seen that the value obtained in the 3 methods tested for cross-validation is Confusion Matrix. So the K-NN method is the most accurate compared to the other 2 methods, namely Naïve Bayes and decision trees with greater Correctly Classified Instances are 83.33% and Incorrectly Classified Instances smaller amounting to 16.67%, for the Confusion Matrix, it

Table 9. Stratified cross-validation

Stratified cross-validation		
Correctly Classified Instances	78	81,25%
Incorrectly Classified Instances	18	18,75%

Table 10. Confusion Matrix

Confusion Matrix						
a	b	c	<--		classified as	
55	2	5		a	=	Negatif
1	4	2		b	=	Positive
4	4	19		c	=	Neutral

is also seen that this method can obtain good accuracy with a value for a harmful class of 57 classified instances. Correctly while 5 instances data are classified in the neutral class, for the positive class it is known that the value obtained in the matrix is 3 instances out of 7data is classified 27 instances incorrectly in the negative class while the correct one is 0. Thus, it can be interpreted that the K-NN method was better at analyzing public comment sentiment. In this research, to test the accuracy of this method, the researcher conducted a test on training data that had used well-confirmed methods, namely K-NN.

Testing and Training

Before testing with testing data, the researcher will perform Class Balancing so that the number of datasets for each balanced attribute. When trying to use testing data, it can make more accurate results. Here the researcher uses a (cost-sensitive classifier) to put the dataset into balance.

55 instances listed in the positive class while 4 other data were sharing on as much negative class 1 and 3 cases of data instances that are in the class neutral, and for style In neutral, it is known that the results of models that are correctly classified are 20 instances of 27 known data. In comparison, the other 7 data are divided into negative and positive classes were for negative 6 instances data and 1 instance data for positive class. The result of prediction provides Correctly Classified Instances of 81.25% and Incorrectly Classified Instances of 18.75%. In the Confusion Matrix for this method, it is known that the Negative Class value in the matrix is 58 instances that are classified correctly. The other 4 are divided into positive classes 1 data instance and 3 neutral samples of data.

For the decision tree method with the results of Correctly Classified Instances 64.58%, Incorrectly Classified Instances 35.42%, and Confusion Matrix in this method. It is known that the class value for negative is 62 data from 62 data instances; for the positive class, it is known that this matrix is 0 classified correctly. In contrast, 7 samples are classified incorrectly in the hostile class and for the neutral class.

Results of Testing and Training

The test results for testing data show that the amount of data used is ten testing data with unknown labels or classes. In this trial, the researcher using training data that has previously been processed using the best classification method of the 3 methods tested, namely the K-NN method. Before conducting the trial, the researcher balances the training data so that the prediction results on the testing data being tested are more accurate. Therefore the researcher uses a cost-sensitive classifier to make the training data balance. After that, the testing data is run using the K method. -NN is the preferred method because it has a better predictive accuracy value and obtains results.

Table 11. Results of Predicted Class

Predicted CLASS	Prediction Margin	
	Statistic	value
	Minimum	-0,995
Netral = 5 komentar	Maximum	0,993
Negatif = 3 komentar	Maen	-0,333
Positf = 2 komentar	StdDev	0,899

Discussion

From data processing using the above classification analysis, the researcher obtained different presentation accuracy values. Where 83% for Correctly Classified Instances and 16% for Incorrectly Classified Instances using the K-NN method. Meanwhile, 81% for Correctly Classified Instances and 18% for Incorrectly Classified Instances using the Naïve Bayes Method. And 64% for Correctly Classified Instances and 34% for Incorrectly Classified Instances using the decision tree method. From this, it is known that the K-NN classification Data Mining Technique has a very accurate level of prediction. This classification can better analyze instances of data for each class attribute with an accuracy rate of 81%. To try the success rate of predicting the K-NN classification, the researcher tested the obtained training data.

In the trial application of this classification, the researcher obtained a prediction result of 81.25% for Correctly Classified Instances and 18.75% Incorrectly Classified Instances, with class prediction results for optimistic two comments,

negative three comments, and neutral five comments. It can be interpreted that the presentation of predictions using the K-NN classification gets excellent results. And from this research, it can be concluded that the features of 550 ml eucalyptus oil packaging such as shape, color, size, efficiency can affect changes to the redesign of the 550 ml eucalyptus oil bottle packaging from the previous design. Where based on comments from social media users for packaging features such as forms that have a negative class value of 47 comments, while positive class is 36 comments, and for neutral class, there are 13 comments. And because the negative is bigger than the other classes. So, it needs to be made as attractive as possible to increase the attractiveness of this packaging by paying attention to several suggestions, namely. Shapes can be made simpler because consumers prefer simple shapes rather than complex shapes. The two shapes regular on the packaging are more concerned with because the common shapes have more appeal. The three forms of packaging can be made smaller than the initial size without reducing the volume of eucalyptus oil. The four forms of packaging can be made more comfortable when used or carried around. The five shapes of eucalyptus oil bottles are preferably convex rather than concave shapes. The packaging size attribute can affect the packaging redesign with a negative class value of 82 comments, a positive class of 11 comments, while the neutral class is only three comments. So from there, the size still has to be adjusted according to the conditions of use, such as if it can be carried as souvenirs, it can be made smaller as in the third part of the form attribute, namely reducing the size without reducing the volume of eucalyptus oil content 550 ml, because for efficiency it is also one attributes that affect the redesign of the packaging that really must be considered 3 things such as, first the addition of a handle or grip to make it easier when used or carried around, secondly, this bottle packaging can be made using alternatives other than glass to reduce the burden on the packaging without losing the properties of eucalyptus oil, and the last one is that the head of the bottle is curved for

easy use. This is because of the attribute efficiency with a negative class value of 94 comments and a positive class with only 2 comments. It is necessary to make improvements to make it easier and help consumers consume this eucalyptus oil product. And for the color attribute on the packaging is very influential in changing the packaging design. Where the packaging color is requested to be more precise or more transparent so that consumers can see the color directly from eucalyptus oil, this proposal is because the color attribute has a negative class value of 46 comments, and a positive class of 32 statements, while for the neutral class there were 18 comments. And because the negative value is more significant, this proposal is made to fulfill consumer desires.

IV. CONCLUSION

Based on the data processing results above, the researcher obtained a presentation accuracy value of 83% for Correctly Classified Instances and 16% for Incorrectly Classified Instances using the K-NN method. Meanwhile, 81% for Correctly Classified Instances and 18% for Incorrectly Classified Instances using the Naïve Bayes Method. And 64% for Correctly Classified Instances and 34% for Incorrectly Classified Instances using the decision tree method. From this, it is known that the K-NN classification Data Mining Technique has a very accurate level of prediction. This classification can better analyze instances of data for each class attribute with an accuracy rate of 81%. To try the success rate of predicting the K-NN classification, the researcher tested the obtained training data. In the trial application of this classification, the researcher got a prediction result of 81.25% for Correctly Classified Instances and 18.75% Incorrectly Classified Instances, with class prediction results for positive 2 comments, negative 3 comments, and neutral 5 comments.
The 550 ml eucalyptus oil packaging features such as Shape, Color, Size, Efficiency can affect changes to the redesign of the 550 ml eucalyptus oil bottle packaging from the previous design. Where based on comments from social media

users for packaging features such as forms that have a negative class value of 47 comments, while positive class is 36 comments, and for neutral class, there are 13 comments. The packaging size attribute can affect the packaging redesign with a negative class value of 82 comments, a positive class of 11 comments. In comparison, the neutral class is only 3 comments for the efficiency attribute with a negative class value of 94 comments and a positive class with only 2 comments. And for the color attribute on the packaging is very influential in changing the packaging design. The color attribute has a negative class value of 46 comments, and a positive class of 32 comments, while for the neutral class, there are 18 comments. And because the results of the comparison of comments have negative values greater than positive and neutral values, this attribute significantly affects the packaging redesign.

REFERENCES

Andini, S. (2013). Klasifikasi Dokument Teks Menggunakan Algoritma Naive Bayes Dengan Bahasa Pemograman Java.

Blanco, E., Moldovan, D. (2011). Some issues on detecting negation from the text. Proceedings in Twenty-Fourth International FLAIRS Conference.

Case, D. A., Dairy, P., Mutsikiwa, M., Marumbwa, J. (2017). The Impact of Aesthetics Package Design Elements on Consumer Purchase Decisions: A Case of Locally Produced Dairy Products.

Cenadi, C. S. (2000). Peranan desain kemasan dalam dunia pemasaran. *Nirmana, 2* (2).

Collie, R., Singh, A. (2016). Power Pivot and Power Bi: The Excel User's Guide to Dax, Power Query, Power Bi & Power Pivot in Excel 2010-2016. Holy Macro.

Putranti, N. D., Winarko, E. (2014). Analisis sentimen twitter untuk teks berbahasa Indonesia dengan maximum entropy dan support vector machine. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 8(1), 91-100..

Dwiningsih, N. (2002). Desain Produk Dan Manajemen Kualitas.

Frank, E., Hall, M. A., Pal, C. J. (2011). Algorithms : The basic methods.

Gusriani, S., Wardhani, K. D. K., Zul, M. I. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes

(Studi Kasus: Facebook Page BerryBenka). In 4th Appl. Bus. Eng. Conf (Vol. 1, No. 1, pp. 1-7).

Han, J. (2012). Data Mining Concepts and Techniques.

Hapsari, Y., Hidayattullah, M. F., Dairoh, D., Khambali, M. (2018). Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivry). Jurnal Informatika: Jurnal Pengembangan IT, 3(2), 233-236.Hidayatullah,

Hidayatullah, A. F., Azhari, S. N. (2014). Analisis Sentimen dan klasifikasi kategori terhadap tokoh publik pada data Twitter menggunakan Naive Bayes Classifier (Doctoral dissertation, [Yogyakarta]: Universitas Gadjah Mada).

Horn, C., (2010). Analysis and Classification of Twitter messages Analyse und Kategorisierung von Twitter Nachrichten. (April).

Sadegh, M., Ibrahim, R., Othman, Z. A. (2012). Opinion mining and sentiment analysis: A survey. International Journal of Computers & Technology, 2(3), 171-178.

Ireland, R., Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. CIRP Journal of Manufacturing Science and Technology, 23, 128-144.

Jackson, J. (2002). Data mining; a conceptual overview. Communications of the Association for Information Systems, 8(1), 19.

Jebaseeli, A. N., Kirubakaran, E. (2012). A survey on sentiment analysis of (product) reviews. International Journal of Computer Applications, 47(11).

Han, J., Kamber, M. (2006). Classification and prediction. Data mining: Concepts and Techniques, 347-350.

Kang, W. G., Ko, E. S., Lee, H. R., Kim, J. N. (2018). A Study of the Consumer Major Perception of Packaging Using Big Data Analysis-Focusing on Text Mining and Semantic Network Analysis. Journal of the Korea Convergence Society, 9(4), 15-22.

Kantardzic, M. (2003). Data Mining: Concepts, Models, Methods, and Algorithms. Technometrics, 45(3), 277.

Kaukab, M. E., Setyanto, R. P., Soedirman, U. J. (2020). Pengembangan Kesesuaian Produk Ekspor (Export Product Fit Development).

Lee, R., Lyu, J. (2020). Sentiment Analysis of Product Reviews Based on JST Model. In Proceedings of the 4th International Conference on Computer Science and Application Engineering (pp. 1-7).

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Shivaprasad, T. K., Shetty, J. (2017, March). Sentiment analysis of product reviews: a review. In 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 298-301). IEEE.

Smith, H., Idrus, S., Ambon, B. I., Ambon, J. K. C. B. M. (2016). Estimasi Produksi Minyak Kayu Putih Pada Industri Kecil Penyulingan Di Maluku. Majalah BIAM, 12(1), 8-13.

Turban, E., Aronson, J. E., Liang, T. P. (2005). Decision Support Systems and Intelligent System,(Sistem Pendukung Keputusan dan Sistem Cerdas) Ed. 7. Jld. 2.

Manalu, B. U. (2014). Analisis Sentimen Pada Twitter Menggunakan Text Mining. Universitas Sumatera Utara Medan.

Zhang, L., Chu, X., Xue, D. (2019). Identification of the to-be-improved product features based on online reviews for product redesign. International Journal of Production Research, 57(8), 2464-2479.