Sentiment Analysis of Madurese Batik Twitter Opinion for Optimize Product Marketing

Fika Hastarita Rachman [1]   Faikul Umam[2]   Imamah[3]

[1,3] Departement of Informatics, University of Trunojoyo Madura, Indonesia
[2] Departement of Mekatronika, University of Trunojoyo Madura, Indonesia
* Corresponding author's Email: hastarita.fika@gmail.com

Abstract: Madurese batik is a typical Indonesian product that has unique motif and colors. To improve the marketing of this product, a customer satisfaction analysis is needed. Social media is a place to write user opinions on the quality of Madurese batik products. This study uses sentiment analysis to optimize the marketing of Madurese products. From the tests conducted, 925 people showed positive sentiment and 25 indicates negative sentiment. The classification results with naïve Bayes also show an accuracy of 94%. The optimal marketing location area is Jakarta and its surrounding areas, because of the dominance of users in these areas to respond positively.

Keywords: sentiment analysis, Madurese batik, product marketing, naïve bayes

1. Introduction

Madura batik is an original Indonesian product which has unique and distinctive motif. The area of its marketing has reached the international level. Good marketing techniques are needed for the sustainable production of certain products [1]. The marketing technique can be seen by adding value to a product or by analyzing areas that are interested in the product. Theoretically, the added value of each product can be seen from the user satisfaction score [2]. The value of user satisfaction can be seen from the public opinion of the Madura batik product users.

Social media is a media that can be used as a means of expressing people's opinions on certain events or products. In addition, social media analysis data is also used to forecast online product sales [3]. In this study, social media is used to find out written opinions of the public who use Madura batik products.

From the social media data, we will analyze how the response of the customer to Madurese batik, how is the overall customer sentiment towards Madura batik? How to map Madura batik marketing based on reviews about Madura batik on social media.

The objectives to be achieved in this study are to obtain information about things that can be used as a consideration for business actors and the government in making policies, by knowing the response of customers to Madurese batik, knowing the overall customer sentiment towards Madura batik. , knowing the market potential of Madurese batik based on the highest frequency of positive reviewers, and knowing the quality improvements that should be made based on negative reviews from customers of Madura batik.

2. Sentiment Analysis

Sentiment analysis is a branch of research in the text mining domain which began to be widely carried out in 2013. Sentiment analysis is computational research of opinions, sentiments and emotions that are expressed textually. Sentiment analysis aims to extract attributes and components from what has been commented on in each document and to determine whether the comment is positive or negative [4].

Sentiment analysis or opinion mining is the process of understanding, extracting and processing textual data automatically to get sentiment information to get the information contained in an opinion sentence. Sentiment analysis is carried out to

see the opinion or tendency of an opinion on a problem or object by someone, whether they tend to have a negative or positive view or opinion [5]. There are several types of sentiment analysis that can be used, including:

1. Fine-Grained Sentiment Analysis

Fine-Grained Sentiment Analysis is an analysis process carried out at the sentence level of a document. There are 2 types of orientation, namely positive and negative, information in the form of papers, journals, and others that contain objective textual information, namely information that describes the author's feelings or opinions on what he writes. Objective information tends to be neutral, whereas information such as opinions, comments, or opinions each has a meaning expressed by the author, the meaning is positive or negative [6].

2. Corse-Grained Sentiment Analysis

Coarse-Grained Sentiment Analysis is an analysis process carried out at the document level. This analysis tries to classify the orientation of a document as a whole. There are 3 types of orientation, namely positive, negative and neutral and some make the value of this orientation continuous / not discrete [6].

3. Naïve Bayes

The Naive Bayes method is a simple probalistic classification based on the application of the Bayes theorem (Bayesian statistics) with the assumption of strong independence. A more descriptive term for the underlying probability model is "independent feature model" [7]. This method requires only a small amount of training data to estimate the parameters (means and variants of the variables) for classification. Since independent variables are assumed, only the variance of each variable for each class requirement will be determined instead of the entire covariance matrix [7].

Naive Bayes is a classification method that uses probability calculations. The basic concept used in the Naive Bayes Classifier is the Bayes Theorem which was first stated by Thomas Bayes [8]. The probability value used is stated simply as follows:

$$P\ (C\mid D) = \frac{P(D|C)P(C)}{P(D)} \tag{1}$$

In conducting text classification the Naive Bayes method takes two stages, namely the training stage and the classification stage [9], namely:

$$P\ (Xk\mid Y) = \frac{P(Y|Xk)}{\sum_{i}\ P(Y|Xi)} \tag{2}$$

note :

$P\ (X_k\mid Y)$: posterior (probability X_k in Y)

$P\ (Y|X_k)$: prior (probability Y in X_k)

$\sum_{i} P(Y|X_i)$: all probability Y in X_i

$$P\ (v1|C = c) = \frac{CountTerms(v1,docs(c))}{AllTerms(docs(c))} \tag{3}$$

note :

v1 : specific words in tweets

CountTerms(v1, docs(c)) : number of words labeled

AllTerms(docs(c)) : the sum of all labeled words in the dataset

$$P\ (v1|C = c) = \frac{CountTerms(v1,docs(c))+1}{AllTerms(docs(c))+|V|} \tag{4}$$

In the above equation, Laplace (add-one) smooting is done to avoid the zero probability value.

4. Methodology

Figure 1 explains the data processing flow using the method, in this process the data that has been taken from Twitter social media will be processed first in the preprocessing stage, at this stage the data will be carried out by the process of case folding, tokenizing, stemming and filtering. After the prepocessing process, the data went through a labeling process using a lexicon based. At this stage the data will be divided into positive and negative data. The next process, the data is divided into two, namely training data and test data, the next stage is the classification process using the Naive Bayes method. Training data and test data will go through a classification process and will get results in the form of a percentage of negative and positive sentiments from the research.
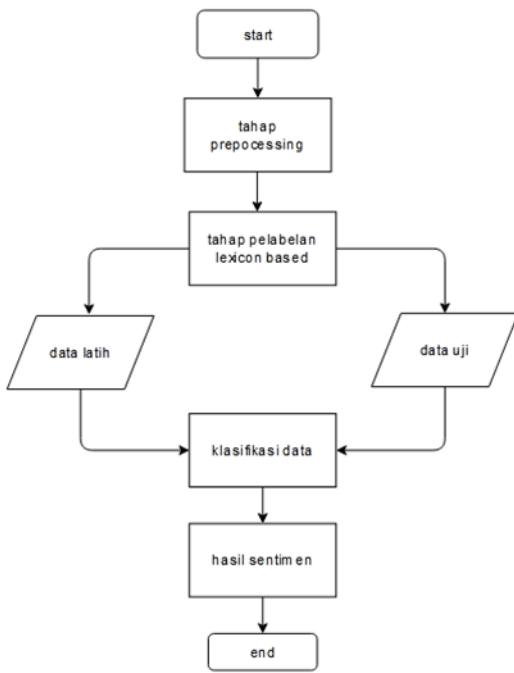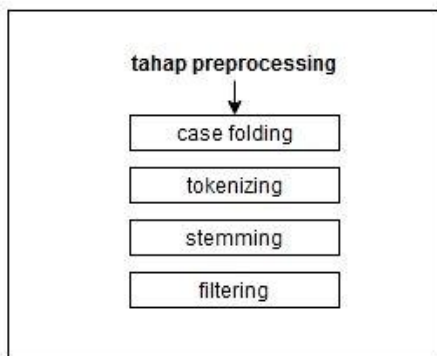
Figure 1. Proposed method

Figure 2. Step of preprocessing data

Testing the accuracy of the system using the Confusion Matrix, to determine the accuracy of the data labeled with the Lexicon Based and the classification of data by the system using the Naive Bayes method. This test will look for the level of system accuracy, system error, precision, recall / sensitivity, and specificity.

$$\text{accuracy} = \frac{(TNegative+TNetral+TPositive)}{count\ of\ data\ testing} \quad (5)$$

$$\text{system error} = 1 - accuracy \quad (6)$$

$$\text{positive precision} = \frac{(TP)}{(TP+FP)} \quad (7)$$

$$\text{negative precision} = \frac{(TN)}{(TN+FN)} \quad (8)$$

$$\text{recall} = \frac{(TP)}{(TP+FN)} \quad (9)$$

$$\text{spesificity} = \frac{(TN)}{(TN+FN)} \quad (10)$$

note :
TP = True Positif
TN = True Negatif
FP = False Positif
FN = False Negatif

5. Result and Discussion

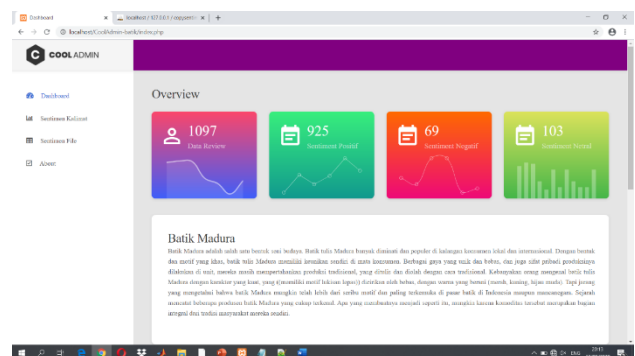The results of the implementation of the Batik Sentiment system development in this study are shown in Figure 3.

Figure 3. Front end of aplication sentiment analysis

It can be seen that the data for reviewing comments on Madura batik drawn from Twitter is 1097 data. From this data, 925 data were detected to have positive sentiment, 69 data were detected as negative sentiment and 103 data were neutral sentiment.
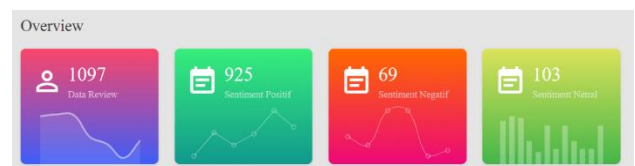

Figure 4. Sentiment analysis results info

Figure 4 shows the amount of data categorized as positive, negative and neutral sentiment. In addition, information about the distribution of Madura batik users based on location can also be seen in the application.
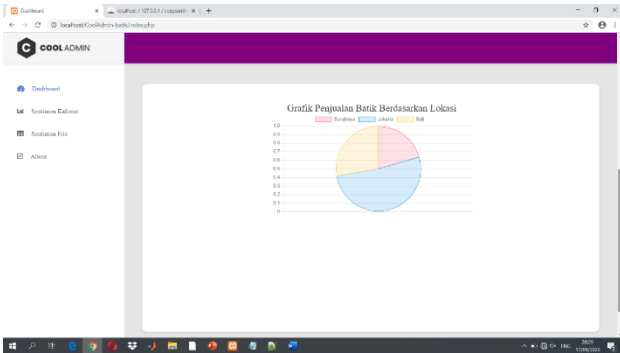
Figure 5. Graph of Number of Batik Madura users

From the dashboard, it can be seen that the highest number of Madurese batik users is in the Jakarta and surrounding areas, followed by users in Bali and surrounding areas as well as Surabaya and its surroundings.

Sentiment testing on sentences entered in the system is carried out to test whether the system is running correctly or not.
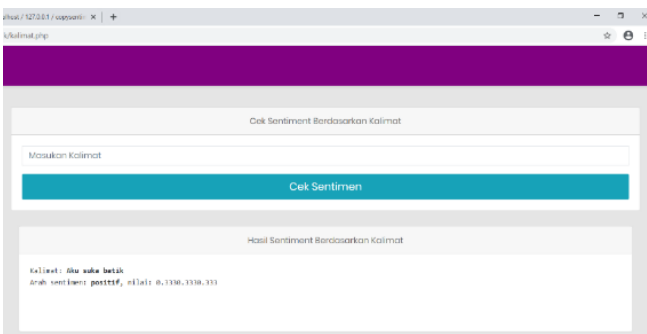

Figure 6. Sentence sentiment testing

Apart from sentence testing, sentiment testing on several sentences contained in the xls file can also be done. As shown in Fig7.
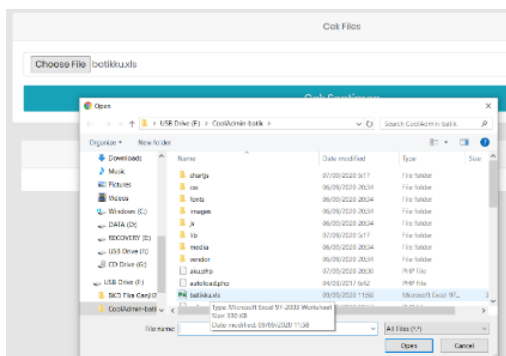

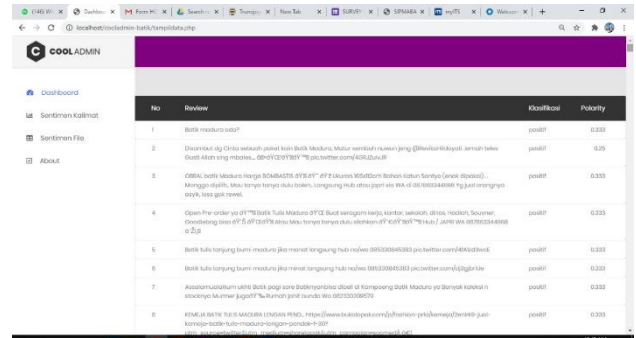
Figure 7. Sentiment testing from file xls

Fig.8 Result of sentiment analysis from input file xls

Figure 9. Accuracy

The accuracy obtained by using the Naïve Bayes classification obtained an accuracy of 0.94. This shows that the use of the naïve Bayes method can be used in the sentiment analysis process.

6. Conclusion

Based on the analysis of test results with an accuracy of 94%, it can be concluded that the use of sentiment analysis to optimize marketing Madura batik is very effective. The results showed that there are certain locations that have the highest levels of positive and negative sentiment, so that they can be used as a marketing reference, namely Jakarta and its surroundings.

If the highest negative review is obtained, it can be seen based on comments or public opinion regarding the cause of the negative review. This can be very useful for improving Madura batik products based on the negative sentiment of the community in certain areas. Meanwhile, if the highest positive reviews are obtained, then the promotion must be intensified in that area, because it is clear that the people's buying interest in that area is high, based on the positive reviews that have been generated.

References

[1]   E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "Managing Marketing

Decision-Making with Sentiment Analysis : An Evaluation of the Main Product Features Using Text Data Mining," pp. 1–19, 2019.

[2]   P. Kualitas, P. Dan, J. Pengiriman, T. Kepuasan, O. Di, and Z. Indonesia, "THE EFFECT OF PRODUCT QUALITY AND DELIVERY SERVICE ON ONLINE- CUSTOMER SATISFACTION IN ZALORA INDONESIA," vol. 4, no. 1, pp. 1189–1199.

[3]   Z. Fan, "Product sales forecasting using macroeconomic indicators and online reviews : a method combining prospect theory and sentiment analysis Product sales forecasting using macroeconomic indicators and online reviews : a method combining prospect theory and sentiment analysis," *Soft Computing*, no. May, 2020.

[4]   H. C. Soong, N. B. A. Jalil, R. Kumar Ayyasamy, and R. Akbar, "The essential of sentiment analysis and opinion mining in social media : IIntroduction and survey of the recent approaches and techniques," *ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, pp. 272–277, 2019.

[5]   I. F. Rozi, S. H. Pramono, and E. A. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, vol. 6, no. 1, pp. 37–43, 2012.

[6]   C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse- and Fine-Grained Sentiment Analysis of Social Media Text," vol. 30, no. 1, pp. 22–30, 2011.

[7]   N. Bayes, "Naive Bayes classifier," pp. 1–9, 2006.

[8]   J. Aldrich, "R. A. Fisher on Bayes and Bayes' Theorem," *Bayesian Analysis*, vol. 3, no. 1, pp. 161–170, 2008.

[9]   F. Nurhuda and S. W. Sihwi, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," vol. 2, no. 2, 2014.