Sentiment analysis of tokopedia application review to service product recommender system using neural collaborative filtering for marketplace in Indonesia

Restu Meifitrarah1*, Irfan Darmawan1, Oktariani Nurul Pratiwi1

1 Information System, Faculty of Industrial Engineering Telkom University, Bandung
*Email: restumeifitrarah@student.telkomunivesity.ac.id

Abstract. Tokopedia is one of the leading e-commerce companies in Indonesia and ranks second in the top 10 e-commerce in Indonesia in 2018 based on Statista data. Large companies like Tokopedia need to find out what users think of the products or services offered. User opinions on the Tokopedia application can be seen in the review column on the Google Play Store, but processing a review is not easy. To overcome this we need a sentiment analysis method or technique. Sentiment analysis is performed using the Naive Bayes algorithm. The results of positive sentiments obtained are used as a reference to maintain service quality and the results of negative sentiments can be used as an evaluation material in improving Tokopedia services and applications.

1. Introduction

1.1 Preliminary
Tokopedia is the most popular technology company in Indonesia with a mission to achieve economic equality digitally. Tokopedia allows individuals and business owners to open and manage online stores easily and for free. Tokopedia ranks second in the top 10 e-commerce sites in Indonesia still have the opportunity to occupy the first position and can continue to maintain its popularity by increasing the quality of service. To improve the quality of its services, Tokopedia can take advantage of user reviews on the Tokopedia application found on Google Playstore. Reviews contain lots of suggestions, praise and even user complaints. These reviews can be used as input for the Tokopedia service. To find out the tendency of comments and information contained in a review is not easy because it requires a method that can process the data quickly and automatically to categorize the reviews both positive and negative. The method that can be used to determine the tendency of comments in the review is to do sentiment analysis. To be able to do sentiment analysis of Tokopedia application user reviews, appropriate steps and algorithms are required.

To do sentiment analysis of Tokopedia application user reviews, appropriate methods and algorithms are needed. The algorithms used to do sentiment analysis are very much, which are often used include the Naive Bayes Classifier, Support Vector Machine, Decision Tree. Each algorithm has a different level of accuracy in each case. In a previous study by [1] about the Personality Classification of Twitter users found different levels of accuracy, are Naive Bayes 63%, KNN 60%, and SVM 61%. In research [2] about product analysis sentiments using machine learning by comparing several algorithms, namely Naive Bayes, SVM, and MaxEnt, the results were 81.33%, 78.67% and MaxEnt classifiers 72%. The advantage of Naive Bayes compared to...
other algorithms is that simple and intuitive Naive Bayes can produce good performance with little
testing data [3]. For maximum results, a TF-IDF calculation is required. Based on considerations
and observations from previous studies the author decides to do a sentiment analysis using the
Naive Bayes algorithm.
1.2 Theoretical basis
1.2.1 Sentiment analysis
Sentiment Analysis or commonly called opinion mining is one branch of Text Mining research.
Sentiment Analysis can be distinguished based on the data source, several levels that are often used
in Sentiment Analysis research are Sentiment Analysis at the document level and Sentiment
Analysis at the sentence level. Based on the level of data sources Sentiment Analysis is divided
into 2 major groups namely Coarse grained Sentiment Analysis and Fine Grained Sentiment
Analysis [4].

1.2.2 TF-IDF
The TF-IDF method is a method for calculating the weight of each word that is most commonly
used in information retrieval. This method will calculate the value of Term Frequency (TF) and
Inverse Document Frequency (IDF) on each word (token) in each document in the corpus [5].

\[
\text{Tf-idf} (t,d) = \text{tf}(t,d) \times \text{idf}(t)
\]

1.2.3 Naive bayes algorithm
Naive Bayes Classifier is a simple probability classification method in which to apply the Bayes
theorem with high independent assumptions. Referring to the basic concept of Naive Bayes
classifier namely Bayes' Theorem which was first stated by Thomas Bayes [6]

\[
P(C|D) = \frac{P(D|C) \times P(C)}{P(D)}
\]

Annotation :
D = Data training
C = Classification
\(P(C|D)\) = Posterior Probability
\(P(C)\) = Prior Probability
\(P(D|C)\) = Probability of Data Occurrence
\(P(D)\) = Opportunity classes in the data classification

1.2.4 Accuracy
Evaluation of accuracy needs to be done to ensure the accuracy of the results obtained. Evaluation
can be measured using Confusion Matrix.

1.2.5 K-fold cross validation
Cross validation is one method that can be used to evaluate the performance of an algorithm. The
dataset will be divided by the value of K, and will be tested alternately until all parts are met. In
determining the K value, 10-fold validation is widely used because it is proven to produce a more
good performance.

1.2.6 The state of the art
Here is a comparison table for algorithms in various studies that have been done before:
Table 1. The accuracy of various algorithms use in previous researches

| No  | Journal                                                                 | Author                      | Accuracy | NB     | SVM    | DT  | KNN     |
|-----|-------------------------------------------------------------------------|-----------------------------|----------|--------|--------|-----|---------|
| 1.  | Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM | Bayu Yudha Pratama, Riyanarto Sarno |          | 63 %   | 61 %   | -   | 60 %    |
| 2.  | Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB, SVM and MaxEnt | Monali Bordoloi, S.K. Biswas |          | 81.33% | 78.67% | -   |         |
| 3.  | Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers | Nur Azizah Vidyaa, Mohamad Ivan Fanany, Indra Budi |          | 78.90% | 82.40% | 72.90% | -       |

In the first study by [1] Naive Bayes had the best 63% accuracy than SVM 61% and KNN 60%, this happened because SVM and KNN were considered to have difficulty separating word classes because the existing datasets were less accurate.

In the second study by [2] regarding mobile product review with several attempts using different algorithms also produced the highest level of accuracy by the Naive Bayes algorithm 81.33% compared to SVM and MaxEnt, but the three algorithms are still classified as algorithms that have the best level of accuracy.

For the third study [7] on the reputation of cellular phone providers giving different results from the two previous studies, here SVM 82.40% has a better accuracy rate compared to Naive Bayes 78.90% and Decision Tree 72.90%. Nevertheless, Naive Bayes still has a better level of accuracy compared to the Decision Tree.

2. Data Retrieval
2.1 Dataset
Tokopedia application review data is obtained by doing web scraping on the google play store website. The reviews taken to serve as data in this study are the most recent reviews taken on February 10, 2020, totaling 6000 reviews.

2.2 Labeling
The data that has been obtained is then labeled manually. The given labels are 1, 2 and 3. Label 1 represents positive sentiment, label 2 represents negative sentiment, and label 3 represents neutral sentiment.

2.3 Text Preprocessing
Text preprocessing is the stage for preparing text data in a more structured form. The stages used in this research are punctuation removal, case folding, tokenizing, stopword removal, and stemming. Step of text preprocessing can be seen in this figure:
Based on the figure, it can be seen that the stages of the text preprocessing used in this study consisted of 5 stages, are case folding, punctuation removal, tokenize, stopword removal, and stemming.

2.3.1 Case Folding
Case folding is a process to change all letters to lower case.

| Table 2. Example of case folding |
|----------------------------------|
| Before | After |
| I LOVE shopping on this app, good items, and fast shipping | i love shopping on this app, good items, and fast shipping |

2.3.2 Punctuation Removal
Punctuation removal is the process of removing symbols or punctuation contained in text.

| Table 3. Example of punctuation removal |
|----------------------------------------|
| Before | After |
| i love shopping on this app, good items, and fast shipping | i love shopping on this app good items and fast shipping |

2.3.3 Tokenizing
Tokenizing is the stage to separate input strings based on words that compose them using spaces.

| Table 4. Example of tokenizing |
|-------------------------------|
| Before | After |
| i love shopping on this app good items and fast shipping | i love shopping on this app good Items And Fast shipping |

2.3.4 Stopword Removal
Stopword removal is the process of removing unnecessary words by checking the stoplist, the stoplist used is Indonesian which is taken from the NLTK library.
### Table 5. Example of stopword removal

| Before       | After      |
|--------------|------------|
| i love       | love       |
| shopping on  | shopping   |
| this app     | fast       |
| good items   | items      |
| and Shipping | fast shipping |

### 2.3.5 Stemming
The stemming stage is done to return the words to their basic form words. This process is done by removing the prefix, suffix, or both. The stemming process is assisted by using literary python for Indonesian-language text.

### Table 6. Example of stemming

| Before | After |
|--------|-------|
| love   | love  |
| shopping | shop |
| app    | app   |
| good   | good  |
| items  | item  |
| fast   | fast  |
| shipping | ship |

### 2.4 Text Processing
After all data review through the preprocessing stage will be classified using the Naive Bayes algorithm. The results obtained from this classification are calculated for accuracy, and their performance can be seen through the confusion matrix to ensure the validity of the results, the validation stage is performed using k-fold cross validation with a value of K= 10.

### 3. Result Analysis

#### 3.1 Dataset
In this study the dataset containing user reviews will be classified into three labels that are done manually, 1 for positive labels, 2 for negative labels, and 3 for labels with neutral comments.

### Table 7. Number of comment in dataset

| Label | Number of Comment |
|-------|-------------------|
| 1     | 2.670             |
| 2     | 2.123             |
| 3     | 1.207             |
| Total | 6.000             |

After text preprocessing, the amount of comment data is reduced, the change in the amount of comment data after the preprocessing stage can be seen in the following table:

### Table 8. Number of comment after text preprocessing

| Label | Number of Comment |
|-------|-------------------|
| 1     | 2.552             |
| 2     | 2.121             |
| 3     | 1.187             |
| Total | 5.860             |
To process the data using an algorithm, the data is divided into training data and test data. The comparison ratio for training data and testing data is 80:20, which means 80% of the total data will be used as training data and 20% of the total data will be used as testing data.

### 3.2 Analysis Sentiment

To do analysis sentiments, the data which has been processed through the preprocessing stage, is divided into training data and testing.

| No | Label | Number of Comment |
|----|-------|-------------------|
| 1. | Positif | 527 |
| 2. | Negatif | 420 |
| 3. | Netral | 225 |

After it did a selection of features using the tf-idf method then was processed using a naive bayes algorithm. Selection of features by using tf-idf can do a good job, the way it works is if the weight is gained, the higher the similarity rate of the document.

**Figure 2.** Amount of comment after analysis

Analysed comments amounted to 1172 comments made by 601 positive comments, 533 negative comments, and 38 neutral comments. From results obtained by user reviews tend to show positive comments, with an accuracy rate of 76%.

It can be seen that there is a considerable difference between the data of the results of the manual label and the data of the prediction results of the system. The biggest difference is in data with a neutral sentiment. This happens because of the misinterpretation of sentences, for example in neutral sentences there are positive words that make the system tend to read the data is positive.

### 3.3 Accuracy

In this study the test calculates accuracy using a confusion matrix

| Actual | Predict  |
|--------|----------|
| Positive | Negative | Netral |
| Positive | 489  | 36  | 2 |
| Negative | 32  | 384 | 4 |
| Netral  | 80  | 113 | 32 |
And below are the results of precision measurements, recall, f1 score using the Naive Bayes algorithm and TF-IDF feature extraction. We can see 79% precision value, 77% recall and 73% f1 score so we get an accurate value of 76%.

| Table 11. Confusion matrix report |
|-----------------------------------|
| Precission | Recall | F1-score |
| Positive   | 81%    | 93%     | 87%     |
| Negative   | 72%    | 91%     | 81%     |
| Neutral    | 84%    | 14%     | 24%     |
| Avg        | 79%    | 77%     | 73%     |

3.4 Validation
To ensure that the accuracy value obtained is good and taken from the sample, validation is done by using the K-Fold method, in this method the data will be tested as many as the K value, the ideal K value used is 10. This means that 10 tests will be conducted. Following are the accuracy results obtained for the value of K = 10

| Table 12. 10-fold scores |
|--------------------------|
| K      | Score  |
| 1      | 0.815  |
| 2      | 0.796  |
| **3**  | **0.832** |
| 4      | 0.815  |
| 5      | 0.747  |
| 6      | 0.691  |
| **7**  | **0.605** |
| 8      | 0.715  |
| 9      | 0.773  |
| 10     | 0.774  |
| Avg    | 0.76   |

The smallest accuracy value is 0.60 in the 7th experiment, and the greatest accuracy is in the 3rd experiment with a value of 0.83. With an average accuracy value obtained is 0.761

3.5 Visualization
The following is a picture that displays words that often appear in each class in this study.

![Wordcloud](image)

**Figure 3.** Positive class wordcloud

In positive class, the words that appear most often are barang, mudah, bagus, cepat, aman, layan, toko, bantu, sukses, and some other words that indicate that the user discusses satisfaction
when shopping because of the fast service process, the process of goods service is easy, safe, and good quality

Figure 4. Negative class wordcloud

The negative class shows the tendency of dissatisfied and disappointed user comments when buying and selling goods due to bad transactions, payment processing, and shipping with couriers. Users also complain about the process of downloading old applications and updating too often, and the absence of CS that serves complaints from users.

Figure 5. Netral class wordcloud

For neutral classes, users provide a lot of input to Tokopedia services, such as establishing a COD system for buying and selling transactions, choosing the right courier for fast delivery, shipping cost cuts. And many also provide support comments for Tokopedia for the better.

4. Conclusion
The conclusions that can be drawn from this research are as follows:
1. Sentiment analysis is done by the naive bayes algorithm and TF-IDF feature extraction
2. Based on the results of tests conducted, the accuracy of the Naive Bayes algorithm and TF-IDF is 76%. In validation with the 10-fold method, the highest accuracy is 83% and an average of 76%
3. User opinions about positive comments can be maintained and negative opinions can be an evaluation. In neutral comments there are many inputs for service improvement.

Suggestions for this research are expected in future studies to add Slankword to help the preprocessing process, and are expected to make improvements to writing errors.
Reference

[1] Pratama B Y and Sarno R 2015 Personality classification based on twitter text using naive bayes, KNN and SVM *International Conference on Data and Software Engineering* 170-174.

[2] S K B and Bordoloi M. 2018 Sentiment analysis of product using machine learning technique: a comparison among NB, SVM and MaxEnt *International Journal of Pure and Applied Mathematics* 71-83

[3] Pablo Gamallo and Marcos Garcia 2014 Citius: A naive-bayes strategy for sentiment analysis on english tweets *International Workshop on Semantic Evaluation* 8 171-175

[4] Fink C R, Chou DS, Kopecky J J, and Llorens A 2011 Coarse and fine grained sentiment analysis of social media text *Johns Hopkins Apl Technical Digest* 30 22-30

[5] Maarif A A 2015 Penerapan algoritma tf-idf untuk pencarian karya ilmiah *Jurnal Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang*.

[6] Ling J, Eka I P dan Oka T B 2014 Analisis sentimen menggunakan naive bayes classifier dengan seleksi fitur chi square *E-Jurnal Matematika* 3 92-99

[7] Vidya N A, Fanany M I, and Budi I 2015 Twitter sentiment to analyze net brand reputation of mobile phone providers *The Third Information Systems International Conference* 72 519-526