Labeling Analysis in the Classification of Product Review Sentiments by using Multinomial Naive Bayes Algorithm

V O Tama¹, Y Sibaroni² and Adiwijaya³

¹Informatics Engineering, School of Computing, Telkom University, Bandung, Indonesia
²³Computational Science, School of Computing, Telkom University, Bandung, Indonesia

Abstract. Along with the development of technology, e-commerce also experienced a fairly rapid development. The existence of e-commerce becomes another consumer alternative to make it easier for them to fulfill their needs. After buying the goods, consumers are free to assess the products they buy. Product reviews and ratings provided by consumers are one means that can be used to increase sales and can also be used to determine the decision in purchasing a product by reading the product reviews. However, using ratings and reviews alone is not enough to summarize one's opinion. Therefore, in this Final Project built a system that can classify opinions on product reviews into positive and negative sentiments by utilizing the rating. The dataset used is Grocery and Gourmet Food from Amazon as much as 50,000 which will then be labeled using Labeling Methods Average and Binary. The classification of this opinion uses the approach of Supervised learning Algorithm Multinomial Naive Bayes. The result of this research shows that labeling using Method Average is suitable for processing Grocery and Gourmet Food Dataset and proves that the best ratio of feature selection usage is 20% succeed to produce 80.48% accuracy.

1. Introduction

As time goes by, technology is undergoing rapid development. This makes the community must be able to adapt to these developments. One of the changes we can see is the shifting behavior of people in buying goods needs offline shop into an online shop or E-commerce [1]. The underlying reason for this behavior shift is the ease offered by the online shop. But in addition to having the ease, there are also shortages that shop online, which is associated with payment security and quality of goods that do not match the expectations [1]. Based on the description of the problem, required a medium that can cause a level of trust in other consumers before buying a product, one of them is by using product reviews. Reviews of the product are very important and about 73% -87% can influence the decision in purchasing a product only through reading reviews of online products for food products, hotels, and services [2]. It can bring a good impact for consumers and producers. Shopping online can help consumers to see which products they will buy have good product reviews or opinions. While the advantages of product reviews for producers are able to classify buyer reviews into negative class and positive class.

To make it easy to translate consumer reviews, e-commerce takes advantage of rating. The rating used is a star from scale of 1 to 5 [14]. But using the rating alone is not enough to be a reference for determining sentiment, because sometimes there is a discrepancy between ratings and reviews provided. Therefore, it is necessary to classify the text of product reviews to determine a sentence including into positive and negative sentiments. To classify the opinion, a sentimental analysis technique was used. Sentiments or opinion mining analysis is the process of automatically understanding, extracting and processing textual data to obtain sentiment information contained in an opinion [9]. Sentiment analysis is very important to use in product marketing, measuring market prediction and can also be used in politics.

In the classification of sentiments there is one thing to note, namely labeling. Labeling is the activity of labeling a dataset. The most commonly used method of labeling is manually using the help of language experts [3]. Labeling manually produces accurate data because humans can distinguish
precisely a sentence including into a sentiment [4]. However, it is very difficult if done by using large amounts of data, as it will require many linguists and long enough time [3]. Another method that can be used for labeling is labeling automatically. The advantage of auto-labeling is the relatively short time required. However, in previous studies relating to automatic labeling and text classification using Support Vector Machine (SVM) yielded a low accuracy of 67.83% [4]. Therefore, in this study proposed the method of automatic labeling by using other labeling methods, which is Average [5] and Binary [6]. The automatic labeling that will be used is labeling the text manually based on specific references. After conducting the labeling then will be classified product sentiment including into positive sentiment and negative sentiment by using approach of Supervised learning.

In previous studies relating to rating predictions discussed the average use of labeling on several methods such as Naïve Bayes and SVM [5]. From the research, it is found that the best rating prediction accuracy is 78.00% using Naïve Bayes method [5]. So, the classification algorithm to be used in this research is Multinomial Naïve Bayes. The Naïve Bayes Multinomial Algorithm is a development of the Naïve Bayes Classifier (NBC) that is suitable for text data processing, easy to implement, and has a minimum error rate [7].

The purpose of this study is to know the most appropriate labeling method used to label test data, to know the effect of using feature selection and without feature selection, and to know the result of system classification by implementing labeling method and using the best feature using Multinomial Naive Bayes Algorithm.

2. Literature study
In the study [10] explained that as the development of technology, the public will be much to express their opinions through the internet. To be able to know the intentions of public opinion is necessary to processing opinion called sentiment analysis. The analysis of sentiment or opinion mining is a person's attitude, thought, or judgment of an object arising from their natural feelings [12]. The beginning of 2001 marked the beginning of a growing research on sentiment analysis and opinion mining. One application of sentiment analysis is to process product reviews.

According to two surveys of more than 2000 adults in America yielding 73% -87% simply by reading reviews of restaurant products, hotels, and other services can influence their purchasing decisions [10]. Then as many as 20% -99% of respondents are willing to pay for a 5-star product rather than a 4-star. The most popular sentiment analysis is to classify those opinions into positive and negative classes. An example of the classification of positive and negative sentiments is [10].

- Positive sentiment
  ...zany characters and richly applied satire, and some great plot twists
  ...awesome caramel sauce and sweet toasty almonds. I love this place!
- Negative sentiment
  It was pathetic. The worst part about it was the boxing scenes...
  ...awful pizza and ridiculously overpriced...

Before undertaking a sentiment classification, it should be noted that the dataset being processed has the correct label. In fact, datasets that already have very small number of labels are hard to find [3]. The labeling process can be done manually alone or together and will be easy to do if the dataset is small [3]. But for the case of sentiment analysis should be needed enough data for modeling. Another solution for labeling is automatically [4]. Labeling is automatically performed on research [4] using the SVM algorithm by predicting positive, negative, and neutral classes based on patterns of training outcomes created by Vladimir Vapnik. Classification is done by a dividing line (hyperlane) that separates between positive and negative classes. Training on the SVM classification then produces a pattern that will be used in the testing process aimed at labeling sentiments. Preprocessing applied in this research that is, tokenization, case folding, filtering, and stopword removal. Accuracy result is 67.83% at k-fold to 6 with implementation of TF-IDF and k-fold cross validation with k as much as 7.

Based on the results of the application of automatic labeling with SVM [4] resulted in low accuracy. Another automated labeling method that can be used for labeling by rating is the Average method [5]
The Average labeling method was once implemented to determine the correctness of rating predictions using the Yelp review data by applying the Naive Bayes and SVM Multinomial Algorithms. They divide the labeling into positive and negative by using the numbers 1 and 0. Initially the result of the rating of the majority Yelp data is in the range of 3.5 to 4.5. From the ratings obtained, the study attempted to predict the ratings given by consumers tailored to extract the review satisfaction and to process the review data, since the text review has a more qualitative value than the rating. To label the review data, they consider each document to be related to one review and have a single rating. Finally, each rating is rounded to 1, 2, 3, 4, and 5. The average rating score is about 3.6. Then the rating below 3.6 is labeled 0 and above 3.6 is labeled 1. From the research obtained the truth of the rating prediction of 7000 data with accuracy of 78% using Multinomial Naïve Bayes method. While using SVM obtained an accuracy of 74%.

In addition to the Average labeling method, there is also labeling by the Binary method. The method is applied to research on rating predictions based on user reviews by binary labeling and Multi class on Amazon electronic data. Binary labeling is done by labeling 0 for rating 3 and label 1 for rating. Preprocessing done for Amazon data is remove punctuation, removing numbers, case folding, removing extra whitespace, removing stopwords. To apply text classification, unstructured text format should be converted into a structured format because the computer is much easier to handle numbers than text. So before entering the classifier, the text will be converted into a numbered vector using Bag of words, meaning every unique word in the document will be saved. With such models, the term frequency (the occurrence of each word) is used as a feature to train training data. But using the term frequency it is feared that all terms are considered equally important. To solve the problem, it is implemented tf-idf. Then processed with Multinomial Naïve Bayes, SVM, Random Forest and Logistic regression algorithms. The results of the SVM and Naïve bayes algorithm resulted in good performance, but SVM is better. The best F1-score is the application of Binary Labeling to 0.84, while Multi class produces 0.57 [6].

In a study of other rating predictions was done by applying Multinomial Naïve Bayes method with add one smoothing and the use of Categorical Proportional Difference (CPD) as feature selection. The use of CPDs is chosen because this method is able to find a lot of words appearing on documents using the positive and negative document frequencies. Feature selection is the stage used to select what features are most important by storing the word with the highest value according to the measurement of the degree of importance of a word. The purpose of feature selection is to improve the performance of text classification. This research implements feature usage of 25%, 50%, 75%, and 100%. Multinomial Naïve Bayes is chosen because it assumes independence between the occurrence of words in the document without taking into account the order of words and document information in general. The data used are 500 data divided each 100 data train on five rating and processed with preprocessing stages, among others tokenisation, filtering, case folding, and stemming. Accuracy was obtained at 87% with feature usage of 50%. These results are the best results after a trial of using feature ratios of 25%, 75%, and 100% [11].

3. Proposed system
The built system can classify the Grocery and Gourmet Food Dataset into positive and negative sentiments using the Multinomial Naïve Bayes Algorithm. The system built in this study is shown with figure 1.
Figure 1. System overview.

Next to the description of the system overview will be described as follows.

3.1 Dataset
The dataset used is Grocery and Gourmet Food Dataset. This dataset amounted to 151,254, but used only 50,000 data. Dataset reduction is done because of the limitations of hardware capabilities used during running feature feature extraction takes a long time. Another reason is because of the time for labeling of data testing and power that helps labeling a bit. Amazon data is known to have many attributes. Below is one example of an unprocessed dataset [13].

{"reviewerID": "AWX394GL61KBC", "asin": "9742356831", "reviewerName": "seriousshopper Ely", "helpful": [0, 0], "reviewText": "Being newbie to THai cooking, these are great addition to my kitchen. Its tastes really good for vegetable and chicken curry.", "overall": 5.0, "summary": "Taste great", "unixReviewTime": 1369785600, "reviewTime": "05 29, 2013"} The data attribute description will be described in table 1 [13].

| Attribute      | Explanation               |
|----------------|---------------------------|
| reviewerID     | ID of the reviewer        |
| asin           | ID of the product         |
| reviewerName   | Name of the reviewer      |

Table 1. Amazon data attribute,
3.2 Read data
Of the many attributes, only the attributes of reviewText and Overall or rating are used.

3.3 Labeling
Labeling will be done automatically on 50,000 data using the reference of two labeling methods. The data will be labeled with the number 1 which means positive, and the number 0 which means negative. Labeling Average [5] divides the label according to the average rating. The average rating of this data is 4.3. So, the rating below 4.3 goes into label 0 and rating above 4.3 is labeled 1. In the data labeled as average, the amount of data labeled 1 is 25,025, while the data labeled 0 is 14,975. Binary labeling [6] utilizes ratings by classifying rating 3, ie 1 and 2 as label 0 and rating 3, i.e. 3, 4, and 5 as label 1. In binary-labeled data, the number of data labeled 1 is 36,687, and data labeled 0 as much as 3312.

3.4 Split data
The 50,000 dataset labeled will be divided into training data and data testing with a ratio of 80:20. To handle the accuracy of the system is objective, the data testing compositions for labels 1 and 0 are made with a ratio of 50:50 [10]. Then to improve the accuracy results, then the data testing is done data sharing randomly five times each containing 2000 data.

3.5 Preprocessing
Based on related studies, preprocessing to handle Amazon data is remove punctuation, case folding, removing stopwords [6]. Then because Amazon data has many related sentences, then to make this data has a form in accordance with the dictionary will be done lemmatization. Differences in the use of lemmatization with stemming done in research [11] is, lemmatization can handle more complex affixed words. While stemming only remove the affixes at the beginning and at the end only [10]. Then, according to the research reference [4] tokenization will be performed. The preprocessing step will be explained in table 2.

| Attribute          | Explanation                                      |
|--------------------|--------------------------------------------------|
| helpful            | Helpfulness rating of the reviewer               |
| reviewText         | Text of the reviewer                             |
| overall            | Rating of the product                            |
| summary            | Summary of the review                            |
| unixReviewTime     | Time of the review (unix time)                   |
| reviewTime         | Time of the review (raw)                         |

**Table 2. Preprocessing process.**

| No | Process            | Input                                                                 | Output                                                |
|----|--------------------|----------------------------------------------------------------------|-------------------------------------------------------|
| 1  | Lemmatization      | Being newbie to THai cooking, these are great addition to my kitchen. Its tastes really good for vegetable and chicken curry. | Be newbie to THai cooking, these be great addition to my kitchen. It tastes really good for vegetable and chicken curry. |
| 2  | Stopword removal   | Be newbie to THai cooking, these be great addition to my kitchen. It tastes really good for vegetable and chicken curry. | newbie THai cooking, great addition kitchen. taste really good vegetable chicken curry. |
| 3  | Case folding       | newbie THai cooking, great addition kitchen. Taste really good vegetable chicken curry. | newbie thai cooking, great addition kitchen. taste really good vegetable chicken curry. |
| 4  | Remove punctuation | newbie thai cooking, great addition kitchen. taste really good vegetable                                                      | newbie thai cooking great addition kitchen taste really good vegetable |
3.6 Feature extraction

After preprocessing, the selected features will be stored in the Bag of words. [6] The example of forming a Bag of words in a text document is described in table 3, where d1 and d2 represent the document.

| Doc1: This is creamy white chocolate |
|-------------------------------------|
| Doc2: This matcha is so creamy |

Table 3. The example of forming bag of words.

| creamy | white | chocolate | this | is | matcha | so |
|--------|-------|-----------|------|----|--------|----|
| Doc1   | 1     | 1         | 1    | 1  | 0      | 0  |
| Doc2   | 1     | 0         | 0    | 1  | 1      | 1  |

3.7 Feature selection

Feature selection is used to improve text classification performance [11]. Based on previous research, feature selection method is using CPD and tf-idf. However, in this research we will try to use feature selection by selecting the lowest feature using term-frequency. The term-frequency used is logarithmic, since in this dataset the term appearance does not appear only once or twice, and the ten occurrences of term in a document are not necessarily ten times as important as a term that appears only once in the document [15]. So, the logarithm is used to adjust the frequency in the document [15].

The ratio to be used is 0, 20%, 40%, 60%, 80%, 90%, and 95%. The system will then issue the number of features before using feature selection and after using feature selection. The term frequency formula used is described in equation (3.1) [15].

\[
L_{ij} = \begin{cases} 
1 + \log f_{ij}, & \text{if } f_{ij} > 0 \\
0, & \text{if } f_{ij} = 0
\end{cases}, \quad (3.1)
\]

Where \( f_{ij} \) is the term \( i \) occurrence frequency in the document \( j \).

3.8 Development of model classifier

After the feature selection, the system will build a classifier model using Multinomial Naïve Bayes Algorithm. Multinomial Naïve Bayes is a supervised learning method that predicts a member class with a probability that is focused on text classification. This algorithm uses multinomial distribution in conditional probability [9]. The probability calculation document \( d \) having class \( c \) can be seen in the equation (3.2) [10].

\[
P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c)
\]

\( P(t_k|c) \) is the conditional probability of word \( t_k \) that appearing on a document having class \( c \). In the equation \( P(t_k|c) \) is likelihood probability \( t_k \) in class \( c \). While \( P(c) \) is the prior probability of documents appearing in class \( c \). The class determination is to compare the posterior probability results obtained, then the class with the largest posterior probability is the class chosen as the predicted result [10]. The prior probability formula can be seen in the equation (3.3).
\[ P(c) = \frac{N_c}{N} \tag{3.3} \]

\( N_c \) is the sum of category \( c \), while \( N \) is the sum of all categories. The formula for likelihood probability can be seen in the equation (3.4).

\[ P(t_k|c) = \frac{T_{tc}}{\sum_{\text{all} T_{te}}} \tag{3.4} \]

Example of calculation and classification with this algorithm will be explained in table 4 below.

**Table 4.** The example of Multinomial Naive Bayes calculation.

| Doc | Words             | Class |
|-----|-------------------|-------|
| Training | 1 | Really Good Tea Awesome | Pos |
|        | 2 | Really Bad Flavor     | Neg |
| Testing | 3 | Tea Flavor           | Pos |

Calculation:

i. Prior probability
\[
P(\text{pos}) = \frac{1}{2} \\
P(\text{neg}) = \frac{1}{2}
\]

ii. Likelihood probability
\[
P(\text{Tea}|\text{pos}) = \frac{1+1}{4+6} = \frac{2}{10} = \frac{1}{5} \\
P(\text{Flavor}|\text{pos}) = \frac{1+1}{4+6} = \frac{1}{10} \\
P(\text{Tea}|\text{neg}) = \frac{0+1}{3+6} = \frac{1}{9} \\
P(\text{Flavor}|\text{neg}) = \frac{1+1}{3+6} = \frac{2}{9}
\]

iii. Posterior probability
\[
P(\text{pos}|\text{doc3}) \quad \alpha \frac{1}{2} \times \frac{1}{5} \times \frac{1}{10} = \frac{1}{300} = 0.01 \\
P(\text{neg}|\text{doc3}) \quad \alpha \frac{1}{2} \times \frac{1}{9} \times \frac{9}{162} = \frac{1}{81} = 0.0124
\]

Since the neg-class probability value is greater than the post class, then document 3 belongs to the neg class.

### 3.9 Evaluation

Evaluation of the development of this system will be expressed by using the accuracy of the confusion matrix in table 5. Accuracy is chosen because to know the percentage of system successes made [10].

**Table 5.** Confusion matrix.

| Actual positive | Actual negative |
|-----------------|-----------------|
| Predicted positive | TP | FP |
| Predicted negative | FN | TN |

For each description in table 2, it is described as follows.

- TP is true positive, meaning the system predicts a positive and in accordance with the original data
- TN is true negative, meaning the system predicts a negative and in accordance with the original data
- FP is false negative, meaning the system predicts a positive while the original data is negative.
• FN is false negative, meaning the system predicts negative while the original data is positive. The calculation of accuracy can be seen in the equation (3.5).

\[ \text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \]  

(3.5)

4. Analysis and result

The test scenario of this study is described in table 6.

Table 6. Test scenario.

| Scenario | Goals | Explanation |
|----------|-------|-------------|
| Scenario 1 | Compare the accuracy from implementation of both labeling methods | Test the training data of average and binary labeling methods without feature selection |
| Scenario 2 | Compare the results of the use of feature selection and without the feature selection of both training data | Implementation of the best labeling training data without feature selection and using the feature selection |
| Scenario 3 | Know the results of sentiment classification and accuracy | Implementation of the best labeling training data by using the best scenario process from the second purpose |

The first test result is described in table 7.

Table 7 The first test result

| Data testing | Labelling method |   |   |
|--------------|-----------------|---|---|
|              | Average         | Binary |   |
| 1            | 81.20%          | 74.25% |
| 2            | 80.85%          | 73.75% |
| 3            | 79.90%          | 74.70% |
| 4            | 80.09%          | 75.18% |
| 5            | 80.15%          | 74.50% |
| Average      | 80.43%          | 74.47% |

From the first average test results can be known the best average accuracy of 80.43% by using the method of Labeling Average without using feature selection. While the average accuracy by using Binary Labeling method without feature selection of 74.47%. From these results it can be concluded that the implementation of the most appropriate labeling method used for Dataset Grocery and Gourmet Food is the method of labeling in Average. This happens because the average labeling is more objective in doing labeling based on the average rating that is adjusted to the amount of data available. In contrast to binary labeling that divides labeling only by using certain assumptions regardless of the composition of the rating and the amount of data.

The second test is to implement the average label dataset by using feature selection and without feature selection. The feature selection ratios used are 20%, 40%, 60%, 80%, 90%, and 95%. So, in the second test is done comparison without feature selection and the percentage sum of feature selection that use is 80%,60%,40%,20%,10%, and 5%. The result of using the second feature selection test can be seen image data 2.
From the second test result, the graph in the figure shows that in the implementation of the test using the Labeling Average, the highest accuracy is 81.65% in the test data testing 1 with a feature reduction of 80% and the lowest accuracy of 79.20% in the test data testing 3 with a feature reduction of 90%. The use of selected feature selection ratios is a multiple of 20%, starting with features of 20%, 40%, 60%, and 80% so that researchers better understand the effect of using feature selection in detail, compared with the use of a feature selection of 25%. Then when the 80% feature is used, the accuracy obtained increases. So, it was decided to increase the use of features to 90% and 95% to know more detailed results. For an analysis of each feature ratio is used as follows.

- Without feature selection produces good accuracy between 79% to 81%. It shows the feature selection process is working maximal enough in generating the classification after the preprocessing process.
- A 20% reduction in feature count on four test data has increased accuracy. This means the feature selection is able to classify by eliminating 20% of the features and improving the accuracy results for the better although there is one test data that decreased accuracy.
- A 40% reduction in the number of features in the three data tests has decreased accuracy and the other two have increased slightly in accuracy. It shows that the system is still trying to select the right term for the classification.
- Reduced number of features by 60% increased accuracy in three data testing, meaning that the system is still trying to select terms that are prioritized for classification.
- 80% reduction in feature count has increased accuracy in two test data and resulted in the best accuracy of 81.65% generated by data testing 1. At this ratio, the term frequency is less successful in selecting the priority term for classification, although there is accuracy highest resulting in one data testing.
- 90% reduction in feature count has increased accuracy in three data testing. However, two data testing experience decreased accuracy. It proves that term-frequency begins to eliminate terminology that should be important for classification.
- Reduced number of features by 95%, all data testing has decreased accuracy. This is because the use of features that cause too much information that the system needed to classify too little. So, it could be a feature that is important for the classification even wasted.

From this research, it can be concluded that the use of optimal feature selection is in the reduce 20% features, because the graph improves on almost all test data where compared with without feature selection and the best accuracy is obtained at 80% feature reduction. It shows that the term-frequency can work by removing unimportant terms and prioritizing the important term for classification.
After the first and second tests, the third test implements the dataset with the best labeling and best method, namely Average Labeling using feature selection of 20%. Use of this feature is chosen because in the second test, all the data testing has improved accuracy well. The test results are described in table 8.

| Data testing | The accuracy of the use of features 20% |
|--------------|----------------------------------------|
| 1            | 81.35 %                                |
| 2            | 80.75 %                                |
| 3            | 79.95 %                                |
| 4            | 80.14 %                                |
| 5            | 80.25 %                                |
| Average      | 80.48 %                                |

From the third test result by using the average method labeling and the use of feature selection of 20% obtained an average accuracy of 80.48%. This proves that the dataset labeling using the Average Method by implementing the Multinomial Naïve Bayes Algorithm and the use of feature selection by 20% succeeded.

5. Conclusion

Based on the results of research and analysis has been done, it can be concluded The most appropriate labeling used to label test data is the method of labeling Average with the average result accuracy of 80.43%. This happens because the average labeling is more objective in doing labeling based on the average rating that is adjusted to the amount of data available. In contrast to binary labeling that divides labeling only by using certain assumptions regardless of the composition of the rating and the amount of data.

The effect of using feature selection on the system is capable of affecting accuracy. The optimal use of feature selection is in the ratio between 80% because produces the best accuracy in data test 1. But the best amount of feature selection is reducing the 20% features, because produces the best average of accuracy and the graph has increased for all test data. It shows that the term-frequency can work by removing unimportant terms and prioritizing the important term for classification.

The result of system classification by implementing the method of labeling average and the use of feature selection by 20% which is implemented by using Multinomial Algorithm Naïve Bayes able to classify positive and negative sentiments that produce accuracy of 80.48%.

References

[1] Farki, Ahmad, Imam B dan Berto M W 2016 Pengaruh online customer review dan rating terhadap kepercayaan dan minat pembelian pada online marketplace di Indonesia 2016 (Surabaya)
[2] Pang, Bo, et al 2008 Opinion mining and sentiment analysis 2008 Foundations and Trends® in Information Retrieval 2 1-135
[3] Rachmat A and Lukito Y 2016 Sentipol: dataset sentimen komentar pada kampanye pemilu presiden Indonesia 2014 dari Facebook page 2016 (Yogyakarta)
[4] Valonia I S, Gloria V and Yuan L 2017 Penerapan sentiment analysis pada hasil evaluasi dosen dengan metode Support Vector Machine Jurnal Transformatika 7276
[5] Jason J 2011 Predicting rating with sentiment analysis 2011
[6] Marie M 2017 Predicting ratings of Amazon reviews - techniques for imbalanced datasets 2017
[7] Andrew K, James W and Heath R 2015 A practical guide to text mining with topic extraction 2015 Wiley Interdisciplinary Reviews: Computational Statistics 326-340
[8] Matthew D and Arthur S 2017 Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it 2017 Cambridge University Press on behalf of the Society for Political Methodology 26
[9] Imam F R, Pramono S H, Dahlan and Erfan A 2013 Implementasi Opinion mining (Analisis Sentimen) untuk ekstraksi data opini publik pada perguruan tinggi Jurnal EECCIS 37-43
[10] Jurafsky And James H M 2014 Speech and language processing (London: Pearson)
[11] Rosi, Fathor, Ali F dan Rizal S P 2017 Prediksi rating pada review produk kecantikan menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD)
[12] Sida W, Christopher D M 2012 Baselines and bigrams: Simple, good sentiment and topic classification 2012 Proc. of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 90-94
[13] Julian M 2016 Amazon product data. Accessed: 03 Oktober 2017 Retrieved from http://jmcauley.ucsd.edu/data/amazon/
[14] Reddy, Sarath C, et al. 2017 Prediction of star ratings from online reviews Region 10 Conference, TENCON 2017-2017 IEEE 2017 1857-1861
[15] Erica C and Tamara G K 1999 New term weighting formulas for the vector space method in information retrieval Oak Ridge National Laboratory Report.
[16] Pratiwi A I and Adiwijaya 2018 On the feature selection and classification based on information gain for document sentiment analysis 2018 Applied Computational Intelligence and Soft Computing
[17] Reynaldi A P, Mohamad S M, Nanang S H and Adiwijaya 2018 A Multi-Lable Classification on Topics of Quranic Verses in English Translation using Multinomial Naive Bayes 2018 6th International Conference on Information and Communication Technology (ICoICT.IEEE.)