Sentiment Analysis Of Hotel Review Using N-Gram And Naive Bayes Methods

Conference Paper · October 2019

DOI: 10.1109/ICIC47613.2019.8985946

3 authors:

Triyanna Widiyaningtyas
State University of Malang
60 PUBLICATIONS **220** CITATIONS
SEE PROFILE

Ilham Ari Elbaith Zaeni
State University of Malang
75 PUBLICATIONS **213** CITATIONS
SEE PROFILE

Riswanda Al Farisi
Universitas Gadjah Mada
2 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Sentiment Analysis Of Hotel Review Using N-Gram And Naive Bayes Methods

Triyanna Widiyaningtyas, Ilham Ari Elbaith Zaeni, Riswanda Al Farisi
Electrical Engineering Department
Universitas Negeri Malang
Malang, Indonesia
triyannaw.ft@um.ac.id, ilham.ari.ft@um.ac.id, riswandaalfarisi@gmail.com

***Abstract*— Hotel booking service providers in the form of websites or online-based applications have provided features where consumers can provide a review regarding their assessment of the hotel. But the number of reviews available makes users unable to filter out all the reviews. Sentiment analysis can be used as a solution to overcome this by classifying reviews into positive or negative sentiments. This study aims to determine the application of n-gram and naive bayes methods in sentiment analysis classification. The research phase includes: (1) the hotel review data collection was obtained from the TripAdvisor.com website, (2) the data preprocessing process is data cleaning and case folding, (3) the process of tokenization using the n-gram method consists unigram, bigram, and trigram, (4) the process of word weighting using Term Frequence Inverse-Document Frequency (TF-IDF) method, (5) the process of classification using the Naive Bayes method to classify hotel reviews to be positive or negative, (6) the evaluation process to determine the results of the performance of the algorithm using a confusion matrix that will produce the value of precision, recall, accuracy and error rate. Based on these results, classification using the Naive Bayes and unigram methods obtained precision results of 94%, recall 100%, accuracy 97% and error rate 3%. The bigram methods obtained precision results of 89%, recall 94%, accuracy 92% and error rate 8%. The trigram methods obtained precision results of 52%, recall 80%, accuracy 58% and error rate 42%. Based on accuracy results, It can be concluded that tokenization unigram method better than other tokenization methods.**

Keywords*— Classification, Sentiment Analysis, Naive Bayes, N-Gram, TF-IDF.

I. INTRODUCTION

The development of information through the website is currently growing rapidly along with the increasing needs of the community, especially in the hospitality sector. At this time, hotel bookings can also be through websites or smartphones, making it easier for hotel marketing to attract consumers. However, with an online ordering system consumers do not see the hotel directly to be ordered, so it takes consideration to book a hotel that suits the needs of consumers. Growing consumer dependence on online product reviews and services has caused so much user-generated content that does not have the potential for customers to filter all of these reviews [1].

One tourist site, namely *TripAdvisor.com* is a site that provides reviews or reviews from consumers at each hotel. *TripAdvisor.com* is not a booking agency or tour tour organizer. *TripAdvisor.com* has partners who provide their own hotel booking services and only help consumers choose hotels, restaurants or airline tickets at the best prices. One feature on *TripAdvisor.com* is to provide reviews or reviews from consumers from hotels or restaurants that consumers will order later, so consumers can be helped by the review feature.

Review represent what is called user-generated content and this is a concern and a rich resource for marketing teams, sociologists, psychologists and others who may care about opinions, views, public moods and public or private attitudes [2]. Some people do not express opinions in the same way. Most reviews will have positive and negative comments. This can be managed by analyzing sentences one by one [3]. With the growing popularity of the internet, electronic word of mouth (eWOM) on social media has become an important tool for customers who seek and share information on products and services. Reviews from online customers as a particular form of eWOM have become the most important source of information in making customer decisions. eWOM is considered more successful in influencing consumer behavior than traditional marketing[4]. Usually, reviews and user opinions are allocated on blogs, forums and short social networks. For example, the database from TripAdvisor.com has an average length of 108 words. In addition, his style is usually informal and full of different orthographic and grammatical errors, misspellings, typographical errors, and so on. This poor text quality causes a lot of noise [5].

Because of the amount of information or reviews given in the form of text, it will be difficult to classify the review without reading the text as a whole. Sentiment analysis aims to overcome this problem by grouping user reviews into positive or negative opinions [6]. Sentiment analysis studies the perspective, behavior and feelings or emotions of a person towards an individual, problems, activities, and subjects [7]. Grouping of reviews from consumers is influenced by emotions (sentiments) that are grouped or classified to determine their polity, namely positive or negative [8]. Sentiment analysis is a fairly popular research field, because it can provide benefits for various aspects, ranging from predictions to sales, politics, and decision-making of investors [9].

Several studies have been carried out on sentiment analysis, including the sentiment analysis on a restaurant review with Indonesian text conducted by Muthia in 2017. The study used the *naive bayes* method and genetic algorithm as a feature selection method. In this study using the n-grams tokenization process, which uses a combination of three words or trigrams. The results of this study showed an increase in naïve bayes accuracy from 86.50% to 90.50% after using the information gain feature selection [6].

As one successful method, *Naive Bayes* is very popular in text classification because of its computational, efficient, and relatively good predictive performance[10]. The research conducted by Chen in 2009 explained that naïve bayes classifier is very simple and efficient and very sensitive to the selection of features, so the research conducted by selecting special features. The research method uses two

feature evaluation metrics for the naïve bayesian classifier that is applied to multi-class text datasets: multi-class odds ratios (MOR), and class discriminating measures (CDM). Text classification experiments with naïve bayesian classifiers were carried out in two multi-class text collections. The results of the two data sets indicate that CDM and MOR are one of the best performing metrics for naïve bayes classifiers applied to multi-class text datasets.

Other studies with the title of automatic rating prediction on reviews of beauty products with naïve bayes and n-gram methods were carried out by Pujadayanti, Fauzi, and Sari in 2018. In this study using the n-gram method, namely unigram, bigram and a combination of unigram and bigram. The study gave results that Unigram had an accuracy of 93%, bigram 83% and a unigram bigram combination had an accuracy of 96%. One disadvantage that affects the results of the classification is that the data obtained is not good because the review data does not use standard words, so that it can affect the results [11].

Word weighting is a mechanism for scoring the frequency of occurrence of a word in a text document. One popular method for doing word weighting is TF-IDF (Term Frequency-Inverse Document Frequency)[12]. In the research conducted by Hadna in 2016 with the title of literature study on comparison of methods for the sentiment analysis process on twitter. From the results of the comparison it was found that the *Naive Bayes* algorithm had an average accuracy of 82.06% and a support vector machine of 82.49%.

Based on the background above, this study will apply the n-gram method to the tokenization process consisting of unigram, bigram and trigram. Because the data used is in the form of text, it will apply the word-weighting method TF-IDF. In the classification process will apply the *Naive Bayes* method for analyzing sentiment review of the hotel. The evaluation process will compare the best accuracy results from three types of n-gram tokens, including unigram, bigram and trigram.

II. METHODS

In this study using six stages, (1) data collection, (2) data preprocessing, (3) N-Gram tokenization (4) TF-IDF (5) *Naive Bayes* classification, (6) evaluation.

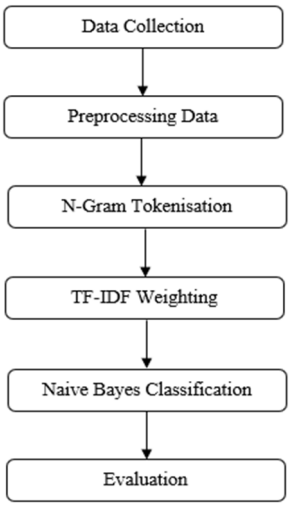

Fig. 1. Stages of Research

A. Data Collection

The data used as a dataset is hotel review that obtained from *Tripadvisor.com* website. Data taken as many as 200 data consists of 100 positive data and 100 negative data. The data is taken randomly only at hotels in Indonesia, the time range of the data set is taken randomly. On *TripAdvisor.com* website, consumers who want to review a hotel have to write the title of the review and the given review must consist of at least 200 characters and also provide ratings ranging from one to five at the hotel. Where one rating means very bad, two means bad, three means on average, four means good and five means very good. Positive or negative review data is taken based on the rating given in the review. If the rating on the review is between four stars and five stars, it will include positive review data. Whereas if the review has a one to two star star rating, it includes negative review data [13].Hotel review data obtained only from hotels in Indonesia, where hotel review data taken is not in all cities in Indonesia but taken randomly from various hotels in various cities.

B. Preprocessing Data

At this stage the hotel review data will go through the preprocessing stage to be ready to be used for the classification phase. These stages include the process of data cleaning (cleaning), and the process of uniformizing the shape of the letter (case folding). Some of the steps carried out in the data preprocessing process includes:

1. *Cleaning*
   Cleaning is a step in preprocessing to delete all punctuation marks in a sentence. Examples of deleted punctuation are like exclamation points, periods, commas and question marks. At this stage also removes all numbers in the sentence.

2. *Case Folding*
   Case folding is a step in converting capital letters to lowercase letters. At this stage all the letters in the sentence will be changed to lowercase letters. This is because the virgin can be processed to the next stage. This process is usually called uniformity of letters.

C. N-Gram Tokenization

N-gram is a tokenization process to separate words based on the type of token used. In this study three types of n-gram tokens were used, namely unigram, bigram, and trigram. Each sentence will be typed based on the three types of n-grams. Unigram is a breakdown of words in a review sentence with n = 1 or a single term. Bigram is a n-word solution in a review sentence with n = 2. While the trigram is a n-word solution in the review sentence with n = 3[11].

D. TF-IDF Weighting

Hotel review data that has gone through the preprocessing and n-gram stages will be a set of terms stored as an index. The next step is to weight the text. This stage is used to find the term with the highest frequency of occurrence and has a high weight. Term is an important word that can represent the class of a document. The term

comes from the results of n-gram tokenization which includes unigram, bigram and trigram.

To calculate the weight of each term with the TF-IDF method, it is necessary to first calculate the values of TF, DF and IDF. TF value is the number of frequency of occurrence of terms in one document d. DF value is the number of documents containing term t. IDF value is the logarithm value of the total number of documents divided by the value of DF. After calculating these values, then do the calculation from the TF-IDF method. The TF-IDF value is the product of the TF value with the IDF value calculated previously. So that the weighting of the TF-IDF will calculate the weight of each term in the types of unigram, bigram and trigram tokens. The TF-IDF weighting formula can be seen as follows:

$$\mathrm{TF(d,t)} = \mathrm{f(d,t)} \quad (1)$$

$$\mathrm{IDF(t)} = \log(\mathrm{Nd/df(t)}) \quad (2)$$

$$\mathrm{TF-IDF} = \mathrm{TF(d,t).IDF(t)} \quad (3)$$

f (d,t) : appear frequency term t on document d
Nd: the total number of all document
df(t): the total documentcontained term t

E. Naive Bayes

Naive Bayes classifier is a simple classifier based on Bayes theorem conditional probabilities and assumptions of strong independence. As a successful method, naïve bayes is very popular in text classification because of its computational nature, efficiency and relatively good predictive performance [10]. Naive Bayes is a family of simple probabilistic classifiers based on the general assumption that all features are independent of each other, given categorical variables, and are often used as a basis for classification of texts[14]. The advantage of Naive Bayes Classifier is that it requires small training data for classification. Naive Bayes classifier is easier to implement, faster to be classified and more efficient[15].

In the training process, hotel review data has determined its sentiment. Then it will be processed to form knowledge in the form of probability values in each word. This process will produce a word for each review data that characterizes the review as a positive or negative sentiment. Naive Bayes calculations can be seen in the following equation:

$$P(C_j|\,X) = \frac{P(X|C_j)P(C_j)}{P(X)} \quad (4)$$

P(Cj|X) :The probability of hypothesis C is based on condition X
P(X|Cj) :X probability is based on coniditions in hypothesis
CP(X) : probability X
P(Cj): The probability of hypothesis Cj

The value of P (X) can be ignored because the value is constant for all Cj, so the equation (4) can be written as follows:

$$P(C_j|X) = P(X|C_j)P(C_j) \quad (5)$$

$$p(c_j) = \frac{n(doc_j)}{n\,(sampel)} \quad (6)$$

p(cj): probability of category documents
n (docj): the total number of documents in a category
n (sample) : the total number of document training

$$p(x_i|c_j) = \frac{1+n_i}{n+|x|} \quad (7)$$

p(xi | cj) : probability of words in each category
ni : frequency of occurrence of words in each category
n : the number of all words in the document in a particular category|x| : total number of words (distinc) in all training data

$$c_{MAP} = argmax\ p(c_j)\ p(x_i|c_j)$$

After obtaining the results of multiplication in each category of documents, then compare and look for the largest probability value (CMAP) that is used to classify document categories.

F. Evaluation

The performance evaluation of the Naive Bayes algorithm in this study was carried out by calculating the value of Precision, Recall, Accuracy and Error Rate by using confusion matrix. With reference to Table I.

TABLE I. CONFUSION MATRIX

Correct Classification	**Clasified as**	
	+	-
+	True Positif (TP)	False Positif (FP)
-	False Negatif (FN)	True Negatif (TN)

$$Precision = \frac{TP}{TP+FP} \text{ x } 100\% \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \text{ x } 100\% \quad (9)$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \text{ x } 100\% \quad (10)$$

$$Error\ Rate = 100\% - Accuracy \quad (11)$$

Precision is the number of groups of relevant documents from the total number of documents found by the system and can measure the level of effectiveness of the system. Recall relates to the ability of the classification system to call relevant documents. Recall values and precision can have different values (Widaningsih & Suheri, 2018). While accuracy is the level of accuracy or accuracy of a prediction and error rate is the level of error or a prediction error.

III. RESULTS AND DISCUSSION

Classification using the N-Gram and Naive Bayes methods is carried out using three types of n-gram tokens including unigram, bigram and trigram. The classification results will be tested using three trials comparing training data and testing data. From the comparison the accuracy and best precision results will be used as a reference.

A. Comparing Data Sampling

Before deciding on the distribution of training data and testing which data will be used, the results of accuracy and

precision will be carried out on several experiments in sharing training data and testing data. The results of the accuracy of the distribution of data training and data testing can be seen in Table II.

TABLE II. ACCURACY RESULT

No	Data Sampling	Unigram	Bigram	Trigram
1	120:80	97%	92%	58%
2	140:60	98%	90%	60%
3	160:40	100%	95%	67%

From the test results in Table II it is known that the accuracy results of the Naive Bayes classification use several experiments comparing training data and testing data. It can be seen in the first experiment the value accuracy of unigram, bigram, and trigram tokens, namely 97%, 92%, and 58%. On the second experiment with the accuracy value of unigram, bigram, and trigram tokens, namely 98%, 90%, and 60%. In the third experiment with the accuracy value of unigram, bigram, and trigram tokens, namely 100%, 95%, and 67%. It can be seen that the accuracy of the bigram in the second experiment has decreased. So the best results are in the third experiment with a comparison of training data of 160 data and testing data for 40 data. The results of comparison of precision values can be seen in Table III.

TABLE III. PRECISION RESULT

No	Data Sampling	Unigram	Bigram	Trigram
1	120:80	94%	89%	52%
2	140:60	96%	80%	51%
3	160:40	100%	94%	60%

Based on the results of Table III, it can be see that the results of bigram and trigram precision in the second experiment is decreased. So the best results are in the third experiment with a comparison of training data of 160 data and testing data for 40 data. So the comparison of the dataset used is the third experiment.

B. Evaluation Result

The results of the Naive Bayes classification test have been obtained using a comparison of data in the third experiment. Then the next step is the evaluation process by calculating the value of Precision, Recall, Accuracy, and Error rate using confusion matrix. The calculation results can be seen in Table IV.

TABLE IV. EVALUATION RESULT

Evaluation Calculating	Result		
	Unigram	***Bigram***	***Trigram***
Precision	100%	94%	60%
Recall	100%	94%	83%
Accuracy	100%	95%	67%
Error Rate	0%	5%	33%

From the evaluation results in Table IV, it is known that the greatest accuracy value is found in the type of unigram token with the value of accuracy of 100%. The biggest precision value is also found in unigram tokens with precision values of 100%. This shows that using a type of unigram token in this study can provide high accuracy and precision compared to other types of tokens. The type of bigram tokens has an acccuracy value of 95% and precision of 94%. The type of trigram token has the lowest accuracy value compared to other types of tokens, which is 67% and precision value of 60%. For unigram and bigram tokens it has a good accuracy value because there are many similar terms in the dataset used in the two types of tokens, so it has a relatively high TF-IDF value.

IV. CONCLUSION

Classification of sentiment analysis hotel reviews can be done using the method of naive bayes and n-gram. The process of sentiment analysis is done by classifying data in the form of hotel reviews into positive or negative sentiments. The effect of the type of n-gram tokens is the more the same terms on the type of token used, the better the level of accuracy. Classification of types of unigram tokens can produce accuracy of 100%. Bigram token type provides accuracy of 95%. And trigram type tokens give accuracy results of 67%. Based on the accuracy value, the type of unigram token is better than the type of bigram or trigram token.

REFERENCES

[1] R. R. L. Barbosa, S. Sánchez-Alonso, and M. A. Sicilia-Urban, "Evaluating hotels rating prediction based on sentiment analysis services," *Aslib J. Inf. Manag.*, vol. 67, no. 4, 2015.

[2] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013.

[3] R. Safrin, K. R. Sharmila, T. S. S. Subangi, and E. A. Vimal, "Sentiment analysis on online product review," *Int. Res. J. Eng. Technol.*, vol. 4, no. 4, pp. 2381–2388, 2017.

[4] P. Phillips, S. Barnes, K. Zigan, and R. Schegg, "Understanding the Impact of Online Reviews on Hotel Performance : An Empirical Analysis," *J. Travel Res.*, vol. 56, 2017.

[5] A. García, S. Gaines, and M. T. Linaza, "A Lexicon Based Sentiment Analysis Retrieval System for Tourism Domain," *e-Review Tour. Res.*, vol. 10, no. 2, pp. 35–38, 2012.

[6] D. A. Muthia, "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Mengunakan," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017.

[7] A. Samad, H. Basari, B. Hussin, and I. G. P. Ananta, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *MalaysianTechnical Univ. Conf. Eng. Technol.*, pp. 545–552, 2012.

[8] A. Ridok and Indriati, “Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN),” *J. Environ. Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 23–32, 2016.

[9] V. Chandani and P. Romi Satria Wahono, “Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film,” *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.

[10] J. Chen, H. Huang, S. Tian, and Y. Qu, “Expert Systems with Applications Feature selection for text classification with Naïve Bayes,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, 2009.

[11] I. Pujadayanti, M. A. Fauzi, and Y. A. Sari, “Prediksi Rating Otomatis pada Ulasan Produk Kecantikan dengan Metode Naïve Bayes dan N-gram,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. November, 2018.

[12] N. Muchammad, S. Hadna, and P. I. Santosa, “Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter,” *Semin. Nas. Teknol. Inf. dan Komun.*, no. February, 2016.

[13] N. D. Pratama, Y. A. Sari, and P. P. Adikara, “Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, no. July, 2018.

[14] S. Xu, “Bayesian Naiive Bayes classifiers to text classification,” *J. Inf. Sci.*, no. 15, 2016.

[15] R. P. Rajeswari and K. Juliet, “Text Classification for Student Data Set using Naive Bayes Classifier and KNN Classifier,” *Int. J. Comput. Trends Technol.*, vol. 43, no. 1, pp. 8–12, 2017.