

PAPER • OPEN ACCESS

Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier

To cite this article: Arif Abdurrahman Farisi *et al* 2019 *J. Phys.: Conf. Ser.* **1192** 012024

View the [article online](#) for updates and enhancements.

You may also like

- [Typological and Functional Planning Features of Sports and Tourist Hotels](#)
D S Borodai, A S Borodai, S P Borodai et al.
- [Prerequisites for the formation and the small hotel's architectural typology development](#)
A A Badavi and V M Molchanov
- [Implementation of Green Hotel Management 5-Star Hotel in Jakarta, Case Study: The Dharmawangsa Hotel Jakarta](#)
I Ardiansyah and H Iskandar



The Electrochemical Society
Advancing solid state & electrochemical science & technology

243rd Meeting with SOFC-XVIII

Boston, MA • May 28 – June 2, 2023

Early registration discounts end **April 24!**

Accelerate scientific discovery!

Learn More & Register



Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier

Arif Abdurrahman Farisi ¹, Yuliant Sibaroni ² and Said Al Faraby ³

^{1,2,3} School of Computing, Telkom University

E-mail: ¹ arifabdurrahmanf@student.telkomuniversity.ac.id,

² yuliant@telkomuniversity.ac.id,

³ saidalfaraby@telkomuniversity.ac.id

Abstract. In this modern age where the internet is growing rapidly, the existence of the internet can make it easier for tourist to find any information. In the field of tourism hotel, internet is very helpful in promotion of hotel. Tourists usually tell the experience during the hotel by writing reviews on the internet. Hence many hotel's reviews are found on the internet. The impact on hotel owners is that they can take advantage of reviews on the internet to improve and evaluate their hotels. With the availability of reviews on the internet with large numbers, tourists can't understand all the reviews they read whether they contain positive or negative opinions. It takes a sentiment analysis to quickly detect if the reviews is a positive or negative reviews. This study provides a solution by classifying positive opinion reviews and negative opinions using the Multinomial Naïve Bayes Classifier method and comparing models using preprocessing, feature extraction and feature selection. The best experimental results using preprocessing and feature selection with 10 fold cross validation have an average F1-Score more than 91%.

1. Introduction

When tourists want to choose a comfortable hotel for their trip, they will look for reviews from other travelers [1]. 93% of hotel management states that online tourist reviews are critical to the future of their property [2]. Reviews available on the internet are more relevant, actual and detailed than the reviews found in hotel brochures [3]. But with the sheer number of reviews, they can't understand and summarize all the hotel reviews whether they contain positive opinions or negative opinions.

Some hotel reviews websites only provide ratings that are considered not objective and therefore cannot be used as comparisons between hotels [3]. For an example on reviews of Ibis Hotel, there are reviews: "Bad service and disappointing facilities" [15]. But the hotel guest gave the score 4 stars. While at Aston Hotel there are reviews: "I like to stay here a lot. The wall and lobby decoration full of Indonesian reliefs and carving " [16]. The hotel guest gives a score 3 stars. Because of that, from the results of the review Ibis Hotel is not necessarily better than Aston Hotel and vice versa. The above example can be concluded that the rating system valuation is different from the written reviews so it cannot be used for hotel comparison because the information provided is not clear.

From the previous research, the classification of sentiments classification using NLP and Bayesian Classification [4]. The Naïve Bayes algorithm method has proved effective in sentiment classification and performs better than any other method. Multinomial Naïve Bayes is used because of its speed and



simplicity in text classification [5]. Beside the Multinomial Naïve Bayes method also follows the principle of the multinomial distribution used for text processing. In the previous study used bag of word for its feature extraction [1]. Bag of Word has weaknesses such as running time is longer and results are not optimal because it accommodates many features. In text classification, feature selection can improve efficiency and accuracy [18].

In this research the authors conducted a study on the classification of automatic sentiments for hotel reviews provided by hotel guests. The dataset is taken from Datafiniti's Business Database which is the site of product reviews, hotels and property reviews database. This study choose the Multinomial Naïve Bayes method on the hotel reviews by expecting a more achievable accuracy. Research provides a solution to model feature selection on text classification with Multinomial Naïve Bayes to compare model performance on bag of word. Selection of the proposed feature there are two namely the selection of frequency-based feature and feature selection by removing features that have a minimum difference value of positive and negative probability. Furthermore the authors add preprocessing to determine the effect of performance on each model.

Structure of this paper is as follows: in section 2 presents related work, in section 3 is about the research method. The experimental results will be presented in section 4. The conclusion of this research will be presented at the end section.

2. Related work

In this research using some references from previous studies. This chapter describes some of the research related to our research. Research Feature Based Sentiment Classification for Hotels Review Using NLP and Bayesian Classification [4] background with the classification of sentiments is very influential on the decision of a person in choosing a hotel. The focus of this research is to improve feature extraction and add term weighting to the word reviews of hotels. Then use Naïve Bayes algorithm to determine the positive and negative classification easily. Naïve Bayes is used because it includes a simple method and has good performance. The results of this study obtained an accuracy of 96.1%.

Research on A Sentiment Analysis Model for Hotel Reviews Based on Supervised Learning [1] discusses the sentiment analysis of hotel reviews with the SVM method using bag of word frequency and TF-IDF. It was shown in the experiment that the bag of word frequency was lower and less effective. Obtained F1-Score at bag of word frequency of 86.4% and F1-Score on TF-IDF of 87.2%. Research on Comparative Studies on Breast Cancer Classification with K-Fold Cross Validation Using Machine Learning Techniques [14] discusses the analysis of breast cancer with k-fold cross validation algorithm. Conduct a comparison analysis with the Decision Tree method, Naïve Bayes, Neural Network and SVM. In general, this study used $k = 10$ but the authors did experiments for other k . The result of the experiment is the best SVM (Linear) and DT accuracy algorithm using $k = 3$, SVM (MLP) and NB best accuracy using $k = 5$, SVM (RBF) and NN best accuracy using $k = 10$. Overall the most optimal accuracy is to use the SVM (RBF) 10 fold method with amount of 98.3%.

Research on Term-Frequency-based features Selection methods for Text Categorization [19] research describes research on the use of feature selection for text categorization research. Feature selection is an important step to remove features so as to make the system more effective and improve performance. Selection of features used are document frequency, information gain and mutual information. Then the authors want to improve the performance of DF, IG and MI by using term-frequency. The experimental results using Naïve Bayes in DF resulted in 80.9281% micro-averaging precision, in IG resulting in micro-averaging precision of 83.9%, in MI resulting in micro-averaging precision of 68.3%.

3. Research methods

3.1. System Overview

The system to be created in this final project is an automatic classification system in the hotel reviews. The process of sentiment analysis is taken from the dataset of English hotel reviews which is then done preprocessing to the dataset. Analysis of the classification of this hotel reviews will result in a positive

opinion orientation and negative opinion using the Multinomial Naïve Bayes Classifier. Added feature extraction and feature selection on the classification as a comparison of model performance. The process of this study is illustrated in figure 1:

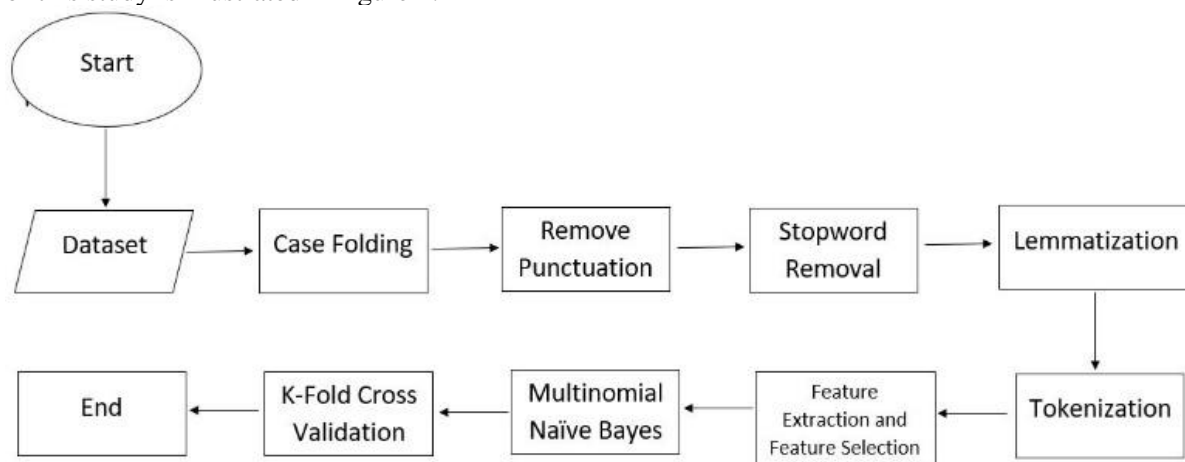


Figure 1. Flowchart system overview.

3.2. Dataset

In this research the dataset is derived from Data finitis's Business Database which contains hotel reviews of as many as 5000 english sentences in csv. The attributes that are in the dataset consist of city, country, hotel name, rating, and review. Then do attribute reduction to fit the research needs. The attribute used is the text review attribute. Then do the manual labeling according to the review sentence to label 1 (positive) and 0 (negative). The result of manual labeling is 3946 sentences is label 1 and 1053 sentences are label 0. To validate the research, k-folds cross validation is used for data sharing with $k = 10$. Furthermore this dataset is done preprocessing to facilitate the system in the process of classification. Here is an overview of the dataset used in the study. Some examples of hotel reviews in data set can be seen in table 1.

Table 1. Illustration of dataset.

Text Review	Label
Lovely view out onto the lagoon. Excellent view. Staff were welcoming and helpful.	1
Really lovely hotel. Stayed on the very top floor and were surprised by a Jacuzzi bath we didn't know we were getting! Staff were friendly and helpful and the included breakfast was great! Great location and great value for money. Didn't want to leave!	1

3.3. Preprocessing

Data Preprocessing is a process for making low quality data into high quality data making it easy to process [6]. There are several data preprocessing techniques used in this research which are reduction of dataset dimension, case folding, remove punctuation, stopwords removal, lemmatization, and tokenization. Dimensional reduction is the selection of the dimensions needed for research. In this case the dimensions taken are text review. Case folding is the process whereby all the letters are converted into lowercase. Only the letters 'a' to 'z' are accepted. In addition the letter is considered a delimiter or word separator.

Remove punctuation is the process of removing punctuation in a sentence. Tokenization is the process of cutting the input string based on each compiler word. The principle works to separate every word that compiles a document. In this process the removal of numbers, punctuation and characters, because the character is considered as a separator of the word and has no effect on text processing. Stopword removal is the process of removing less important words that often appear on documents. To shorten the classification process, it can eliminate stop words such as "which", "the", "and".

Lemmatization is the process of converting a word into a word or root word for each word of tokenization. With the process of lemmatization, each word affixed will be removed and turned into a basic word so that it can further optimize when done text processing [8]. Examples of the use of lemmatization are the words "applied", "words", "saw" to "apply", "word", "see". Table 2 is exsample preprocessing of the research dataset.

Table 2. Illustration of preprocessing process.

Text Review	Preprocessing
Lovely view out onto the lagoon. Excellent view. Staff were welcoming and helpful.	'lovely', 'view', 'onto', 'lagoon', 'excellent', 'view', 'staff', 'welcoming', 'helpful'
Really lovely hotel. Stayed on the very top floor and were surprised by a Jacuzzi bath we didn't know we were getting! Staff were friendly and helpful and the included breakfast was great! Great location and great value for money. Didn't want to leave!	'really', 'lovely', 'hotel', 'stayed', 'top', 'floor', 'surprised', 'jacuzzi', 'bath', 'didnt', 'know', 'getting', 'staff', 'friendly', 'helpful', 'included', 'breakfast', 'great', 'great', 'location', 'great', 'value', 'money', 'didnt', 'want', 'leave'
Room was tiny-bed saggy-bAthroom door didn't work. Good breakfast and convenient location. Wouldn't return or recommend.	'room', 'tiny', 'bed', 'saggy', 'bathroom', 'door', 'didnt', 'work', 'good', 'breakfast', 'convenient', 'location', 'wouldnt', 'return', 'recommend'

3.4. Feature extraction

Bag of word is a way to extract features with text used for modeling like machine learning algorithm [9]. Bag of words is also a representation that describes the text of a sentence or document. Steps to get a Bag of words collect data containing text, create a list containing all the vocabulary that appears in the sentence, each sentence is described as a token by ignoring the order of words, calculating the frequency of occurrence or the number of occurrences of words present in the sentence. Example of the bag of word process is the sentence 1 "The best hotel in USA" and sentence 2 "The Aston hotel is very dirty". Then a built dictionary contains tokens ("The", "best", "hotel", "in", "USA", "Aston", "is", "very", "dirty"). The above sentence has 9 different words so that each sentence is described as follows, sentence 1 [1,1,1,1,1,0,0,0,0] and sentence 2 [1,0,0,0,0,1, 1,1,1].

Laplace smoothing is a technique of adding probability of one or small numbers to avoid zero probability on feature [12]. Laplace smoothing is used in this study. With this technique the final result will not be zero. It takes Laplace Smoothing to avoid the process of failing analysis. The way it works is to add one or more small numbers into the feature so as not to generate a zero probability. In this study using add 1 smoothing.

3.5. Feature extraction

Feature selection is the process of selecting features for training data and using only some features in the text classification to simplify when building a model [10]. Additionally applied selection features to implement a more efficient classifier as it reduces vocabulary and can improve accuracy as it can eliminate noise features. In this study using two feature selection to be used as comparison of model performance. The first feature is frequency-based [19] or deletes features that have the lowest word occurrence frequency and the second feature is to remove features that have a minimum difference of positive and negative probability values.

3.6. Learning and classification using Multinomial Naïve Bayes Model

Multinomial Naïve Bayes Classifier is a supervised learning method that uses probability and is focused on text classification cases. This method follows the principle of multinomial distribution in conditional probability [10]. Although using multinomial distributions, this algorithm can be applied to text cases

by converting to a nominal form that can be computed with an integer value. The probability calculation is described in the equation 1:

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

$P(t_k | c)$ is the conditional probability of the word tk that appears in the document having class c . In the equation $P(t_k | c)$ is likelihood probability tk in class c . While $P(c)$ is the prior probability of the document 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. The prior probability formula can be seen in the formula 2:

$$P(c) = \frac{N_c}{N} \quad (2)$$

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 formula 3:

$$P(t_k|c) = \frac{T_{tc}}{\sum_{t' \in V} T_{ct'}} \quad (3)$$

T_{tc} is the number of occurrences of the word t in the document having class c , and $\sum_{t' \in V} T_{ct'}$ is the total number of occurrences of all words in class c . Next it will explain the example calculation Multinomial Naïve Bayes used in this study. Examples of datasets are described in table 3, while the results of conditional probability calculations can be seen in table 4.

Table 3. Illustration of preprocessing process.

No	Dataset	Text	Label
1	Data Training	The hotel is clean and great!	Positive
2	Data Training	The hotel owner is very helpful.	Positive
3	Data Training	Overall Aston Hotel's experience was great!	Positive
4	Data Training	The condition of the hotel was very bad.	Negative
5	Data Training	A HORRIBLE EXPERIENCE FOR ONE WEEK!!!	Negative
6	Data Test 1	The hotel view was great	Positive
7	Data Test 2	My holiday experience stay in usa so horrible	Negative
8	Data Test 3	Overall the hotel in aston very clean and great	Positive

Calculation of prior probability

Total of sentences = 5
 P(positive) = 3/5
 P(negative) = 2/5

Calculation of posterior probability and classification of data test

Data Test 1:

- $P(\text{positive}|\text{sentence}) \propto P(\text{hotel}|\text{positive}). P(\text{great}|\text{positive}). P(\text{positive}) = 0.0100$
- $P(\text{negative}|\text{sentence}) \propto P(\text{hotel}|\text{negative}). P(\text{great}|\text{negative}). P(\text{negative}) = 0.0016$

The conclusion : Data test 1 is classification as positive.

Table 4. Example of probability after add 1 smoothing obtained from data training.

After Smoothing					
No	Words	Positive	Negative	P(word Positive)	P(word Negative)
1	Hotel	4	2	4/26	2/22
2	Clean	2	1	2/26	1/22
3	Great	3	1	3/26	1/22
4	Owner	2	1	2/26	1/22
5	Very	2	2	2/26	2/22
6	Helpful	2	1	2/26	1/22
7	Overall	2	1	2/26	1/22
8	Aston	2	1	2/26	1/22
9	experience	2	2	2/26	2/22
10	condition	1	2	1/26	2/22
11	Bad	1	2	1/26	2/22
12	Horrible	1	2	1/26	2/22
13	One	1	2	1/26	2/22
14	Week	1	2	1/26	2/22
Total	14	26	22	1	1

Data Test 2:

- $P(\text{positive}|\text{sentence}) \propto P(\text{experience}|\text{positive}).P(\text{horrible}|\text{positive}).P(\text{positive}) = 0.0017$
- $P(\text{negative}|\text{sentence}) \propto P(\text{experience}|\text{negative}).P(\text{horrible}|\text{negative}).P(\text{negative}) = 0.0033$

The conclusion : Data test 2 is classification as negative.

Data Test 3:

- $P(\text{positive}|\text{sentence}) \propto P(\text{overall}|\text{positive}).P(\text{hotel}|\text{positive}).P(\text{aston}|\text{positive}).P(\text{very}|\text{positive}).P(\text{clean}|\text{positive}).P(\text{great}|\text{positive}).P(\text{positive}) = 0.0100$
- $P(\text{negative}|\text{sentence}) \propto P(\text{overall}|\text{negative}).P(\text{hotel}|\text{negative}).P(\text{aston}|\text{negative}).P(\text{very}|\text{negative}).P(\text{clean}|\text{negative}).P(\text{negative}) = 0.0016$

The conclusion : Data test 3 is classification as positive.

From the above calculation obtained:

Prediction = [1, 1, 1]

Test = [1, 0, 1]

TP = Predict (1) & Test (1)

FN = Predict (0) & Test (1)

FP = Predict (1) & Test (0)

TN = Predict (0) & Test (0)

Note = positive (1) & negative (0)

Model performance evaluation results are calculated using the confusion matrix[11] as seen in table 5. based on the confusion matrix, we can calculate the value of precision, recall, F-Measure and accuracy.

Table 5. Confusion matrix.

	Actual positive	Actual negative
Predicted positive	2	1
Predicted negative	0	0

3.7. K-Fold Cross Validation

K-Fold Cross Validation is a cross validation technique to evaluate a model where data is split into two subsets ie learning process data and validation data [13]. Models are trained by a learning subset and validated by a subset of validations. Usually this method is used because it can reduce computational time while maintaining the accuracy of the estimation. For the selection k can be based on the size of the research needs. In general the study chose and used $k = 10$ to obtain optimal accuracy [14]. In this study using 10 fold cross validation as validation of system model that already created.

4. Result and discussion

4.1. Experiment Scenario

On this research, we are doing sentiment classification to the hotel review with positive and negative labels. To get the optimized results, there were 2 scenarios in table 6 built for testing that's made for model comparison, that is:

Table 6. Research scenarios.

No	Scenario	Purpose
1	Calculate the performance model using preprocessing and not using preprocessing	To know the comparative performance of the preprocessing effect on the model
2	Calculate the bag of word performance, feature selection 1 that is frequency based, and feature selection 2 that is erasing the feature with difference of the lowest positive and negative probability	To know the most optimal performa for the model classification

This research is done by a few stages, that is preprocessing on the dataset, after producing, token will be doing comparison using the feature extraction, feature selection 1 and feature selection 2. Feature extraction that is used is called bag of word. While feature selection is using 2 methods that is feature selection 1 with frequency based or erasing feature that has low word appearance frequency and feature selection 2 with erasing feature with the minimum difference of positive and negative value. The classification stage using Multinomial Naïve Bayes, and evaluation with 10-fold cross validation.

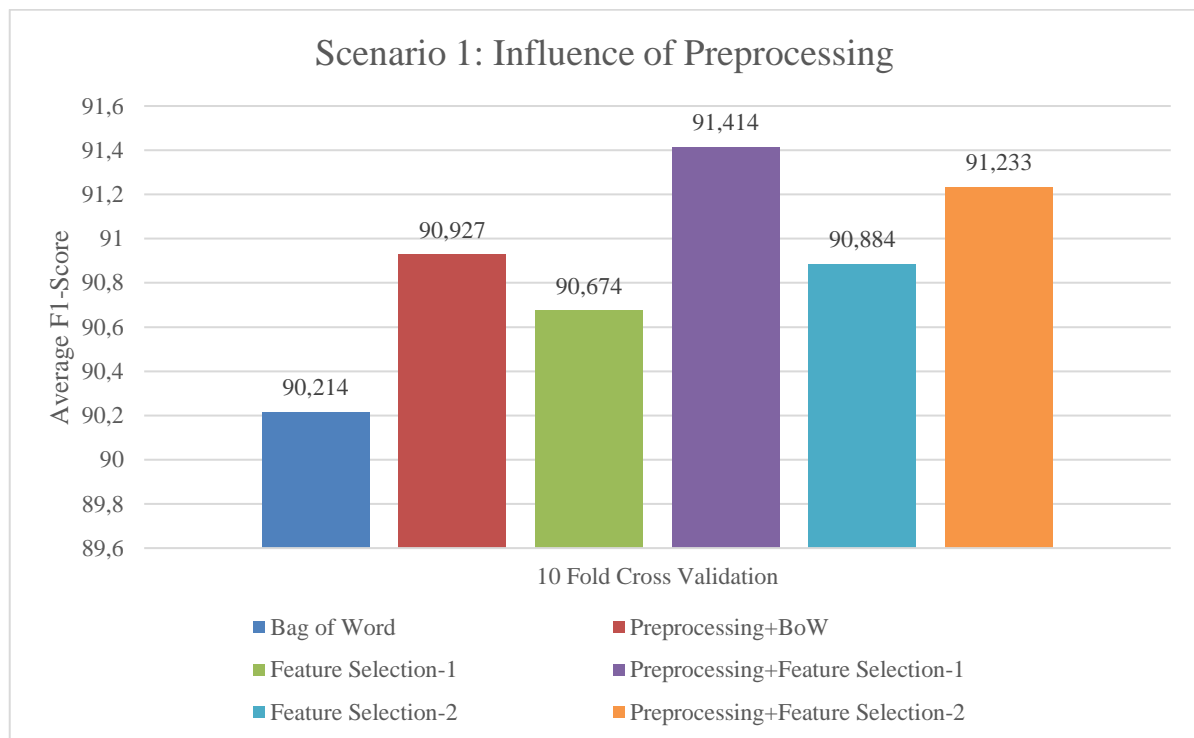
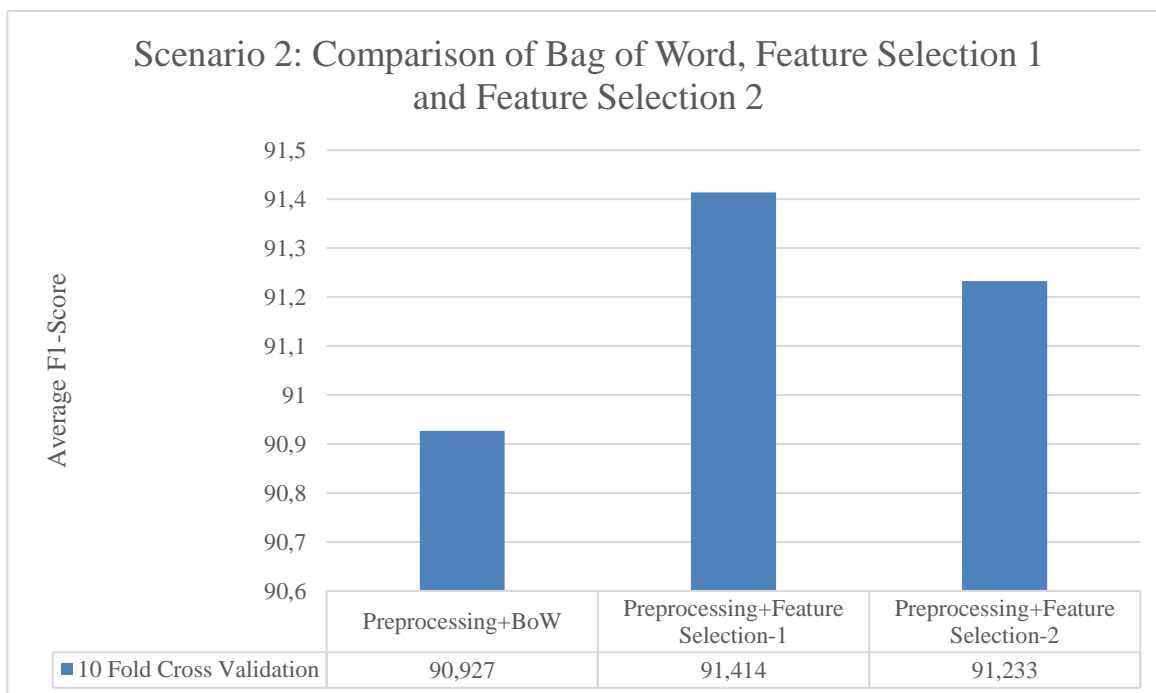
5. Experiment and Analysis Result

The results of scenario 1 can be seen in figure 1 while the results of scenario 2 can be seen in figure 3.

By doing a test scenario will result in different performance because each scenario can affect the model built. Each scenario is validated using 10 fold cross validation. In figure 2 we can see the results of testing that preprocessing can improve system performance. It is proven that only tokenization uses only average F1-Score 90.2%, 90.7%, 90.9% lower than using preprocessing which has average F1-Score 90.9%, 91.4%, 91.2%. Because preprocessing can eliminate unneeded words and make the word more structured in the created model so that it is more efficient and performance will increased.

Then in figure 3 the most optimal result is the feature selection method 1 with average F1-Score 91.4%. While the lowest is the bag of word method with average F1-Score 90.9%. Because the bag of word all the features can be accommodated so as to make the model less effective in classification. In addition to the many features without the selection of features can reduce system performance and require more time for running time.

When viewed from the two figures, when without using preprocessing the largest average F1-Score is the feature selection method 2 is 90.884%. This is interesting because when compared to preprocessing, feature1 selection is higher than selection feature 2 with average F1-Score 91.4%, so it can be concluded that preprocessing will improve performance on feature selection 1 and get better value than feature 2 selection.

**Figure 2.** Scenario 1 result.**Figure 3.** Scenario 2 result.

6. Conclusions

Based on the results of testing and analysis that has been done, it can be concluded as follows: The system is built successfully classified positive sentiment orientation and negative sentiment with the most optimal result is F1-Score average of 91.4% using preprocessing and feature selection 1 with frequency based on the classification process. The use of preprocessing in figure 1 greatly affects the

performance of the system so it can be seen if using preprocessing F1-Score average results can improve performance optimally. The system that uses the feature 1 selection and feature 2 selection has a higher F1-Score average than the bag of word in this case.

7. References

- [1] H X Shi and X J Li "A sentiment analysis model for hotel reviews based on supervised learning," in *International Conference on Machine Learning and Cybernetics* China 2011
- [2] "6 Tren Wisata Utama Tahun 2016," tripadvisor 14 December 2015 [Online] Available: <https://www.tripadvisor.co.id/TripAdvisorInsights/w665> [Accessed 19 February 2018]
- [3] W Kasper and M Vela "Sentiment Analysis for Hotel Reviews," in *Proceedings of the Computational Linguistic-Applications Conference* 2011
- [4] T Ghorpade and L Ragha "Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian Classification," in *International Conference on Communication, Information & Computing Technology (ICCICT)* Mumbai India 2012
- [5] A Goel J Gautam and S Kumar "Real time sentiment analysis of tweets using naïve bayes," *2nd International Conference on Next Generation Computing Technologies (NGCT)* pp 257-261 2016
- [6] J Han M Kamber and J Pei Data Mining: Concepts and Techniques 3rd Edition Elsevier 2011
- [7] A Valdivia M V Luzon and F Herrera "Sentiment Analysis in TripAdvisor," *IEEE Intelligent Systems* pp 72-77 2017
- [8] Maliffauzi "Text Pre-Processing," 27 02 2016 [Online] Available: maliffauzi.lecture.ub.ac.id/files/2016/02/Text-Pre-Processing-v2.pptx [Accessed 22 10 2017]
- [9] J Brownlee "A Gentle Introduction to the Bag-of-Words Model," 09 10 2017 [Online] Available: <https://machinelearningmastery.com/gentle-introduction-bag-words-model/> [Accessed 21 02 2018]
- [10] C D Manning P Raghavan and H Schütze An Introduction to Information Retrieval Cambridge: Cambridge University Press 2009
- [11] D M Powers "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *International Journal of Machine Learning Technology* 2 pp 37-63 2011
- [12] P Goyal "What is Laplacian smoothing and why do we need it in a Naive Bayes classifier?," Quora 27 September 2017 [Online] Available: <https://www.quora.com/What-is-Laplacian-smoothing-and-why-do-we-need-it-in-a-Naive-Bayes-classifier> [Accessed 22 April 2018]
- [13] A Wibowo "10 FOLD-CROSS VALIDATION," BINUS 24 November 2017 [Online] Available: <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/> [Accessed 5 May 2018]
- [14] Z Nematzadeh R Ibrahim and A Selamat "Comparative Studies on Breast Cancer," in *IEEE Kinabalu* 2015
- [15] "Ibis Singapore," tripadvisor 14 08 2017 [Online] Available: https://www.tripadvisor.com/Hotel_Review-g294265-d1229624-Reviews-Ibis_Singapore_on_Bencoolen-Singapore.html [Accessed 03 04 2018]
- [16] "Aston," tripadvisor 02 05 2016 [Online] Available: https://www.tripadvisor.com/Hotel_Review-g187147-d228765-Reviews-Aston_Hotel_Paris-Paris_Ile_de_France.html [Accessed 03 03 2018]
- [17] R WP A Novianty and C Setianingsih "Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy," in *e-Proceeding of Engineering* Bandung 2017
- [18] C Jingnian H Huang S Tian and Y Qu "Feature selection for text classification with Naïve Bayes," *Expert Systems with Applications: An International Journal* pp 5432-5435 2009
- [19] Y Xu and L Chen "Term-frequency Based Feature Selection Methods for Text Categorization," in *Fourth International Conference on Genetic and Evolutionary Computing* Shenzhen 2010

- [20] Sibaroni Y Widyantoro D H and Khodra M L (2015) Survey on research paper's relations 2015 International Conference on Information Technology Systems and Innovation (ICITSI)