APPLICATION OF OPINION CLASSIFICATION FOR FASTFOOD RESTAURANT POPULARITY USING INSTAGRAM

Elvandio Frinandy[1)], Dedi Trisnawarman[2)], Ery Dewayani[3)]

[1) 2) 3)] Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta
email: elvandio.825160039@stu.untar.ac.id[1)]

ABSTRACT

The purpose of this study is to identify several opinions regarding Instagram account of fast food restaurant in Indonesia. By using a system that can provide the information about the popularity ranking of fast food restaurants, researcher conduct an analysis using Naïve Bayes Classifier. To identify the popularity ranking, researcher collected restaurants data from Instagram, preprocessing text, using Naïve Bayes classification, and implementing the programing. The design of this system be able to determine the popularity ranking of fast food restaurants using collected data (number of likes and comments on Instagram's post). All of the predictions obtained exceeds of 90%, indicating a fairly accurate prediction of the popularity ranking, where all results almost touch the maximum score. Data collected from the restaurants' Instagram account, indicating positive results and ratings, based on number of likes and comments per post. The result of this research shows that, Burger King restaurants are the most popular, second is Kentucky Fried Chicken, third is McDonalds, fourth is Hokben, and fifth are Wingstop..

Kewyords

Instagram, Naïve Bayes, Fast Food Restaurant, Sentiment Analysis.

1.Introduction

In 2019, there were 150 million active social media users in Indonesia out of 368.2 million residents. 60% of the adult population in Indonesia use smartphone to access internet. The majority of internet users have used it for 2-3 years, with 93.9% of users connected via smart phones every day [1]. 24% of internet user, using Yelp, Zomato, TripAdvisor, Facebook and Instagram, on selecting restaurant in comparison to other platform. Average rating is the most influential variable in restaurant selection through social media platforms, as much as 24%. These results indicating that, it is essential for consumers to share their experiences on social media platforms in the form of reviews. The next most influential factor of restaurant selection is social media platform with the percentage of 19%. Social media is a platform for restaurants to publish written information and images to provide visual stimulation and provide a proper perception of the food and atmosphere in the restaurant [2].

The number of reviews written online by visitors has a significant positive impact on restaurant performance. In addition, the overall rating by visitors, numbers of visitor served per hour, and the quality of the food offered by restaurants, have an impact on restaurant performance. Restaurants with good certification, the total number of great reviews or assessment or a good assessment, can increase the number of sales, and the number of visitors [3]. Sentiment analysis is a type of research that focused on developing information explosion situation on the internet. Sentiment analysis or often also called opinion mining is a computational study of people's opinions, appraisal and emotions through entities, events and attributes owned. The previous studies related to sentiment analysis include research on the Naïve Bayes method for sentiment analysis of Spanish tweets on Twitter. The previous study analyzed Twitter as a corpus linguistics, using data collected from language materials, as the source for sentiment analysis. And the result developed a system that could identify and classify people's sentiments to predict interesting products in marketing [4].

In this study, opinion classification is used to determine restaurants' popularity ranking based on data collected, such as number likes, and comments, and classification of comments or opinions on Instagram, about fast food restaurants. This study, use a classification strategy based on the Naïve Bayes method, for the reason that it is a simple and intuitive method, and which performance is similar to other approaches. In addition, based on previous studies, Naïve Bayes combines efficiency (optimal performance time), and accuracy.

2. Related Work

There are several studies that have examined sentiment analyst using social media data that can help in analyzing opinions based on comments where opinions are highly unstructured, either positive, negative, or neutral. the first research to analyze public opinion or called sentiment analyst that can be extracted from social media data. The research method used is data crawling from Facebook, Instagram, and Twitter. For popularity ranking obtained from the comment count, Facebook likes count, Facebook were here count, Facebook talking about and Instagram visitors [4]. the second research shows a variety of methodologies. Modeling for data in the form of tweets must be different from data in the form of customer reviews, this is because the form of text and information taken from tweets is very limited compared to customer reviews. How to find the appearance of this word using TF-IDF calculation. This research uses several sources to compare models from Facebook, Google+, Yelp, etc [5]. The third study uses the SVM classification algorithm to categorize customer reviews of their products. Although only using one algorithm, this study focuses on improving the text before doing the classification, namely with feature classification and polarity classification. This is considered to help eliminate unnecessary information in the data mining process. Benefits of preprocessing before documents are classified as reducing size when modeling, and this is rarely the focus of other studies [6].

3. Method

The data collected based on restaurants data on Instagram, preprocessing text using Naïve Bayes classification, classification, and testing. (Figure 1) is the framework of thinking.

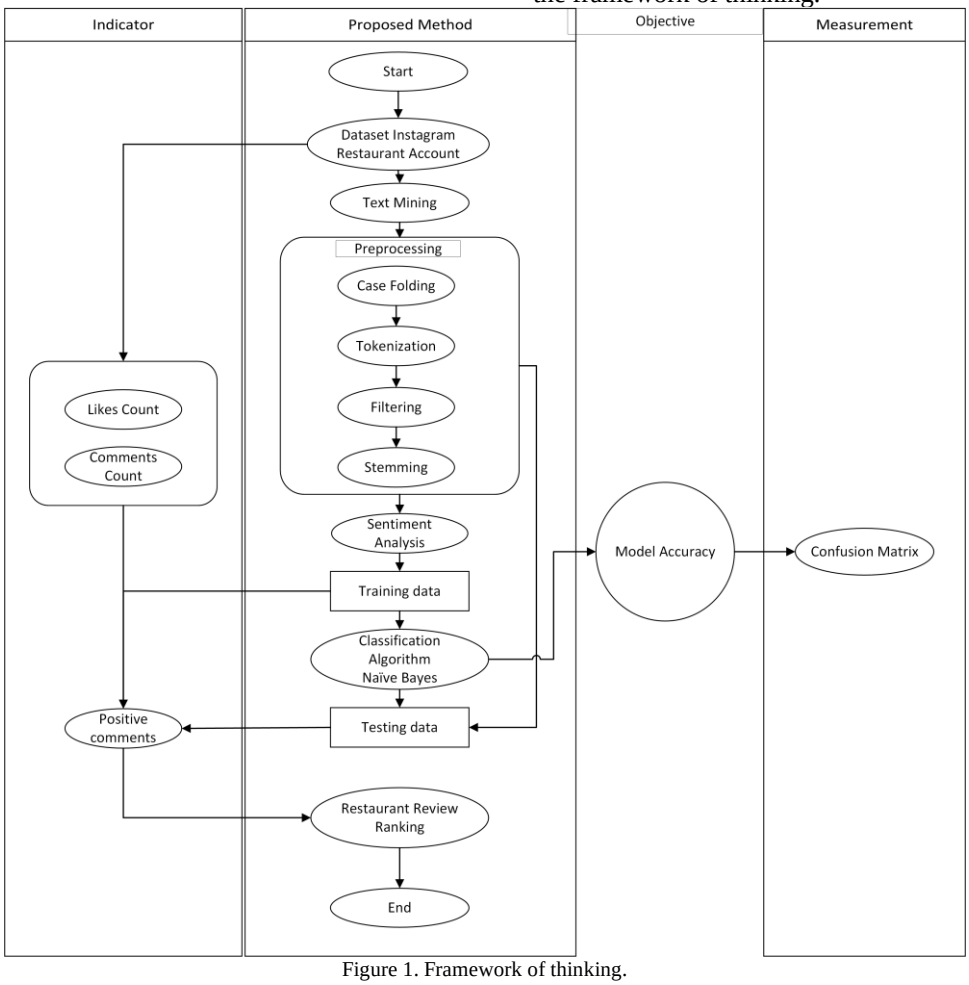

Figure 1. Framework of thinking.

Below are some explanations of the method used:

3.1. *Collecting data*

The data used in this research, collected from Instagram comments. The software used to scraping data is RStudio. To collect the data, several usernames were collected from active social media restaurants accounts. Restaurant's account that were used in this research are, kfcindonesia, mcdonaldsid, wingstopid, burgerking.id, and hokben_id. Then the study was conducted on April 29, 2020. For the search process there are several tools that will be used in RStudio, namely Instagram-scraper. After running the existing

code, the data that have been collected, will be stored in the form of a CSV file, and after that the data will be labeled as positive and negative manually.

3.2. *Preprocessing text*

- Case Folding is a step that change all the letters in a document into lowercase [7]
- Tokenization is the process of separating text into small pieces and input strings based on each word or token [7].
- Filtering Stop words are commonly used in sentiment analysis research to prepare texts, and therefore could be easily processed [8].
- Stemming is a procedure in the IR system that transform the words contained in a document into its root words using certain rules [9].

3.3. *Classification*

The Naïve Bayes Classification method is a simple model for classification. This classification works well in text classifications. Given the probabilistic classification based on the application of Naïve Bayes the independent assumption has a strong theorem [10]. Naïve Bayes is a simple method but has a high accuracy, even better than other classifier models [11]. The Naïve Bayes method is located in the testing section after the sentiment analysis process, with the aim of predicting sentiment from the comments that have been obtained. The results of this method are casted into 20% of randomly determined testing data.

3.4. *Testing*

After the result has been gathered, testing is needed to prove the performance and accuracy of these results. There are 5 tests in this study, Confusion Matrix, Accuracy, Precision, Recall, and F-Measure.

4. Result & Discussion

4.1. *Collect Data*

After scraping data using RStudio, 4217 comments were found on Kentucky Fried Chicken restaurants, 673 comments on Wingstop Restaurants, 4228 comments on McDonalds Restaurants, 2564 comments on Hokben restaurants, and 6521 comments on Burger King restaurants. From the total amount of comments that have been collected, some users have different opinions in each restaurants' account.

4.2 *Preprocessing text*

It can be seen in that the comment data after preprocessing text is cleaner because there are no symbols, conjunctions, uppercase, and so on.

4.3. *Classification*

The result of the Naive Bayes classification is 1 indicating for positive aspect and 0 for negative.

4.4. *Testing*

Below are five results of testing:
Confusion Matrix:

Table 1. Kentucky Fried Chicken Restaurant Confusion Matrix.

Nilai Prediksi	**Nilai Aktual**	
	Positive	**Negative**
Positive	768	29
Negative	7	24

Table 2. Wingstop Restaurants Confusion Matrix.

Nilai Prediksi	**Nilai Aktual**	
	Positive	**Negative**
Positive	131	5
Negative	1	1

Table 3. McDonalds Restaurant Confusion Matrix.

Nilai Prediksi	**Nilai Aktual**	
	Positive	**Negative**
Positive	824	9
Negative	9	15

Table 4. HokBen's Restaurant Confusion Matrix.

Nilai Prediksi	**Nilai Aktual**	
	Positive	**Negative**
Positive	506	13
Negative	2	7

Table 5. Burger King Restaurant Confusion Matrix.

Nilai Prediksi	**Nilai Aktual**	
	Positive	**Negative**
Positive	1200	41
Negative	13	17

Table 6. Accuracy Score.

Restoran	**Nilai**	**Persentase**
Kentucky Fried Chicken	0.9565217391304348	96%
Wingstop	0.9565217391304348	96%
McDonalds	0.9789964994165694	98%

HokBen	0.9715909090909091	97%
Burger King	0.957513768686074	96%

Table 7. Precision Score.

Restoran	Nilai	Persentase
Kentucky Fried Chicken	0.9636135508155583	96%
Wingstop	0.9632352941176471	96%
McDonalds	0.9891956782713085	99%
HokBen	0.9749518304431599	97%
Burger King	0.9669621273166801	97%

Table 8. Recall Score.

Restoran	Nilai	Persentase
Kentucky Fried Chicken	0.9909677419354839	99%
Wingstop	0.9924242424242424	99%
McDonalds	0.9891956782713085	99%
HokBen	0.9960629921259843	99%
Burger King	0.98928276991756	99%

F-Measure:

Table 9. F-Measure Score.

Restoran	Nilai	Persentase
Kentucky Fried Chicken	0.9770992366412213	98%
Wingstop	0.9776119402985075	98%
McDonalds	0.9891956782713085	99%
HokBen	0.9853943524829601	98%
Burger King	0.9779951100244499	98%

4.5. Implementation on Web Platforms

In order to produce information on the popularity of Restaurants destinations the results of sentiment analysis are implemented on a website that can be viewed publicly at the address reviewrestoran,xyz. There are two main pages the Restaurant information page and the ranking as shown in (Figure 2).

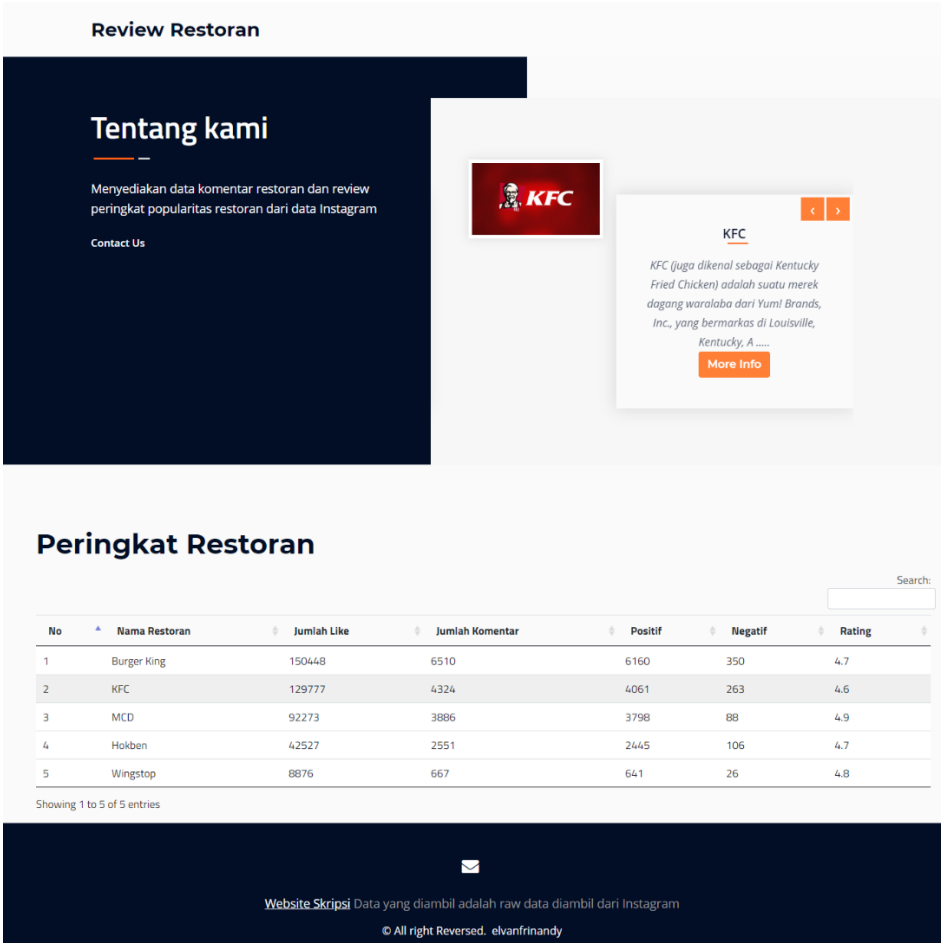

Figure 2. Homepage

Rating method:

The method that were used to rank the restaurants popularity, is based on a very simple method, there is no weighting aspect on each assessment component.
a. The scale that were used ranging from 0-1.
b. The component that were used to identify the number of likes, and positive comments.
c. Then a ranking can be made based on the number of likes, positive, negative comments, and ratings.

5. Conclusion

The result of this study shows that the prediction of the restaurant's popularity exceeds as much as 90%. In accordance with the outcome, indicating a fairly accurate prediction of the restaurants' popularity, which all results almost reach the maximum score. This can be concluded that restaurants that have positive results, and ratings based on number of likes and comments on Instagram, are the most popular restaurant. In order, the most popular restaurant based on this research are, Burger King restaurant, the second Kentucky Fried Chicken, the third McDonalds, the fourth Hokben, and the fifth is Wingstop,

REFERENCES

[1] BOC statistik pengguna digital dan internet Indonesia. https://www.boc.web.id/statistik-pengguna-digital-dan-internet-indonesia-2019/.

[2] Santos, D. P. D. (2017). The influence of online social media platforms on the choice of restaurants (Doctoral dissertation).

[3] Timilsina, M. (2017). Impacts of social media in restaurant businesses: A case study of restaurants based on Oulu region.

[4] Murnawan, M. (2017). Utilization of Sentiment Analysis for Improving Popularity of Tourist Destinations . Journal of Post and Information Research, 7 (2), 109-120.

[5] Zhang, Y., & Desouza, P. (2014). Enhance the Power of Sentiment Analysis. International Journal of Computer, Information, Systems and Control Engineering

[6] Soliman, T. H. A., Elmasry, M. A., Hedar, A. R., & Doss, M. M. (2012, October). Utilizing support vector machines in mining online customer reviews. In 2012 22nd International Conference on Computer Theory and Applications (ICCTA) (pp. 192-197). IEEE.

[7] Proses Preprocessing: Case Folding, Tokenizing, Filtering, Stemming. (Triawati 2009). http://tugasakhirkami.blogspot.com/2012/04/preprocessing.html.

[8] Jianqiang, Z. and Xiaolin, G., (2017). Comparison research on text pre-processing methods on twitter sentiment analysis. IEEE Access, 5, pp.2870-2879

[9] Ledy Agusta. 2009. Comparison of Porter's Stemming Algorithm with Nazief & Adriani's Algorithm for Stemming Indonesian Text Documents. https://yudiagusta.files.wordpress.com/2009/11/196-201-knsi09-036-perbandingan-algoritma-stemming-porter-dengan-algoritma-nazief-adriani-untuk-stemming-dokumen-teks-bahasa-indonesia.pdf

[10] C. Tseng, N. Patel, H. Paranjape, T. Y. Lin, S. Teoh (2012), "Classifying twitter data with *Naïve Bayes* classifier," in IEEE International Conference on Granular Computing,.

[11] Xhemali, D., Hinde, C.J. dan Stone, R.G. (2009). *Naïve Bayes* vs. decision trees vs. neural networks in the classification of training web pages.