

PUBLIC SENTIMENT ANALYSIS OF PASAR LAMA TANGERANG USING K-NEAREST NEIGHBOR METHOD AND PROGRAMMING LANGUAGE R

¹Hustinawaty, ²Rama Al Azis Dwiputra, ³Tavipia Rumambi

^{1,2,3}Department of Information Systems, Faculty of Computer Science and Information Technology, Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

^{1,3}{hustina, tavipia } @staff.gunadarma.ac.id, ²ramaalazis@yahoo.co.id

Abstract

Pasar Lama Tangerang is a tourist attraction in the city of Tangerang. With the development of current technology, the public can provide an overview of how the facilities and services are provided by expressing opinions on the internet. However, it is difficult to distinguish which opinions belong to positive or negative opinions. Sentiment analysis is needed to overcome this problem. The stage in sentiment analysis starts with collecting data first, then the data is processed. Furthermore, the data that has been propagated is given a sentiment classification using the K-Nearest Neighbor (KNN) algorithm. Then the classification results obtained an accuracy of 83% with a value of $k = 1$ of 120 data divided by 92 positive and 28 negative comments. Sentiment analysis is made using the R and Rstudio programming languages as supporting software

Keywords: Classification, KNN, opinion, Pasar Lama, sentiment

INTRODUCTION

In line with the rapid development of current information technology, the utilization of technological advances can be seen from the ease with which all people in society search for information that is spread freely on the internet. Only by entering keywords will get the desired information, a series of information about it will appear very quickly and easily. Pasar Lama Tangerang is a tourist attraction in the city of Tangerang. With the amount of information that is spread free of charge on media information, especially those on the internet such as the web, blogs, and social media, most people can assess how the Pasar Lama Tangerang in terms of facilities

and services provided by providing comments on media websites that can be done freely and easily. Information written by the public is sometimes not only in the form of opinion, but it is not uncommon to reveal facts about something or only express emotions.

A comment written by the public about a matter is an opinion about the evaluation of a service quality and in general that opinion is about satisfaction or actually expresses the disappointment of consumers from a quality of service that is being enjoyed by consumers. However, it will be difficult to determine which opinion is classified as positive or negative opinion. The number of positive and negative opinions will determine the decision of the new consumer will take in

visiting the Pasar Lama Tangerang, to solve this problem can be done by making a sentiment analysis. One site that can be utilized in this analysis is a review on Google. The reason to use Google, is because almost everyone uses Google as a search engine. Google has a feature that can track the whereabouts of each place visited by Google Maps location and provide feedback services for users in the form of ratings and comments from each user.

This analysis uses the K-Nearest Neighbor (K-NN) method. There are several advantages of the K-NN method, namely the K-Nearest Neighbor classification algorithm which is proven to achieve good accuracy results and in accordance with the calculations applied in an application [1].

The performance of K-NN as a classification algorithm is quite good shown by several studies that use it. The K-Nearest Neighbor algorithm is very commonly used to categorize texts [2]. This is known because the algorithm is easy and efficient for text classification. Not only easy and efficient, the nature of the K-Nearest Neighbor algorithm is self-learning, where the algorithm can study existing data structures and categorize itself.

RESEARCH METHODOLOGY

Sentiment analysis starts with reading a dataset in the form of comment data that has been made previously in Microsoft Excel totaling 120 data, after reading the dataset

then the preprocessing stage that consists of the case of folding, tokenizing, stopword removal and stemming.

From the preprocessing stage will be presented in the histogram and wordcloud visualization. Then the preprocessed dataset is divided into two data: training data with 102 data and testing data with 18 data. Furthermore, at the classification stage, calculations will be performed based on the K-Nearest Neighbor method to display the results of the classification of comments and system accuracy.

RESULTS AND DISCUSSION

In this research, there will be testing and analysis of the results of the analysis of public sentiments towards the Tangerang Old Market using the K-Nearest Neighbor method. The data tested were 18 data with a value of $k = 1$. Confusion matrix calculation in RStudio is calculated based on the classification of actual and predictive sentiments, and will then be read back by the system forming predictive and actual data such as Table 1. Classes that are predicted manually are displayed at the top of the table, while classes that are read by the system are displayed on the left, so that the table can be calculated for accuracy [3].

1. Calculate *Sensitivity*

$$Sensitivity = \frac{TP}{P} = \frac{13}{16} = 0,8125$$

2. Calculate *Specificity*

$$Specificity = \frac{TN}{N} = \frac{2}{2} = 1$$

3. Calculate *Precision*

$$PrecisionPositive = \frac{TP}{TP+FP} = \frac{13}{13+0} = 1$$

$$PrecisionNegative = \frac{TN}{TN+FN} = \frac{2}{2+3} = 0,4$$

4. Calculate *Accuracy*

$$Accuracy = Sensitivity \frac{P}{(P+N)} + Specificity \frac{N}{(P+N)}$$

$$Accuracy = 0,8125 \frac{16}{(16+2)} + 1 \frac{2}{(16+2)} = 0,722 + 0,11 = 0,832$$

After doing the classification, the values of accuracy, sensitivity, specificity,

and precision. Comparison of review data contained in Table 1 as many as 18 reviews, consists of 3 negative testing data and 15 positive testing data. While those that have been classified by the system, as many as 16 reviews are positive sentiments and 2 reviews are negative sentiments. Based on accuracy, the results of the review classification obtained from the sentiment analysis system using K-Nearest Neighbor with k = 1 value of 83% with 100% positive precision and 40% negative precision.

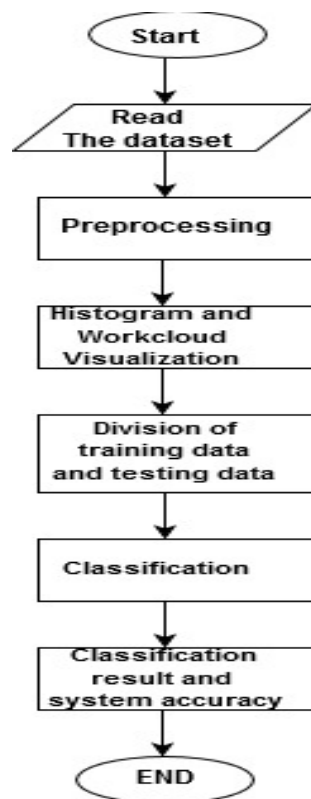


Figure 1 Sentiment Analysis

Table 1. Confusion Matrix

Amount of testing data : 18	Based on Actual Analysis of Negative Sentiments	Based on Actual Analysis of Positive Sentiments	Amount of Predicted Sentiments
Based on Negative Sentiment Prediction Analysis	TN (True Negative) = 2	FP (False Positive) = 0	2
Based on Positive Sentiment Prediction Analysis	FN (False Negative) = 3	TP (True Positive) = 13	16
Amount of Actual Sentiments	5	13	18

CONCLUSION AND SUGGESTION

This study analyzes people's sentiments towards the Pasar Lama Tangerang because it is to find out how the public's response to the Pasar Lama Tangerang in the facilities and services provided is in the form of positive and negative opinions. The sentiment analysis process begins with reading a dataset that has been made as much as 120 data, preprocessing to process raw data into data that is ready to be classified, histogram and wordcloud visualization, training data distribution and testing data in the form of 102 training data and 18 data testing, making data models , classification using the K-Nearest Neighbor algorithm, and finally the results of classification and system accuracy using the

Confusion Matrix produces an accuracy of 83% with 100% positive precision and 40% negative precision. From

120 data generated with 92 positive comments and 28 negative , with the many positive comments can show that the Pasar Lama Tangerang is worth visiting as a tourist attraction of the city of Tangerang for the people of the city of Tangerang and outside the city of Tangerang as consumers.

BIBLIOGRAPHY

- [1] R. R. Sani, J. Zeniarza, and A. Luthfiarta, "Pengembangan aplikasi penentuan tema tugas akhir berdasarkan data abstrak menggunakan algoritma *K-Nearest Neighbor*", In Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (Sendi_U) ke-2, 2016, pp. 103 - 111.
- [2] Y. Samuel, R. Delima, and A. Rachmat, "Implementasi metode *K-Nearest Neighbor* dengan *decision rule*

- untuk klasifikasi subtopik berita”, *Jurnal Informatika*, vol. 10, no. 1, pp. 1 – 15, 2014.
- [3] I. Menarianti, “Klasifikasi *data mining* dalam menentukan pemberian kredit bagi nasabah koperasi”, *Jurnal Ilmiah Teknosains*, vol. 1 No. 1, pp. 36 – 45, 2015.
- [4] R. R. Sani, J. Zeniarza, and A. Luthfiarta, “Penerapan algoritma *K-Nearest Neighbor* pada *information retrieval* dalam penentuan topik referensi tugas akhir”, *Journal of Applied Intelligent System*, vol. 1, no. 2, pp. 123 – 133, 2016.
- [5] Mustakim and G. Oktaviani, “Algoritma *K-Nearest Neighbor Classification* sebagai sistem prediksi predikat prestasi mahasiswa”, *Jurnal Sains, Teknologi dan Industri*, vol. 13, no. 2, pp. 195–202, 2016.
- [6] F. Agus, H. R. Hatta, and Mahyudin, “Pengklasifikasian dokumen berbahasa Arab menggunakan *K-Nearest Neighbor*”, *Jurnal SIFO Mikroskil*, vol. 18, no. 1, pp. 43–56, 2017.