Sentiment Analysis for Go-Jek on Google Play Store

Sitaresmi Wahyu Handani1*, Dhanar Intan Surya Saputra2, Hasirun3, Rizky Mega Arino4, Gita Fiza Asyrofi Ramadhan5

Department of Informatics Engineering, STMIK Amikom Purwokerto, Central Java 53123, Indonesia12345
sita.handani@amikompurwokerto.ac.id1*, dhanarsaputra@amikompurwokerto.ac.id2, hasirun84@gmail.com3, rizqimegarino@gmail.com4, gitaasyrofiramadhan@gmail.com5
* corresponding author

Abstract. Go-Jek is one of the most popular providers of online transportation services in Indonesia that has now grown to become the on-demand mobile platform and the leading application that provides a full range of services ranging from transportation, logistics, payments, food delivery services, and various other services. Compared to similar services from its competitors, Go-Jek is still ranked first. In order to maintain the best service to consumers, Go-Jek must be able to improve service in various aspects continuously. One of them is the GoJek application on the Google Play Store. This study aims to measure the extent of the sentiment analysis results given by customers to Go-Jek through the comments column on the Play Store. Customer opinion is taken to get positive, negative or neutral comments.

1. Introduction

Technology is growing fast, as well as applications in smartphones based on Android, Windows or iOS operating systems. The rapid growth of technology has accumulated a lot of data, data with a lot of valuable information will be useful when the data is processed and used [1]. Those developments continue and increase in various aspects, such as economics, education and also transportation. There are two important elements in transportation that is transfer and changing the place of goods and passengers to other places [2].

Good transportation can help improve the economy of a region. The conveniences offered on online transportation make the level of public trust. Online transportation service providers in Indonesia are increasingly growing such as Gojek, Grab, Uber, My Blue Bird, or in some areas there are Blue Jek, Lady Jek, and OmJek. Some of the main reasons for using online transportation is due to ease of access through existing smartphone applications, cheap, and safe. In addition, due to the use of online transport services that are more likely to use the concept of on-demand business, the service is based on the existence of consumer demand [3] so that the service providers will provide better convenience to consumers [4] customer satisfaction and loyalty is one of the main goals of a company [5]. Measurement of customer satisfaction on the performance of information systems needs to be done as one of the evaluation and improvement of the system used. The results of performance measurement of the system used in the form of recommendations for the company so that in the future the system used can be implemented better [6].

Go-Jek gojek as one of the popular online transportation services in Indonesia is managed by Go-Jek Indonesia or PT. Applications Karya Karya Bangsa. GoJek was founded in 2010 as a two-wheel transportation company through phone calls, Go-Jek is now grown into on-demand mobile platform and the leading application that provides a full range of services ranging from transportation, logistics, payment, food delivery, and various other on-demand services [7][8]. Currently the Go-Jek mobile app
is available for Android and iOS smartphone. Users can download the app for free on PlayStore or AppStore.

Until the end of 2017, Go-Jek apps on the Google Play Store have been downloaded more than 10 million times and rated 4.4. This number shows that Go-Jek is very acceptable by users. Until early July 2018 it was recorded more than 1 million comments in the PlayStore Review column [9]. Their comments ranged from positive comments, complaints, criticisms and suggestions. If further investigation of the comment will be obtained a sentiment that if collected will be able to make conclusions like what User Experience than Go-Jek Application users.

Sentiment analysis is a computational study of opinions, sentiments, and emotions expressed in text [10]. The basic task in sentiment analysis is to classify the polarity of the text present in documents, sentences, or opinions. Polarity means whether the text present in documents, sentences, or opinions has positive or negative aspects [11]. Measurement of the similarity of text that is comparing text with available references to show the level of similarity between these objects. There are many studies on the similarity of texts and produce various approaches and algorithms [12]. In recent years, several methods of sentiment analysis have been developed for many domains such as the health sector both in terms of disease and health services, entertainment fields such as film reviews, music, to the political field [15][13][14][15][16].

One of the machine learning techniques for sentiment analysis is the Naïve Bayes Classifier (NBC). NBC is a probabilistic-based machine learning technique. NBC is a simple method but has high accuracy and performance in text classification [7][18][19][20]. The purpose of sentiment analysis on Go-Jek application in this article are to find out the level of user satisfaction based on comments in the Google Play Store review column and to find out the level of service success, deficiencies and weaknesses of Go-Jek applications based on community sentiments.

2. Methods
2.1 Sentiment Analysis Process

Sentiment analysis process consist of identifying data, performed the pre-processing phase, classify all the data using Naïve Bayes method [22] then the classification results sorted according to the type that has been determined.

![Flowchart of Sentiment Analysis Process](image)

Fig. 1. Flowchart of Sentiment Analysis Process

2.2 Pre-processing

The pre-processing stage is required to clear data from unnecessary data, with the aim of entering into the Naïve Bayes Classifiers method more optimally in its calculations. This stage involve
the recognition of the content and the structure of the text adapted to its grouping [23]. Here are the steps of pre-processing:

- Cleansing. This is a process of cleaning up unnecessary words to reduce noise. The omitted words are URL, hashtag (#), username (@username), and email and also punctuation marks.
- Case folding. This is the stage of changing the form of words into the same shape, whether it be all the lower case or the upper case.
- Tokenizing. This stage works to identify words in the text into several sequences that are cut off by spaces or special characters.
- Filtering. This stage plays a role to remove the words that often appear and are general, showing less relevance to the text. The words to be discarded are defined in the stop word list.
- Stemming. This is the stage to create a word that affixed back to its original form.

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier is a classification using simple statistics based on the Bayes theorem that assumes that the existence or absence of a class with other features [24][25]. There are several stages done such as the formation of features (words that have sentimental value), calculate the probability of each sentiment class, calculate the probability of each feature, calculate the Vmap of each class, calculate the maximum Vmap. Vmap is a calculation used by Naïve Bayes Classifier to determine the probability of data test from each class based on the learning process [25].

The Naïve Bayes Classifier mathematical model, equation (1):

$$P(C|F_1, ..., F_n) = \frac{p(C)p(F_1, ..., F_n|C)}{p(F_1, ..., F_n)}$$ (1)

After the formation of features, then proceed by calculating the probability of each class, with the following equation, equation (2):

$$p(c_i) = \frac{fd(c_i)}{|D|}$$ (2)

Information, equation (2):

- fd(c_i) = number of documents belonging to the class
- c_i|D = amount of training data

After the probability of each class is obtained, then calculate the probability of each feature in the sentiment class in the following way, equation (3):

$$p(w_k|c_i) = \frac{f(w_k, c_i)+i}{f(c_i)+|W|}$$ (3)

Information, equation (3):

- f(w_k, c_i) = value of occurrence of word w_k on class c_i
- f(c_i) = total number of word occurrences in class c_i
- |W| = total amount of w_k

3. Results and Discussion

3.1 Data Input

The input data used in this study was obtained from the comments of Go-Jek on the Google Play Store. Some comments from Go-Jek users in the Google Playstore are used as data input.

3.2 Implementation of Pre-processing

Before entering the comment clarification process using Naïve Bayes Classifier, the data is processed first in the pre-processing stage such as case folding, tokenizing, stemming, filtering. With the aim that during the classification phase the results will be more optimal.

The result of the preprocessing process can be done the formation of features that is to determine the word that has sentiments, Data trainings calculated the probability of occurrence of words in each class that has been determined. The next step is to determine the probability of each feature in the sentiment class with the results seen on Table I.
After obtaining the training data, the stages are continued into the classification process to test the Vmap value.

Calculates Vmap on the test data to determine the highest probability of each class.

\[ Vmap = p(wk|c)xp(c) \]  \hspace{1cm} (4)

a. Vmap for positive sentiment
Vmap(positive) = P(positive) P(repair| positive) P(error| positive) P(driver| positive)
Vmap(positive) = \(\frac{1}{2} \times \frac{1}{14} \times \frac{1}{14} \times \frac{1}{14} = 0.714286\)

b. Vmap for negative sentiment
Vmap(negative) = P(negative) P(repair| negative) P(error| negative) P(driver| negative)
Vmap(negative) = \(\frac{2}{2} \times \frac{1}{8} \times \frac{1}{8} \times \frac{1}{16} = 1.3125\)

c. Vmap for neutral sentiment
Vmap(neutral) = P(neutral) P(repair| neutral) P(error| neutral) P(driver| neutral)
Vmap(neutral) = \(\frac{2}{2} \times \frac{1}{14} \times \frac{1}{14} \times \frac{1}{7} = 1.285714\)
From the Vmap calculation it can be gathered that the categories of the test data are included in the negative sentiment because it has the greatest chance of occurrence of the word which is 1.3125. So that the test data document will be clarified into the character with the highest Vmap value.

4. Conclusion

The results of research conducted at the end of 2017 (November and December), negative sentiment results were obtained. This study shows the results of user satisfaction carried out at the end of the year, where negative results can occur due to many other aspects besides aspects of GoJek services. It is necessary to study other aspects that might make users feel less satisfied using GoJek services in the next research. But at least these results can be used as material for GoJek's evaluation to improve its services.

References

[1] Sulaiman, S., Rahim, N., & Pranolo, A. (2016). Generated rules for AIDS and e-learning classifier using rough set approach. International Journal of Advances in Intelligent Informatics, 2(2), 103-122. doi:https://doi.org/10.26555/ijain.v2i2.74
[2] Abbas, Salim. 2000. Manajemen Transportasi, Cetakan Pertama, Edisi Kedua. Ghalia Indonesia. Jakarta.
[3] Osterwalder, A., dan Pigneur, Y. (2010). Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons
[4] Fahland, D., Favre, C., Koehler, J., Lohmann, N., Völzer, H., & Wolf, K. (2011). Analysis on demand: Instantaneous Soundness Checking of Industrial Business Process Models. Data & Knowledge Engineering, 70(5), 448-466.
[5] Saputra, D. I. S., & Handani, S. W. (2016) Implementasi Gamifikasi Pada Sistem Informasi Pelanggan Dan Pemesanan Iklan Berbasis Online Pada Surat Kabar. Seminar Nasional APTIKOM (SEMNASTIKOM), 535-540
[6] Mustofa, A., & Handani, S. W. (2017). Pengukuran Kinerja Sistem Informasi Tata Kelola Keuangan Kantor Kecamatan Kemranjen Kabupaten Banyumas Menggunakan Framework Cobit 5.0 Pada Domain MEA (Monitor, Evaluate, and Assess). PRO BISNIS, 10(2).
[7] Eka, R. (2016). Mengenal Valuasi Startup dan Istilah “Unicorn”. retrieved July, 2018, from https://dailysocial.id/post/mengenalvaluasi-star-
tup-dan-istilah-unicorn/
[8] Go-Jek. (2017). One App For All Your Needs. retrieved July, 2018, from https://www.go-jek.com/about/
[9] PT. Aplikasi Karya Anak Bangsa., GO-JEK - Transportasi Ojek, Delivery, Pembayaran retrieved July, 2018, from https://play.google.com/store/apps/details?id=com.gojek.app&hl=ind
[10] Liu, B. (2012). Sentiment Analysis and Opinion Mining. San Rafael: Morgan & Claypool Publishers.
[11] ElMessiry, A., Zhang, Z., Cooper, W. O., Catron, T. F., Karrass, J., & Singh, M. P. (2017, August). Leveraging Sentiment Analysis for Classifying Patient Complaints. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 44-51). ACM.
[12] Thet, T. T., Na, J. C., & Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of information science, 36(6), 823-848.
[13] Salas, J. (2018). Generating Music from Literature Using Topic Extraction and Sentiment Analysis. IEEE Potentials, 37(1), 15-18.
[17] Thanvi, P. R., Sontakke, N. S., Waghmare, S. R., Patel, Z. S., & Gavhane, S. (2017). Sentiment Analysis for Political Reviews using AAVN Combinations. International Research Journal of Engineering and Technology, 4(04).

[18] Routray, P., Swain, C. K., & Mishra, S. P. (2013). A survey on sentiment analysis. International Journal of Computer Applications, 76(10).

[19] Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2016). Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes. Prosiding SNST Fakultas Teknik, 1(1).

[20] Routray, P., Swain, C. K., & Mishra, S. P. (2013). A survey on sentiment analysis. International Journal of Computer Applications, 76(10).

[21] Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2016). Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naive Bayes. Prosiding SNST Fakultas Teknik, 1(1).

[22] Saputra, D. I. S., & Amin, K. M. (2016, August). Face detection and tracking using live video acquisition in camera closed circuit television and webcam. In Information Technology, Information Systems and Electrical Engineering (ICITSEE), International Conference on (pp. 154-157). IEEE

[23] Hidayat, E., Firdausillah, F., Hastuti, K., Dewi, I., & Azhari, A. (2015). Automatic Text Summarization Using Latent Driclet Allocation (LDA) for Document Clustering. International Journal of Advances in Intelligent Informatics, 1(3), 132-139. doi:https://doi.org/10.26555/ijain.v1i3.43

[24] Gull, K., Padhye, S., & Jain, D. S. (2017). A Comparative Analysis of Lexical/NLP Method with WEKA’s Bayes Classifier. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), 5(2), 221-227.

[25] Zhen, R., Jin, Y., Hu, Q., Shao, Z., & Nikitakos, N. (2017). Maritime Anomaly Detection Within Coastal Waters Based on Vessel Trajectory Clustering and Naïve Bayes Classifier. The Journal of Navigation, 70(3), 648-670.