PAPER • OPEN ACCESS

Sentiment Analysis of Online Transportation Service using the Naïve Bayes Methods

To cite this article: M Tika Adilah et al 2020 J. Phys.: Conf. Ser. 1641 012093

View the article online for updates and enhancements.



IOP ebooks[™]

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection-download the first chapter of every title for free.

Sentiment Analysis of Online Transportation Service using the Naïve Bayes Methods

Tika Adilah M^{1*}, Hendra Supendar², Rahayu Ningsih³, Sri Muryani⁴, and Kusmayanti Solecha⁵

¹ Ilmu Komputer, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jl. Kramat Raya No. 18, Jakarta Indonesia

 2,3,5 Teknologi Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Jakarta Indonesia

⁴ Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jl. Kramat Raya No. 18, Jakarta Indonesia

E-mail: 14002364@nusamandiri.ac.id

Abstract. Sentiment analysis is a computational study of opinions and emotions expressed textually. Sentiment analysis will group text in sentences or documents to find out the opinions expressed in the sentence or document, whether negative or positive. This sentiment analysis research was conducted on the online taxi transportation service. Gojek uses a lot of social media to communicate with its customers, one of the social media used is Instagram. This research takes 1000 comments from the Instagram of the Gojek page which is used to see the public opinion of the online Gojek transportation services. Comments from the page are processed by doing text preprocessing and then classified using the Naive Bayes Classifier (NBC) method to obtain the value of the public value for online transportation services. The results of this study using the Naive Bayes method resulted in an accuracy value of 81.00%, which means that from all the comments on the Instagram page, the subject of the NBC method could be accurately classified by 81.00% whether the comments were negative or positive comments.

1. Introduction

Social media is an internet-based application based on ideology and Web 2.0 technology and allows users to exchange messages in real-time [1]. The ability of social media that can produce two-way communication is one of its strengths. In addition, the nature of social media that is able to build two-way communication allows social media to become a forum for sharing ideas and information, working together, meeting new people, communicating, thus making it almost resemble human interaction [2].

The social media referred to in this study is Instagram. Instagram is a social media that is much in demand by the wider community, seen from Instagram's growing social media users. As said by APAC's Brand Development Lead Paul Webster, based on the latest data to date active users of Instagram worldwide reach up to 400 million users. more than half of the 100 million new users who join Instagram come from Asia and Europe. Indonesia itself is one of the countries with the highest Instagram users, around 89% of Instagrammers in Indonesia are aged 18-34 years. in one week they access Instagram at least once [3]. Many companies use Instagram social media as a means to attract the attention of the public to use the services or

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

products they provide. One company that uses Instagram as a tool to attract the attention of the public is an online transportation service company.

Online transportation is a new innovation in the field of m-commerce. Online transportation service is a service where customers can order transportation services such as motorcycles, cars, etc. Through the application on the mobile and the driver can directly respond to orders through the application [4]. some of the benefits provided by online transportation services such as customers and drivers can accurately know each other's location, driver and vehicle information can be clearly seen by customers, and customers can easily find online transportation anywhere to go somewhere. this makes sharing trips among urban communities increasingly popular [5]. Online transportation services in Indonesia are growing rapidly. Gojek is one of the most popular transportation services in Indonesia. Gojek was founded in 2010, managed by PT. Aplikasi Karya Anak Bangsa. Gojek emerged as a two-wheeled transportation company by telephone, along with technological developments, Gojek has now been developed into a mobile-based platform and application that provides various services such as transportation, food delivery, logistic, payment, and other services [6].

The social media platform can be used by transportation service providers to provide service information such as service announcements and updates. In addition, Instagram social media is also used to receive feedback from service users, so as to create two-way communication between service providers and service users. [1]. Many people like to use Instagram social media and use Instagram social media as a means to express their opinions and discuss various issues.

Many online transportation service providers in Indonesia make online transportation service companies compete to improve their services and innovations. to find out public sentiment towards the services provided, namely by analyzing public opinion. Sentiment analysis is a process of extracting and understanding sentiments that are defined in the text of the document [7]. Sentiment analysis is usually done to detect whether positive, negative, or neutral sentiments from the community itself [8]. With the ease in accessing Instagram and so many users, it will be so easy to get opinions that can be utilized. Data analysis aims for better decisions in taking steps forward to improve service, learn about what customers need, and increase customer satisfaction in using their services.

One of the techniques of learning from text mining for sentiment analysis is the Naive Bayes Classifier. The Naive Bayes Classifier method is considered as a potentially good method for classifying data than other classification methods in terms of accuracy and computation [9].

The goal of this paper is to classify public sentiments into positive and negative sentiments based on comments. Crawling data automatically using Exsportcomments.com. This study uses 1000 data obtained from comments on Gojek's Instagram pages. To classify sentiments from public comments about Gojek services into two sentiments, positive and negative using the Naive Bayes method. Research related to text classification using NBC has been done. On research [10] Based on the results of the analysis and testing that has been done, the Naive Bayes Classifier method gives accurate results far better than K-NN for the film review case, which is 82.43%. While the research conducted by [11] It was found that NBC can be implemented with an accuracy value of 94.00%. The research proves that this method can provide high accuracy. In addition to producing high accuracy values, Naïve Bayes classifier methods also have the lowest error rate [12].

2. Research Methods

2.1. Sentiment Analysis Process

This research is divided several steps. First, data is collected automatically from the Gojekindonesia Instagram page. Raw data is prepared for subsequent manuals. Furthermore, pre-processing consists of tokenization, stopwords, transform case, and filtering. After the data is ready in the next process is a weighting process based on the number of words that appear

in the document, the feature used is the Frequency Document Reverse Frequency (TF-IDF) term. Next is the classification of comment data using the Naïve Bayes Classifier. Evaluation is done by comparing the predicted labels with the labels calculated by the system. Review of the research that will be carried out in Figure 1.



Figure 1. Research Methods

2.2. Naive Classification Method

At this stage the data are analyzed, then models are applied according to the type of data. The model proposed in this study is Naive Bayes. This method is a machine learning method that uses probability calculation. The data is divided into two, namely training data and testing data. The Naïve Bayes Classifier mathematical model, equation (1):

$$P(C|F1\dots Fn) = \frac{(p(C)p(F1\dots Fn|C))}{(p(F1\dots Fn))}$$

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

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

Information, equation (2):

fd(ci) = number of documents belonging to the class ci|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):

$$P(wk|ci) = \frac{(f(wki,ci)+i)}{(f(ci)+|W|)}$$

Information, equation (3):

f(wki, ci) = value of occurrence of word wk on class ci f(ci) = total number of word occurrences in class ci |W| = total amount of wk

3. Result and Discussion

3.1. Dataset Retrieval

The data used in this study are Indonesian language opinions about Gojek online transportation services on the Gojek Instagram page. The dataset is in the form of comments which are public expressions of a matter through social media. The crawling process is done automatically through the export comments.com website. The amount of 1000 data that has been retrieved is entered in Excel to facilitate the data processing in the RapidMiner Tools. In Tabel 1 show some samples that has been crawled.

Table 1. Data Sample

No	Original Comment
1	Malam ka, aku ada problem nih sama aplikasi gojek tadi kan aku beli subscription
	nah terus saldo sudah ngurang tapi voucher subscription nya ga dapet
2	Semoga bertambah berkembang pt gojek
3	Gosend bobrok banget min gak bs search map gmn mau nentuin lokasi pick up nya
	bambang
4	Sekarang cari driver susah bgt, nunggunya lama bgttt
5	I LOVE GOJEK TERIMAKASIH VOUCHERNYA, POTONGAN BNYAK
	TIAP HARI DPET VOUCHER, MULAI 8K,10K,12K SEMUA BISA DIGUNAKAN
	DAN GAK HOAXS

The public comments that have been collected are classified into 2 types of sentiments, positive and negative. Positive and negative labeling is given to comments that contain sentiment. and comments that do not contain any sentiments are labeled neutral. Sentiment labeling from public comment data is done manually.

3.2. Preprocessing

This stage is the initial stage for processing text data into sentiment analysis. Preprocessing steps are useful for cleaning high-dimensional text, poor structure and noise. Data derived from Instagram comments has a high complexity because comments on Instagram use language that is not in accordance with language standards and often spelling mistakes. In this research, the first step taken is tokenizing. tokenization process works to eliminate non-letter punctuation, symbols and characters in each comment. Tokenization process can be seen in Table 2.

Table 2. Result from Tokenization					
Text before processing	Text after the process				
Terima kasih @gojekindonesia atas motor HONDA PCX nya semoga makin jaya selalu di INDONESIA	Terima kasih gojekindonesia atas motor HONDA PCX nya semoga makin jaya selalu di INDONESIA				

The process after tokenization is filtering process to delete words in stopword, in this process data that is considered irrelevant will be deleted. every word contained in the comment in the stop words is deleted. The next process is Transform cases, rapidminer will change the character (letters) to be small for all words or letters. After this stage, the entire contents of the comments will become non-capital. The final step in pre-processing is the token filter (by Length) process. At this stage, tokens are selected with a minimum length of 4 characters (letters), although some of them are stop words. If words that are less than 4 letters but are included in stopwords will still be discarded. After preprocessing, the data was scored using TF-IDF.

3.3. Validation

in the process of data Cross Validation is divided into two parts. the first part is a subset of training and the second part is a subset of tests that will be used to test performance. By using K-fold cross-validation, the data will be automatically partitioned, the original dataset becoming subsample K. One subsample is stored as validation data to test the classification model and the remaining subsample is used for training data in Subsample K. The validation process is repeated continuously for K with each k sample used once as test data. Then to produce an estimate of performance of results K from the fold then averaged. To get the best model, the dataset distribution process is done using 10 times Cross-Validations.

3.4. Evaluation

The results of implementing the Naive Bayes model will get a rules. The resulting rule will be used as the basis for predicting the value to be performed. Previously the rule had to be evaluated and validated so that it was known how accurate the prediction results would be. At this stage the evaluation process uses the confusion matrix method.

The confusion matrix model will form a matrix consisting of accuracy, recall, and precision. Table 3 shows the result of the confusion matrix using the Naive Bayes classifier method.

No		true Positif	True Negatif	Class Precission
$\frac{1}{2}$	Prediction Positif Prediction Negatif Class recall	$430 \\ 76 \\ 84.98\%$	95 299 75.89%	81.90% 79.73%

Table 3. Accuracy Analysis Table. Accuracy: 81.00%+/- 3.68% (micro average: 81.00%)

Naive Bayes test accuracy results for sentiment analysis comments on Instagram social media results obtained an accuracy of 81.00% while for the AUC value of 0.747 which can be seen in Figure 2. Based on the results of the accuracy shows that the Naive Bayes Classifier method can produce high accuracy in the process of classifying public comments on Instagram social media. this shows that the Naïve Bayes Classifier method is an appropriate algorithm used to analyze public opinion sentiments towards a service. With analysis using public opinion, online transportation service providers can find out how public opinion is about their services and can be used in taking steps to improve service quality.

1641 (2020) 012093 doi:10.1088/1742-6596/1641/1/012093



Figure 2. Research Methods

4. Conclusion

Based on the results of the analysis conducted on the sentiment of online transportation services using the Naive Bayes method, researchers can draw some conclusions that the Naive Bayes Method is quite good in classifying data mining or text mining. this is because the algorithm can produce a fairly high accuracy value of 81.00%, which means that all the comments on the Instagram page with the NBC method can be accurately classified whether the comment is negative or positive. The results of this study can be used as recommendations to improve the performance of online transportation. For future suggestions, this research is expected to be developed so that the accuracy results can be improved, recommendations can be enriched by using more datasets and using dictionaries that are always updated following with language growth, and also the method used can be done with other algorithms such as the support vector machine (SVM) method to improve the accuracy.

References

- Giancristofaro G T and Panangadan A, 2016 Predicting Sentiment toward Transportation in Social Media using Visual and Textual Features IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC p. 2113–2118.
- [2] Hapsari Y Fikri Hidayattullah M Dairoh D and Khambali M, 2018 Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivry) J. Inform. J. Pengemb. IT 3, 2 p. 233–236.
- [3] Mailanto A, 2016, Pengguna Instagram di Indonesia Terbanyak, Mencapai 89%: Okezone techno.
- [4] Wallsten S, 2015 The Competitive Effects of the Sharing Economy: How is Uber Changing Taxis? Technol. Policy Inst. June p. 1–22.
- [5] Silalahi S L B Handayani P W and Munajat Q, 2017 Service Quality Analysis for Online Transportation Services: Case Study of GO-JEK Procedia Comput. Sci. 124 p. 487–495.
- [6] Gojek, Tentang Kami Gojek. [Online]. Available: https://www.gojek.com/about/. [Accessed: 27-Jun-2020].
- [7] Shivaprasad T K and Shetty J, 2017 Sentiment analysis of product reviews: A review Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017 Icicct p. 298–303.
- [8] Rosenthal S Farra N and Nakov P, 2018 SemEval-2017 Task 4: Sentiment Analysis in Twitter p. 502–518.
- [9] Xhemali D Hinde C J and Stone R G, 2009 IJCSI IJCSI Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages IJCSI Int. J. Comput. Sci. Issues 4, 1 p. 16–23.
- [10] Dey L Chakraborty S Biswas A Bose B and Tiwari S, 2016 Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier Int. J. Inf. Eng. Electron. Bus. 8, 4 p. 54–62.
- [11] Kristiyanti D A Umam A H Wahyudi M Amin R and Marlinda L, 2019 Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018 Citsm p. 1–6.
- [12] Elghazaly T Mahmoud A and Hefny H A, 2016 Political sentiment analysis using twitter data ACM Int. Conf. Proceeding Ser. 22-23-Marc.