



PAPER • OPEN ACCESS

Mapping online transportation service quality and multiclass classification problem solving priorities

To cite this article: Andry Alamsyah and Imam Rachmadiansyah 2018 *J. Phys.: Conf. Ser.* **971** 012021

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

You may also like

- [Development of a solid waste service level index](#)
A A Rumakat, I Juwana and S Ainun
- [The Effect of Supply Disruptions on Customer Service Levels: a Case for Delivering Fertilizer Products using Maritime Transportation](#)
N Siswanto, U Kurniawati, S E Wiratno et al.
- [Building satisfaction and loyalty of student users ojek online through the use of it and quality of service in tangerang city](#)
Siti Asriah Immawati and Abdul Rauf



The Electrochemical Society
Advancing solid state & electrochemical science & technology

247th ECS Meeting
Montréal, Canada
May 18-22, 2025
Palais des Congrès de Montréal

Showcase your science!

Abstracts due December 6th

Mapping online transportation service quality and multiclass classification problem solving priorities

Andry Alamsyah¹, Imam Rachmadiansyah²

School of Economics and Business, Telkom University, Bandung, Indonesia

¹ andrya@telkomuniversity.ac.id, ²imamrachmadian@gmail.com

Abstract. Online transportation service is known for its accessibility, transparency, and tariff affordability. These points make online transportation have advantages over the existing conventional transportation service. Online transportation service is an example of disruptive technology that change the relationship between customers and companies. In Indonesia, there are high competition among online transportation provider, hence the companies must maintain and monitor their service level. To understand their position, we apply both sentiment analysis and multiclass classification to understand customer opinions. From negative sentiments, we can identify problems and establish problem-solving priorities. As a case study, we use the most popular online transportation provider in Indonesia: *Gojek* and *Grab*. Since many customers are actively give compliment and complain about company's service level on *Twitter*, therefore we collect 61,721 tweets in *Bahasa* during one month observations. We apply *Naive Bayes* and *Support Vector Machine* methods to see which model perform best for our data. The result reveal *Gojek* has better service quality with 19.76% positive and 80.23% negative sentiments than *Grab* with 9.2% positive and 90.8% negative. The *Gojek* highest problem-solving priority is regarding application problems, while *Grab* is about unusable promos. The overall result shows general problems of both case study are related to accessibility dimension which indicate lack of capability to provide good digital access to the end users.

1. Introduction

Online transportation service is known for its accessibility, transparency, and tariff affordability. These points make online transportation have advantages over the existing conventional transportation service. Online transportation is an example of disruptive technology that makes customers have better bargaining position than the companies. It changes the way how customers and companies interact towards conclude transactions [1]. Co-occurrence with online transportation increasing popularity, we can easily find those topics discussed on social media, such as *Twitter*, *Facebook*, and *Instagram*.

Gojek and *Grab* are the most popular online transportation company in Indonesia. They have over 500.000 users in 2017, based on playstore unique users. Both provide variety of service, such as delivery, financial, and public transportation. Customers are actively give compliment and complain about both company's service level on social media. Those data are extracted, analyze, and finally turn into knowledge that can help to improve their service level^[1]. For this paper, we investigate which social media are the most popular to hold the discussion about both company. From table 1, we see that on average number of discussion on *Twitter* is higher than others, thus we pick *Twitter* as representative media to get the data. The previous approach of identifying problem priorities mostly based on interviews or questionnaires, which basically more time-consuming effort compared to collecting online opinions.

Table 1. Average Number of Discussion in Each Social Media during Observation Time

Company	Social Media		
	Twitter	Facebook	Instagram
Gojek	761/week	140/week	73/week
Grab	402/week	105/week	239/week



We apply transportation service quality [2] measurement to map each dimension proportion to the global domain problem and sentiment analysis [3] method to classify people's feeling to phenomenon. The purpose of those methods is to identify service satisfaction, to improve service quality, and to increase accuracy of decision support making. To classify service problems, we use multiclass classification method^[4]. The most used multiclass classification methods are *Naive Bayes* (NB) and *Support Vector Machine* (SVM). Both classification methods are applied in this research to evaluate multiclass classification performance.

Our global objectives are first to map problems of online transportation companies and second to set problem-solving priorities. We apply both sentiment analysis and multiclass classification to understand customer thought of each dimension or in other word dimension performance. As the result, we can identify problems from set of negative opinions, then build problem solving priorities based on that information. The problem-solving priorities help companies to decide what problems should be dealt first. Problems-solving priorities arranged based on amount of negative opinions in each dimension.

2. Theoretical Background

Service quality is a conceptual model to assess services based on its performance's perception. There are 5 general service quality dimensions [4]: *tangibles*, *reliability*, *responsiveness*, *assurance*, and *empathy*. However, in this research, since our object is about online transportation, thus, we apply transportation service quality [2] to identify problems classification. Each opinion has contextual similarities to others classified into the same class. Multiclass classification method help to determines customer satisfaction and opinion in each class. Thus, it makes us easier to achieve research objectives. The transportation service quality dimension shown in table 2:

Table 2. Transportation Service Quality Dimensions

<i>Service Quality Dimension</i>	<i>Description</i>
<i>Availability</i>	The availability of transportation services everytime, everywhere, and in any condition.
<i>Accessibility</i>	The application service ease of use in certain time, condition, and area.
<i>Information</i>	Information availability, well informed customer such as travel fee before making the journey
<i>Time</i>	The detail information about departure time, arrival time, and travel duration.
<i>Customer Service</i>	The capabilities of company to handle complaints, suggestions; The capabilities of company to response customer inquiry in reasonable time; The information about promotional activities.
<i>Comfort</i>	Company effort to provide comfort to customer, such as all-weather protection, vehicles hygiene, and driving style
<i>Safety</i>	Company effort to provide safety and security, such as driver preparation, driving attributes, route knowledge, traffic condition awareness
<i>Environment</i>	Vehicle noise, and vehicles contribution in gas emission

Multiclass classification is a classification method which has more than two target classes. The target result shows classes of tweets contain specific issue related to dimensions in Table 2. Sentiment analysis is a method to analyze the user opinion or sentiment by classification. Sentiment analysis performed automatically to classify opinion based on words or phrase that formed during learning procedure. The automated sentiment analysis is important to reduce human error and processing time to identify, extract and summarize text [3]. The previous research [5] address many advantage of applying multiclass classification problem in customer relationship management area, such as faster identification, considerably good accuracy, and support real-time customer complaints. In the face of unstructured nature of online data, text mining based multiclass classification methods is preferable. The learning process become advantage in identifying future complains automatically.

To perform sentiment analysis methods, we need preprocessing procedure which consist of the several following processes [5]. First is data cleaning to remove useless items for the study, such as emoji and noise words. Second is tokenization process to separate the sentences into words, phrases, symbols called tokens. Third is filtering process to remove punctuation, hyphens, characters and high frequency intermediate words, such as '*dari*', '*untuk*', '*belakang*', '*atas*' and '*depan*'. In addition, this step also

transform size letters into lowercase. Fourth is stemming process to remove incremental word or turn word into a basic word [6], like 'bepergian' to 'pergi'. Table 3 shows examples of each process in preprocessing procedure

Tabel 3. Preprocessing Example

<i>Raw Data</i>	Kenapa ya susah banget dapet gocar padahal armadanya banyak di sekitarku, @Gojekindonesia? So saaad :(
<i>Tokenization</i>	Kenapa ya susah banget dapet gocar padahal armadanya banyak di sekitarku, @Gojekindonesia? So saaad :(
<i>Filtering</i>	susah banget dapet gocar armada so saad
<i>Stemming</i>	susah sekali dapat gocar armada sangat sedih

To measure how important a word to a document we use TF-IDF [7] procedure. The main function of weighting process is to enhance term retrieval effectiveness. There are 2 effective factors for effective retrieval. First, terms likely to be relevant to the user's need must be retrieved. Second, terms likely to extraneous must rejected [8]. This process is done before we can classify user sentiments. First step is to give weight to each word in document using with TF-IDF formula.

$$W_{d,t} = TF_{d,t} \quad (1)$$

$tf_{d,t}$ is the term occurrences of the word t in documents d . The calculation of $tf_{d,t}$ is known to increase the recall value but does not increase the precision value. To deal with this problems terms with high frequency are usually removed.

$$idf_t = \log \frac{N}{df_t} \quad (2)$$

N is the number of words contained in the text and df is the number of texts containing the word t . df_t is the number of text that contain the term t . It means idf_t is the inverse of term t in number that existed in the documents. Idf_t represent term specificity which expected to improve precision. So, by using $W_{d,t}$ with $tf_{d,t}$ and idf_t together, the method performs better, because it addresses both aspects of recall and precision. The better recall and precision give us better overall performance. Then TF-IDF will be described as an equation.

$$W_{t,d} = tf_{t,d} \times idf_t. \quad (3)$$

3. Research Methodology

The overall research workflow can be seen in Figure 1. We set the research objective to map service quality and problem classification to determine each opinion class category. The following step is finding the best method to perform the research. In data collection process, we decide to use *Twitter* for the reason explained in Introduction chapter. Our focus to choose the best sentiment analysis methods, between NB and SVM methods in term of the prediction accuracy multiclass classification problem regarding service quality. Both methods have similar performance in computation time, thus we can ignore this aspect.

We collect all *Twitter* data concerning @gojekindonesia which represent *Gojek* and @GrabID which represent *Grab* during the data collection time. Following the data collection process, we do data preprocessing procedure, which consist removal irrelevant tweet; noise removal by tokenizing, stemming, filtering; weighting the words with TF-IDF. After preprocessing procedure, we build training and testing data before applying NB and SVM methods.

Model evaluation are done by measuring recall, precision, f-measure, and kappa value of each methods. The last step is to analyze the result and determine the best way to achieve research goals to establish problem-solving priorities.

Data collecting process is done for the period of one month in March 2017. We collect in total 61,721 tweets, consist of 49,947 tweets associated with @Gojekindonesia and 11,774 tweets associated with @GrabID. After preprocessing step, we obtain 1,463 tweets for @Gojekindonesia and 1,098 tweets for @GrabID. We classify those tweets into 8 different dimensions of transportation service quality. The examples of tweet classification based on transportation service quality dimension is in Table 4. The

example of sentiment classification tweets is in Table 5. The example of preprocessing to of raw data into clean data is in Table 6.

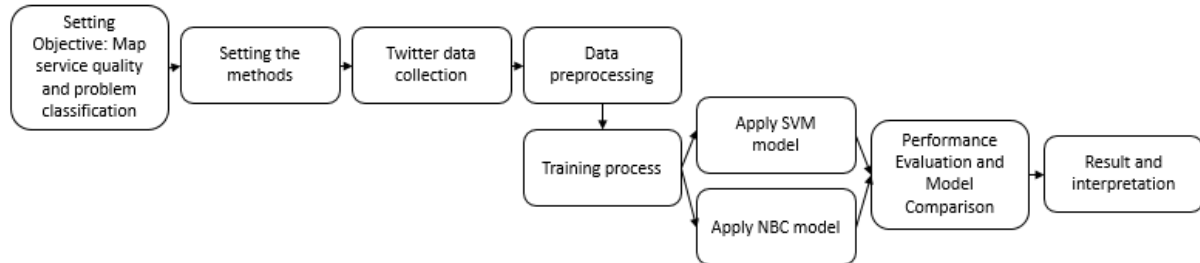


Figure 1. Research Workflow

Table 4. Example of Tweet Classification based on Transportation Service Quality Dimension

Example Tweets	Dimension
Susahnya Go-jek pada jam jam mau berangkat :"	Availability
@Go-jekindonesia min saya gagal verifikasi kode akun Go-jek	Accessibility
Karena go-mart cukup mahal	Information
Pesen Go-jek abangnya lama banget huhu padahal ga terlalu jauh dari lokasi kasian ibu gue nungguin :(Time
Thanks to gopay gratisan dari Go-jek	Customer Service
Pertama kali ini Go-jek ada yg cakep plus wangi....sukaaa	Comfort
Termengerikan adalah naik Go-jek yg super aneh kayak baru belajar naik motor takut	Safety
Go-jek parkir semprawut gak jelas sampai di trotoar	Environment

Table 5. Example of Tweet Classification by Sentiment

Tweet	Sentiment
Terima kasih bapak Go-jek baik hati yang mau antri lama demi saya ngidam ini	Positive
Go-jek ini Tarifnya gak masuk akal	Negative

Table 6. Preprocessing Example

Raw Data	Clean Data
App Gojek giliran lagi dibutuhin error hhhhh sebhul.	Aplikasi error sebal

We apply TF-IDF procedure to weight the words using *Rapidminer* application. The result determines each word into classes and omit words to increase the multiclass classification performance, for example, 'aplikasi' has 0,134 weight in *accessibility* dimension, while 'bayar' has 0 weight in *accessibility* and *information* dimension, so 'bayar' should be deleted to increase the performance of multiclass classification.

Problem-solving priorities applied to see what problem must be dealt first in order. It intended to allocate companies' resource efficiently. We use multiclass classification and sentiment analysis to detect problems by each class negative opinion quantities. The largest negative quantities classes assigned to be the highest prioritized dimension, while the largest negative quantities of words in classes assigned to be the highest prioritized problems. For example, *availability* dimension attains 142 negative sentiment, *accessibility* dimension attains 210 negative sentiment, *information* dimension attains 134 negative sentiment and *time* dimension attains 78 negative sentiment. So, the top 3 prioritized dimension are *accessibility*, *availability* and *information*. Moreover, as example, *time* dimension has 'late pickup' with 130 times mentioned while 'slow travel duration' with 145 times mentioned. The highest prioritized problem should be 'slow travel duration'.

4. Result and Analysis

NB classification methods has good accuracy for both object. The value of both kappa is around 0.80 which predicated a nearly perfect agreement of its classification model, while precision and recall value

reach between 80-95%, with F-measure 91.46% and 89.36% as harmonization of precision and recall value indicating the classification process is going well. NB determine words naively.

SVM classification model has higher accuracy than NB as shown in Table 7 and Table 8. It shows that SVM kappa and F-measure for is better than NB. As the result, for this study SVM is more suitable method to do the classification task. SVM model consider each word class before determining the word included in specific class.

Table 7. Naive Bayes Text Classification Performance

Company	Accuracy	Kappa	Precision	Recall	F-measure
<i>Gojek</i>	93,92%	0.823	88.64%	94.47%	91,46%
<i>Grab</i>	96,58%	0.786	90.77%	88.00%	89,36%

Table 8. Support Vector Machine Text Classification Performance

Company	Accuracy	Kappa	Precision	Recall	F-measure
<i>Gojek</i>	97,26%	0.910	97,66%	93,66%	95,61%
<i>Grab</i>	97,26%	0.809	98.54%	85.00%	91,27%

Both *GoJek* and *Grab* are actively engaged in interaction with customers through social media. These interactions crawled and mapped to determine customer perceived quality. On Table 9, we can see the sentiment proportion of both companies. *Gojek* obtain better opinion for their service than *Grab*, with 19.76% against 9.2% positive.

Table 9. The Tweet Sentiment Proportion

Company	Total Tweet	Positive	Negative
<i>Grab</i>	1462	134 (9.2%)	1328 (90.8%)
<i>Gojek</i>	1098	217 (19.76%)	881 (80.23%)

Table 10. *Gojek* Dimension Proportion

Sentiment	Availability	Accessibility	Information	Time	Customer Services	Comfort	Safety	Environment
Positive	25.3%	1.3%	17.9%	9.6%	18.8%	21.1%	5.5%	0.5%
Negative	24.1%	34.7%	13.6%	8.4%	6.7%	7%	4.9%	0.6%

The top-3 negative dimension in *Gojek* are *accessibility*, *availability* and *information*. *Gojek* problems in *accessibility* dimension are application problems, where 273 times mentioned about ‘*aplikasi*’, ‘*error*’, ‘*sistem*’, and ‘*server*’; difficult to logging in or verification problem, where 182 times mentioned about ‘*akun*’, ‘*log*’, ‘*masuk*’, ‘*keluar*’, ‘*password*’, ‘*kode*’ and ‘*verifikasi*’; ordering transportation service, where 72 times mentioned about ‘*pesan*’ and ‘*invalid*’. Problems in *availability* dimension are driver ignorance and difficult to get transportation service, where 244 times mentioned about ‘*ambil*’, ‘*dapat*’, ‘*susah*’, ‘*pesan*’, ‘*tiada*’, ‘*lama*’, ‘*jemput*’, and ‘*invalid*’; unilateral cancellation or driver’s cancellation request, where 30 times mentioned about ‘*batal*’ and ‘*tolak*’. Problem in *information* dimension are the expensive tariff mentioned 77 times about ‘*mahal*’, ‘*bayar*’ and ‘*tarif*’; gopay truncated due unilateral cancellation from driver mentioned 33 times about ‘*batal*’, ‘*potong*’, and ‘*ambil*’.

We use the result above to set *Gojek* problem-solving priorities. There are 3 problems occurred to be the most mentioned from 2 different dimension. So, the table 11 below are the problems that should be dealt first.

Table 11. *Gojek* Problem-Solving Priorities

No	Problems	Mentioned	Solution
1	Application problems	273 Times	<ul style="list-style-type: none"> • Improve application performance and system optimization • Conduct research about most used phone in the market and user habit
2	Driver ignorance and difficult to get transportation service	244 Times	<ul style="list-style-type: none"> • Create reward and punishment system for the driver

3	Difficult to logging in or verification problem	182 Times	<ul style="list-style-type: none"> • Create auto login from same device • Integrate the application with social media service to get alternatives code access verification • Send login confirmation to email
---	---	-----------	--

Table 12. *Grab* Dimension Proportion

Sentiment	Availability	Accessibility	Information	Time	Customer Services	Comfort	Safety	Environment
Positive	1.5%	3.8%	21.4%	6.6%	13.3%	45.3%	7.4%	0.7%
Negative	10.3%	23%	18.4%	3.4%	33.7%	7.1%	3.6%	0.5%

The top-3 negative dimension in *Grab* are *customer service*, *accessibility* and *information*. The problems in *customer service* dimension are unusable promos, where 375 times mentioned about ‘*promo*’, ‘*diskonseharian*’ and ‘*kode*’; no response from admin, where 134 times mentioned about ‘*admin*’ and ‘*komplain*’; no *grabpay* bonus received, where 364 times mentioned about ‘*bonus*’, ‘*grabpay*’, ‘*top*’, ‘*isi*’, ‘*masuk*’, and ‘*gratis*’. The problems in *accessibility* dimension are unusable *grabpay* credit, where 341 times mentioned about ‘*masuk*’, ‘*grabpay*’, ‘*isi*’, ‘*top*’, ‘*bayar*’, and ‘*lama*’; application error, where 107 times mentioned about ‘*aplikasi*’, ‘*error*’, and ‘*update*’; difficult to order and cancel, where 64 times mentioned about ‘*dapat*’, ‘*batal*’, ‘*invalid*’ dan ‘*pesan*’. The problems in *information* dimension are *grabpay* value reduction due improper cancellation, where 136 times mentioned about ‘*grabpay*’, ‘*potong*’, ‘*batal*’, ‘*pesan*’, and ‘*jemput*’; high tariff, where 55 times mentioned about ‘*word*’, ‘*mahal*’, ‘*harga*’, and ‘*tarif*’.

We use the result above to set *Grab* problem solving priorities. There are 3 problems occurred to be the most mentioned from 2 different dimension. So, the table 14 below are the problem that should be dealt first.

Table 13. *Grab* Problem-Solving Priorities

No	Problems	Mentioned	Solution
1	Unusable promos	375 Times	<ul style="list-style-type: none"> • Arrange the promo frequency wisely • Make sure the promos work exactly as term made • Inform it to customer about the time and usage limit
2	No <i>grabpay</i> bonus received	364 Times	<ul style="list-style-type: none"> • Improve server capacity to build faster response and prevent overload request • Anticipate current and prospective customer before work the bonus plan.
3	<i>Grabpay</i> credit refill process taking long time	341 Times	<ul style="list-style-type: none"> • Improve server capacity, faster response, and prevent overload request. • Co-operate with reliable company, to prevent any obstacle company, such as reload <i>grabpay</i> balance delayed because bad access

Conclusion and Recommendation

Both multiclass classification, NB and SVM are able to classify the datasets well, but SVM performance is better than NB measured by kappa, precision, recall, f-measure and accuracy. So, it is better to use SVM as multiclass classification model for Indonesian *Twitter* dataset.

Gojek sentiment rating with 19.76% positive and 80.23% negative are better than *Grab* with 9.2% positive and 90.8% negative. The proportion of dimensions in *Gojek* are *accessibility*, *availability*, *information*, *comfort*, *customer service*, *time*, *safety* and *environment*. While *Grab* are *customer service*, *accessibility*, *information*, *comfort*, *availability*, *safety*, *time* and *environment*.

The main problems of *Gojek* are application problems, difficult to get transportation service, and difficult to logging in. The main problem of *Grab* are unusable promos, no response from admin, and no *grabpay* bonus received. Both companies suggested that the prioritized top-3 problems should be dealt as soon as possible, before negative opinion among customer get out of control. Overall, arising problems of both online transportation companies are related to *accessibility* which indicate lack of capability to provide good digital access to the society. Hence, we conclude sentiment analysis and multiclass classification are capable to establish problem-solving priorities.

References

- [1] Wilson, T. D. 2013. The e-book phenomenon: a disruptive technology. *Proceeding of the International conference Publishing – trends and contents*. UDC: 655:[002:004](485)=111.
- [2] Poliakova, A. 2010. Application of the Customer Satisfaction Index(CSI) to transport services. *eJournal of Slovak University of Technology in Bratislava*. **5** 208-215.
- [3] Liu, B. 2012. *Sentiment analysis and opinion mining*. (California: Morgan & Claypool Publisher).
- [4] Parasuraman, A., Berry, L. L., and Zeithaml, A. 1991. Perceived Service Quality as a Customer-Based Performance Measure: An Empirical Examination of Organizational Barriers Using an Extended Service Quality Model. *Human Resource Management*. **30** 335-364.
- [5] Arusada, M. D. N., Putri, N. A. S., and Alamsyah, A. 2017. Training data optimization strategy for multiclass text classification. *The 5th International Conferences on Information and Communication Technology*. ISBN: 978-1-5090-4911-0.
- [6] Han, J., Kamber, M., and Pei, J. 2012 . *Data Mining Concepts and Technique*. 3rd Edition (Waltham: Elsevier).
- [7] Tokunaga, T. and Iwayama, M. 1994. Text categorization based on weighted inverse document frequency. *Technical Report of Departemen of Computer Science Tokyo Institute of Techonolgy*. ISSN 0918-2802
- [8] Salton, G. and Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management* **24** 513-523.