

PAPER • OPEN ACCESS

The sentiment analysis of Indonesia commuter line using machine learning based on twitter data

To cite this article: M O Pratama *et al* 2019 *J. Phys.: Conf. Ser.* **1193** 012029

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

You may also like

- [Positive or Negative Thermal Expansion in Anharmonic Vibrational Systems in EXAFS study](#)
Dong-Seok Yang and Jay Min Lee
- [Coupling between feedback loops in autoregulatory networks affects bistability range, open-loop gain and switching times](#)
Abhinav Tiwari and Oleg A Igoshin
- [Convergence to pure steady states of linear quantum systems](#)
Muhammad F Emzir, Ian R Petersen and Matthew J Woolley



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD



Submit your abstract



The sentiment analysis of Indonesia commuter line using machine learning based on twitter data

M O Pratama¹, W Satyawan¹, R Jannati¹, B Pamungkas¹, Raspiani¹, M E Syahputra¹ and I Neforawati²

¹Artificial Intelligence Department, Premier Optima Sattiga, Jakarta, Indonesia

²PoliteknikNegeri Jakarta, Depok, Indonesia

Author's Email: octav@premieroptima.com

Abstract. This paper presents The Sentiment Analysis of Indonesia Commuter Line (KRL) using Twitter data. Some of people are expressing their complaints about public transportation, especially the satisfaction of Commuter Line service on Twitter. We collected the data from Twitter to classify many public opinions into positive or negative label then Machine Learning model will be used as classification model to classify positive or negative opinion. Multinomial Naive Bayes (MNB), Random Forest (RF) and Support Vector Machine (SVM) is used as a model, then we measure all of model performances. And the result is Support Vector Machine produced has the highest accuracy of all with 85%. Net Sentiment Score (NSS) is also computed in order to determine whether KRL features meet customer's satisfaction.

Keyword: machine learning, Naive Bayes, Random Forest, sentiment analysis, support vector machine.

1. Introduction

There are a lot of mass transportation in Indonesia, especially in Jakarta, starting from public transportation, Trans Jakarta and Commuter Line Electric Train (KRL). The last mentioned, KRL, has been widely used by people to connect one location to destination. There are several cities that connected by KRL like Jakarta itself, Bogor, Bekasi, Tangerang, and others. Based on statistical data, the number of passengers throughout 2017 to June 2018, it reaches 481.314 passengers. Because KRL is used by many people, so the level of customer satisfaction of using KRL should be known in order to improve the performance of KRL. There are previous researches to measure the level of customer satisfaction in public transportation like Sonia et. al [1] tried to analyse the satisfaction of online transportation using Twitter data. Vidya et. al also tried to measure level of satisfaction of mobile phone user using Twitter data [2]. Based on both researches, Twitter data can be used as data source. Twitter is one of the most popular micro blogging social media in Indonesia. People usually express their complaints or performance on Twitter. Considering the results of people's thoughts about their writing on Twitter, it is quite simple, this can be a good source for sentiment analysis [3]. Zang et. al [4] shows the data which was obtained from social media and it can be used to measure the satisfaction with customer insight. In



addition, Generally, Yu et. al [3] shows that sentiment on social media has a greater impact in a company than conventional media. More specifically by Twitter which might have positive results.

Sentiment Analysis (SA) is a branch of text data tabulation which focuses in text data analysing. Many of SA model are using Twitter's data to collect sentiments for classification process. In this study, a tweet is classified into two classes (positive or negative), which have one aim, that is to analyse the sentiments performances of commuter line KRL transportation services in Indonesia. There are many sentiment analysis researches in public spaces such as Ye at. al [5] using machine learning to do sentiment classification to reviews travel destinations. Sonia et. al [1] are using sentiment analysis to show which one of online transportation between Grab and GO-JEK that is more satisfying for customer. Moreover, they used NSS calculation as a parameter of customer satisfaction between Grab and GO-JEK. In that research, we are also adapting method from Sonia [1] that has been proved. NSS calculation has the same result as traditional customer. The purpose of this research is to measure how much customer satisfaction of commuter line KRL transportation services in Indonesia. By dividing into two positive and negative classes, we calculate these two sentiments to measure the satisfaction of KRL customer using the Net Sentiment Score (NSS) towards commuter line KRL. NSS is a method to measure customer satisfaction from social media data [6]. The result of this study can be a recommendation for Commuter Line Indonesia to obtain the feedback in order to improve customer satisfaction.

2. Related Work

There are several previous studies on sentiment analysis using various algorithms in machine learning. Sonia et. al [1] used machine learning algorithms such as Support Vector Machine (SVM), Naive Bayes (NB) and Decision Tree (DT) to measure the level of satisfaction of online transportation users in Indonesia, namely Grab and GO-JEK. On their experiment, the level of Grab customer satisfaction is higher than GO-JEK with 72.97% accuracy using the support vector. The classification result then forwarded into NSS calculations which can be decided as the most satisfying online transportation. Della et. al [6] analysed the sentiment towards DKI Jakarta gubernatorial election in 2017 by using the Multinomial Naive Bayes (MNB) and Support Vector Machine (SVM) algorithms. The results showed that both of two algorithms produce 66.99% and 69.11%. Ika et. al [7] did a research that investigating the benefits of the use of hashtags to determine the sentiment of analysis in the political domain using the Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression, Random Forest (RF) and SentiHT. The results showed that the sentiment classification uses SentiHT has a very good accuracy of more than 95%. In fact, SentiHT outperformed the unigram feature when combined with Naïve Bayes (NB), SVM or Logistic Regression. Based on explanation, most of the empirical studies use text data processing technology that combined with machine learning algorithms to process analytical sentiments. The process of processing text data with machine learning techniques can increase the predictive accuracy.

In this study, we analyse the performance sentiments of the Commuter Line KRL transportation service by using the three available algorithms in machine learning, there are Multinomial Naive Bayes (MNB), Random Forest (RF) and Support Vector Machine (SVM). Then, the calculation of Net Sentiment Score (NSS) based on positive and negative tweets is also done. The Net Sentiment Score computes a ratio of positive and negative mentions of a topic [1]. The value of NSS could be 100 or -100. A Net Sentiment of 100 means that all mentions of the topic are positive, and -100 means all mentions of the topic are negative.

3. Methods

In this section, researchers present the methods that the researchers conducted in this Sentiment Analysis study. There are 4 main methods in this study, which are Dataset, Text Processing, Classification, and NSS Calculation.

3.1. Dataset

The dataset was collected by crawling based on request “commuter line”. It was collected from Indonesian language on June 28, 2017 until June 28, 2018. The dataset also equipped with tweets and its sentiment label. We classified the dataset in three categories, which are negative (-1), positive (1), and neutral, irrelevant, or can’t be understood (0). Before classified the dataset, there are some rules for each tweet as can be seen below:

- A. If the tweet is expressing a dislike statement towards Commuter line as a whole including, services, systems, staffs, machinist, etc. As well as sarcasm, critics or complaining tweets towards Commuter line too. Then it is classified as negative(0).
 - “Nggak capek apa jadi tuli dan nggak profesional? Kalau emang pada nggak bias kerja mending dipecat2in aja ganti sama yg bias kerja. @CommuterLine”.
 - “@Commuter Line kiranya berkenan untuk memperbaiki jadwal perjalanan ke Bekasi dan Cikarang. Sangat sering harus menunggu lama. Kereta Jarak Jauh harusnya yang menyesuaikan, karena bias kejar jadwal di luar jalur tersebut. Selain itu tidak ada penumpang yang berdiri”.
- B. If the tweet is praising towards the Commuter line as a whole then it is classified as positive(1).
 - “Hari senin walau ada update sistem di @CommuterLine, tapi persiapan petugas sangat baik, siapkan uang yang pas supaya mengurangi antrian, stasiun bogor padat, stasiun tanjung barat lancar.”.
 - “Alhamdulillah. Terimakasih @CommuterLine, Pak @BudiKaryaS, @perkeretaapian. Terimakasih juga atas kerja keras Tim mbak @meilia_w, @JalitaTangerang, mas @_ibaadurrahmaan, mbak @Iputrii1, mas @aditmailtweet, @arashivhi Semoga jadi solusi yg baik bagi komuter KRL Tangerang-Duri”.

We used 3 annotators for considering tweets which are consists of computer scientist, mathematician, and social scientist. Kappa value [8] is used to test two or more reliable annotators. This is important enough to ensure that the data used in the study has an appropriate representation. The annotators will sort on the entire 2,160 tweets out manually. Sorting out is done based on the rules described above. After sorting out the tweets, we create a new dataset that compares all of the annotators’ sentiments. The new dataset will be calculated with the kappa value for each of the annotators. After obtaining the kappa value for each annotator, then we get the average to see whether the kappa considered as good, moderate, or poor. As mentioned before, this study used two classes which are positive (1) with 147 tweets and negative (0) with 648 tweets. The rest is considered as neutral or irrelevant which don’t use in this research. As mentioned before, this study used four annotators, I, II, III, and IV. First of all, all of the annotators manually sorting out the dataset. Next, researchers created new dataset by comparing sentiments result between each annotator. For example, in line 1 of dataset, comparing which sentiment that most obtained among all annotators. If positive, the label is positive (1). And so on until the last line of dataset. Then, calculate the kappa value between the new dataset and the label dataset by the four annotators to find the average. From the table 1, it can be seen that dataset considered as good based on kappa value.

Table 1. Kappa Value

Annotators	Kappa	Strength
Dataset and I	0,659	good
Dataset and II	0,715	good
Dataset and III	0,553	moderate

3.2. Pre-Processing

Before dataset can be proceed into the model, data should be proceeded into pre-processing step in advance to remove irrelevant words which are not represent any sentiments such as punctuations, symbols, numbers, and links. We also do stemming [9], normalization, and removing stop words too.

1. Removing Punctuations

All punctuation marks such as commas, dots, exclamation points, etc. must be deleted. For Example:

- *“Menyiapkan uang tunai sesuai tariff tiket kertas sih saya sanggup, kalo menyiapkan waktu lebih saya ga sanggup om.”* → *“Menyiapkan uang tunai sesuai tariff tiket kertas sih saya sanggup kalo menyiapkan waktu lebih saya ga sanggup om”*.

2. Removing symbols and numbers

All of numbers [0987654321] and symbols such as [@\$%!] Should be deleted. For Example:

- *“Di Stasiun Kranji e-money ga bisa @CurhatKRL @CommuterLine jalur bekasi”* → *“Di Stasiun Kranji e money ga bias Curhat KRL CommuterLine jalur bekasi.”*

3. Removing link

All tweets have link within its tweet, must be deleted. For Example:

- *“Pengguna KRL Commuter Line Harus Berangkat Lebih Pagi Hari Ini! <http://detik.id/VxFwYQ>”* → *“Pengguna KRL Commuter Line Harus Berangkat Lebih Pagi Hari Ini”*.

4. Stemming

All words in KBBD are not yet standard, they will be converted into standard words. For Example:

- *“Bisa memberikan solusi yg lebih baik”* → *“bias beri solusi yg lebih baik”*

5. Normalization

All abbreviation words will be changed to normal words. Researcher made a dictionary which contains abbreviated words and original words. For example, “bgt:banget”, “tdk:tidak“, and others. For Example:

- *“Mohon maaf bgt nih ngetwit mulu, tapi yg begini tuh mengganggu bgt dan gak aman”* → *“mohon maaf banget nih ngetwit mulu, tapi yang begini tuh mengganggu banget dan gak aman”*.

6. Stop words

This study used to remove any words which not contains any sentiments within the tweet. For instance, “yang”, “ada”, “ikut”, “maka”, etc. All of words that considered as stopwords will be deleted. For example:

- *“Apakah tombol darurat yang dipasang di setiap gerbong itu”* → *“tombol darurat dipasang gerbong”*.

3.3. Classification

We use three kinds of Machine Learning model, Multinomial Naïve Bayes, Support Vector Machine and Random Forest. Multinomial Naive Bayes is a branch of The Naive Bayes Algorithm. from several Naïve Bayes models, Multinomial Naïve Bayes (MNB) is one of the best model to solve text classification problems because of its ability that can get information without seeing the sentence many times in a document[7]. The advantage of using Naive Bayes is that this method only requires a small amount of data training to estimate the parameters in classification. Combining probability distribution of P with fraction of documents belonging to each class. Support Vector Machine (SVM) is one classifier algorithm that often used in both classification and regression. SVM is also used in the Supervised Learning task. How SVM works is by finding a hyperplane on the input, the best hyperplane is the one which located in the middle between two sets of classes, look for the best hyperplane equivalent to maximize the margin or distance between two sets of objects from different classes. Random Forest is commonly used for classifications in large amounts of data. This algorithm combines the tree (tree) by training the sample that is owned. The more using trees will affect the accuracy that will be obtained for the better.

3.4. Net Sentiment Score

The labeled data will also be calculated into the NSS formula. NSS is a method to calculate the level of customer satisfaction with the object to be studied, in this study. The object is the Commuter line Indonesia. NSS has proven having the same results as conventional customer satisfaction measurement methods [5]. NSS scores in the form of numbers with a range from -100 to 100, a higher score indicates that the object under study has been satisfactory.

4. Result

We have established the model using three different classifications method which are Multinomial Naive Bayes, Random Forest and Support Vector Machine. We obtained 0.834%, 0.80% and 0.82% testing accuracy using Multinomial Naïve, Random Forest and SVM. The accuracies are obtained after the data is processed through text-processing. We also tried to calculate the accuracy score from data that is not processed by text-processing. The result is that when the data is not pre-processed, each algorithm outputs has different output. Complete result can be seen in Table 2. In addition, we also obtained measurement evaluation using Precision, Recall and F1-score in order to ensure whether our model can perform well. Complete Classification report can be seen in Table 3.

Table 2. Accuracy Score

Model	Without Text Processing	With Text Processing
Multinomial Naïve Bayes	0.8	0.834
Random Forest	0.81	0.809
Support Vector Machine	0.85	0.820

Table 3. Measurement Evaluation

	Multinomial Naïve Bayes			Random Forest			Support Vector Machine		
	Precision	Recall	F1-Score	Precision	Recall	F1-score	Precision	Recall	F1 Score
Negative	0.82	1	0.9	0.84	0.94	0.89	0.83	0.97	0.90
Positive	1	0.22	0.35	0.6	0.33	0.43	0.71	0.28	0.40
Overall	0.86	0.83	0.79	0.79	0.81	0.79	0.81	0.82	0.79

Table 4. Net Sentiment Score

Keyword	Positive Tweet	Negative Tweet	NSS
Commuter Line	187	648	-55.2

5. Discussion

Based on the performance evaluation result, average score from models is higher than 80%. The process of giving label to dataset by the annotator produces good kappa score 0.675. The dataset is only processed by two classes. The dataset is not balanced between positive and negative sentiments. This results in a positive dataset has lower value of recall and f1score than precision for all models. The SVM algorithm can be seen as an algorithm that produce the highest accuracy score among other algorithms. Because of this research using only 2 class labels (positive and negative), we used SVM model with linear kernel which the linear kernel has a pretty good performance in classifying 2 classes. Pre-processing also plays an important role in text classification, it can be seen from the 3 algorithms that 2 algorithms have a slight increase in accuracy score when the data is done by text pre-processing. Pre-processing is also one method that must be applied in task sentiment analysis or other text-related tasks because doing pre-processing can remove words or sentences that are not related in the task

classification like stop-words, punctuation or symbols, links, and number. Based on NSS calculation in Table 4, we produced negative score that means KRL features like system, service, staff and other don't give satisfaction to customer. This result can be improved or changed because the positive tweet less than negative tweet, so for the next research we would like to increase the number of positive tweet.

6. Conclusion

In this research, we have compared three classification algorithms in determining sentiment from Twitter's data of Commuter Line Indonesia. We applied the same pre-processing on all three classification algorithms. Based on pre-processing result we proceed the data into classification algorithms to compare which algorithms has the highest accuracy. From the discussion session shows that Multinomial Naive Bayes, Random Forest, and Support Vector Machine obtained different result in accuracy as well as in recall and f1-score especially in the positive sentiment which has small score. However, among the three classification algorithms, it can be seen that Support Vector Machine produced the highest accuracy of 85% whereas Multinomial Naive Bayes obtained 82% and Random Forest obtained 84%. Therefore, these algorithms are suitable for sentiment analysis in this research. The dataset also shows that commuterline has many negative sentiments. This can be seen from the amount of negative sentiment data that far exceeds the number of positive sentiment data. This also ensures by NSS calculation. NSS value obtained significantly negative. It means the commuter line features isn't satisfied as a whole whether the system, services, staffs, and other. This performance can be changed, when positive tweet about KRL improve as a time.

References

- [1] AnastasiaS and Budi I 2016 Twitter sentiment analysis of online transportation service providers *2016 Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS)* pp 359–65.
- [2] VidyaN A, Fanany M I and BudiI 2015 Twitter sentiment to analyze net brand reputation of mobile phone providers *Procedia Comput. Sci.* vol 72 pp 519–26.
- [3] Yu Y and Cao Q 2013 The impact of social and conventional media on firm equity value: A sentiment analysis approach *Decis. Support Syst.* vol55 no4 pp 919–26.
- [4] Abbasi A, ChenH and Salem A 2008 Sentiment analysis in multiple languages *ACM Trans. Inf. Syst.* vol26 no3 pp 1–34.
- [5] Ye Q, Zhang Z and Law R 2009 Sentiment classification of online reviews to travel destinations by supervised machine learning approaches *Expert Syst. Appl.* vol36 no3 pp 6527–35.
- [6] Budiono D F, NugrohoA S and Doewes A 2017 Twitter sentiment analysis of DKI Jakarta's gubernatorial election 2017 with predictive and descriptive approaches *2017 Int. Conf. on Computer, Control, Informatics and its Applications (IC3INA)* pp 89–94.
- [7] Alfina I, Sigmawaty D, Nurhidayati F and Hidayanto A N 2017 Utilizing hashtags for sentiment analysis of tweets in the political domain *Proc. of the 9th Int. Conf. on Machine Learning and Computing - ICMLC 2017* pp 43–7.
- [8] Viera A J and Garrett J M 2005 Understanding interobserver agreement: the kappa statistic. *Fam. Med.* vol 37 no 5 pp 360–3.
- [9] Adriani M, Asian J, Nazief B, Tahaghoghi S M M and Williams H E2007 Stemming indonesian *ACM Trans. Asian Lang. Inf. Process.* vol 6 no 4 pp 1–33.
- [10] Le Cun Y, Bengio Y and Hinton G2015 Deep learning *Nature* vol 521 no 7553 pp 436–44.