SENIMENT ANALYSIS OF USERS’ PERCEPTION TOWARDS PUBLIC TRANSPORTATION USING TWITTER

Chua Wan Fen¹, Maizatul Akmar Ismail²*, Tasnim M. A. Zayet³ and Kasturi Dewi Varathan⁴

¹²³⁴ Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, MALAYSIA

*Corresponding author: maizatul@um.edu.my

Accepted: 18 February 2020 | Published: 23 March 2020

Abstract: Twitter is a fast-growing micro-blogging social networking website where people like to express their opinions in a brief and simple way of expression. Today, as the volume of data increases, text mining has become more important for gaining and extracting meaningful insights from the text for certain purposes. Sentiment analysis was applied in this research to investigate the perception of users towards public transport in Malaysia. To improve the public transport system, and offer the passengers a good ride experience, it is crucial to understand the opinions of users. Before analysis, tweets collected from the Twitter API must be pre-processed. The sentiment score of each collected tweet was calculated using three standard lexicons like syuzhet, bing, and afinn. Then, by implementing machine learning algorithms such as Support Vector Machine, Decision Tree and Random Forest, tweets were classified into polarity, either positive, negative or neutral. After that, a confusion matrix was used to investigate the performance of each model. Based on the results of performance metrics, the combination of afinn lexicon with Support Vector Machine was the best classifier with 76.77% accuracy.

Keywords: Text Mining, Public Transportation, Sentiment Analysis, Confusion Matrix.

1. Introduction

Users’ satisfaction is the ultimate aim that any service looking forward to. To achieve this aim, the users’ engagement in any service development and planning becomes a must. Users’ engagement can be assured by getting their feedback, complaints and suggestions. Decision-makers usually used traditional methods like questionnaires and surveys to get users’ point of view. Traditional methods consume time, efforts and personnel. With the evolution of social media in the past years, this process becomes easier (Pournarakis et al., 2017). People published their thoughts and opinions on their social media accounts like Facebook, Twitter, Instagram and others. By scrapping data from these social media platforms and mine the opinions, decision-makers can get users’ feedback. The process of mining the opinion known as sentiment analysis.
Sentiment analysis has been used in multiple domains; in tourism, it is used to get users’ opinion about restaurants (Govindarajan, 2014), hotels (Farisi et al., 2019), airlines (Rane et al., 2018) and others; in e-commerce, it is used to get customers’ opinion about the products (Jabbar et al., 2019); in education, it is used to get students’ feedback (Altrabsheh et al., 2013); in transportation it is used to get commuters’ opinion about the transportation network (Haghighi et al., 2018), road conditions (Kumar et al., 2014), commuters’ complaints (Abali et al., 2018).

In this paper, the perception of users towards public transportation will be investigated. Prasarana Malaysia Berhad (Prasarana) is the organization that drives the development and transformation of the public transport system in Malaysia. Prasarana’s urban rail services include three networks of Light Rail Transit (LRT), KL Monorail and Mass Rapid Transit (MRT). On the other hand, Keretapi Tanah Melayu operates the KTM commuter rail system. This paper is aiming to answer the following questions:

1. How to pre-process and classify tweets into polarity?
2. What is the most appropriate algorithm in classifying tweets into polarity?
3. What is the users' perception towards public transportation?

2. Literature Review

According to the Oxford dictionary, the sentiment is defined as a view or opinion that held or expressed, either a general feeling, opinion or emotion. Dorothy et al., (2016) presented a survey about sentiment analysis methods; several techniques or approaches that have been applied by the researchers in sentiment analysis research are lexicon-based approaches, machine learning approaches and hybrid approaches. Many researches have been done for analyzing users’ opinion toward transportation network.

Gupta et al., (2018) used the hybrid approach to analyze the sentiment of the UK commuters toward the transportation network. Alamsyah et al.,(2018) analysed the users’ opinion toward Uber one of the biggest online transportation service in the world. They used the supervised learning algorithm Naïve Bayes (NB). Anastasia et al., (2016) analysed the users’ sentiment toward online transportation in Indonesia using supervised machine learning techniques Support Vector Machine (SVM), Naïve Bayes (NB) and Decision Tree (DT). Kaur et al., (2018) analysed the users’ opinion toward the airlines in Malaysia using the lexicon approach.

Through the literature, SVM proved its efficiency in text classification. In this study we will screen the efficiency of SVM against other two machine learning algorithm. We will adopt the hybrid approach i.e. the lexicon with supervised machine learning. We will explore the usage of three lexicons with three machine learning algorithms. The lexicons are: Afinn, Bing and Syuzhet and the machine learning algorithms are: SVM, DT and Random Forest (RF). This screening is aimed to find the best combination between lexicon and machine learning technique.
3. Problem Statement

Kuala Lumpur is heading toward being a World-Class city. The transportation infrastructure is an important indicator of the development and economic situations of countries. Therefore, the infrastructure needs to meet the high expectations of the commuters. It has to assure the well-linkage between transportation modes and easy access to them. However, there are many existing public transportation issues such as inadequate station interchange facilities, lack of integration between rail-based stations, and poor quality of services. These issues have caused people to use their own private cars which leads to heavy traffic congestion, especially in rush hours. Irtema et al., (2018) stated that if commuters are pleased with transportation network services, they will efficiently use the transportation modes and even encourage other people to use them. Hence, getting to know the level of commuters’ satisfaction toward the transportation network is important, so that the services can be improved based on the users’ complaints and suggestions.

4. Method

The summarization of the sentiment analysis process is presented in Figure 1. The steps are described in detail in the following sub-sections. R language was used in the whole process.

![Figure 1 The pipeline of sentiment analysis process](image)

4.1 Data Collection

Twitter was the data source for this research. Twitter data was retrieved using Twitter API (Application Programming Interface). The data collection was carried out for 12 weeks, starting from 23rd January 2019 until 14th April 2019. The tweets related to Malaysia’s public transportation were crawled by using annotation such as @kmberhad, @MRTMalaysia, @AskRapidKL, @MyRapidKL and @transitmy. A total of 10,529 tweets were collected. Out of these tweets, only about 10% or 1235 tweets were selected and labeled for the use of this research because most of the tweets collected were irrelevant to the study (e.g. objective tweets).

4.2 Data Pre-processing

Data pre-processing was carried out by removing stop words, white spaces, hashtags, numbers, URLs, annotations and all unfamiliar symbols or words from the tweets. Some misspelled words in the tweets were corrected. Lastly, all words were converted to lower case in the pre-processing process.
4.3 Lexicon-Based Approach

Lexicon-Based sentiment analysis was carried out to investigate the sentiment and polarity of tweets by using syuzhet, bing and afinn lexicons that can be obtained from the syuzhet library.

4.4 Classification Algorithms

Three algorithms were implemented with three different lexicons in order to train the model to classify tweets into a polarity: SVM, RF and DT. Table 1 shows the used models in training the data.

<table>
<thead>
<tr>
<th>Table 1: Models of classifying the tweets into polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>syuzhet (s)</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
</tr>
<tr>
<td>Decision Tree (DT)</td>
</tr>
<tr>
<td>Random Forest (RF)</td>
</tr>
</tbody>
</table>

4.5 Performance Metrics

A confusion matrix was used to compare the accuracy, sensitivity, precision, and F1-score of each model. The best algorithm in predicting users’ perception toward public transportation was determined. Accuracy is the overall performance of an individual classifier; precision is the measurement of correctness; recall is the measurement of completeness; and F1-score is the harmonic mean of precision and recall (Bogdani, 2017).

Figure 2 shows the Equation of Accuracy, Precision, Recall and F1-score, while Figure 3 shows the Example of Confusion Matrix. According to Figure 3, the situation when both actual class and predicted class is labeled as positive, and it is known as True Positive (TP) while True Negative (TN) is the situation when both actual class and predicted class is labeled as negative. When a negative class data is wrongly predicted to be a positive class data, it is known as False Positive (FP) while False Negative (FN) is the condition when a positive class data is wrongly predicted to be a negative class data.

\[
\text{Accuracy} = \frac{TP + TN}{\text{Total Number of Observations}}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Figure 2 Equation of Accuracy, Precision, Recall and F1-score
4.6 Validity and Reliability

Insufficient data is one of the limitations of this research. Most of the tweets related to public transportation are objective tweets, where people ask the administrators about the arrival time of public transport and the way to get to a destination. After eliminating 47 of these tweets, the tweets left for this research is relatively small. The second limitation of this research is time constraints. Due to the limited time given to complete this project, tweets are collected from Twitter over the consecutive twelve weeks. Therefore, it is advisable to collect data for a longer period. In the context of the research method, only three lexicons and three algorithms are applied in classifying tweets into polarity. Extending this research with other lexicons and algorithms is the solution that can be applied to identify the most generalized models in classifying tweets regarding public transportation into polarity.

5. Results and Discussion

In this section, the evaluation of the models (Table 1) will be presented and discussed. The performance of each model can be summarized into three tables as below. Table 2, Table 3 and Table 4 shows the performance of each algorithm implemented with syuzhet, bing, and afinn lexicon respectively, Figure 4 shows the performance grouped according to matrics. The performance of the models was evaluated through the confusion matrix. The best model is then used for analysing the sentiment.

### Table 2 Performance of each algorithm with syuzhet lexicon

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Random Forest</th>
<th>Support Vector Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.6505051</td>
<td>0.7393939</td>
<td>0.7292929</td>
</tr>
<tr>
<td>Precision</td>
<td>0.6015949</td>
<td>0.7339713</td>
<td>0.7172648</td>
</tr>
<tr>
<td>Recall</td>
<td>0.6772383</td>
<td>0.7323950</td>
<td>0.7112304</td>
</tr>
<tr>
<td>F1</td>
<td>0.5938637</td>
<td>0.7247426</td>
<td>0.7106031</td>
</tr>
</tbody>
</table>

### Table 3 Performance of each algorithm with bing lexicon

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Random Forest</th>
<th>Support Vector Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.6545455</td>
<td>0.7090909</td>
<td>0.7050505</td>
</tr>
<tr>
<td>Precision</td>
<td>0.6359227</td>
<td>0.7010123</td>
<td>0.6997060</td>
</tr>
<tr>
<td>Recall</td>
<td>0.7119975</td>
<td>0.7287049</td>
<td>0.7218752</td>
</tr>
<tr>
<td>F1</td>
<td>0.6513149</td>
<td>0.7095137</td>
<td>0.7076313</td>
</tr>
</tbody>
</table>

Copyright © 2020 ACADEMIA INDUSTRY NETWORKS - All rights reserved
Table 4 Performance of each algorithm with afinn lexicon

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Decision Tree</th>
<th>Random Forest</th>
<th>Support Vector Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.7171717</td>
<td>0.7515152</td>
<td>0.7676768</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7105967</td>
<td>0.7482989</td>
<td>0.7637773</td>
</tr>
<tr>
<td>Recall</td>
<td>0.8006952</td>
<td>0.7641503</td>
<td>0.7769997</td>
</tr>
<tr>
<td>F1</td>
<td>0.7128273</td>
<td>0.7523859</td>
<td>0.7665314</td>
</tr>
</tbody>
</table>

According to Figure 4, Afinn lexicon outperformed other lexicons in terms of all metrics: accuracy, precision, recall and F1. Based on Figure 4, Support Vector Machine (SVM) achieves the highest score in terms of accuracy, F1-measure, and Precision compared to Decision Tree and Random Forest algorithms. However, the Recall of Decision Tree is the highest among these three algorithms, which is 80.07%. The accuracy, F1-measure, Precision, and Recall of SVM are 76.77%, 76.65%, 76.38%, and 77.70% respectively.

As Afinn was the best, it was used to assign the polarity to the tweets. Figure 5 illustrates the total number of tweets for each polarity. The polarity of each tweet is assigned to either -1 (negative), 0 (neutral) or 1 (positive). Out of the collected data within twelve weeks, there are 385 negative tweets, 417 neutral tweets and 434 positive tweets regarding public transportation in Malaysia. Notice that the total number of tweets is likely to be equal for each polarity. The total number of positive tweets is slightly higher than that of neutral and negative tweets. This can be due to the small number of retrieved tweets in 12 weeks period.
Throughout the analysis, frequencies of words were found. Figure 6 shows the word cloud of the words, the size of the word reflects its frequency. The most frequent word that appeared in the collected dataset is the word ‘thank’ which conveys a positive meaning. The most frequent word that conveys a negative meaning is the word ‘wait’. In addition, the most frequent bigrams in the negative and positive polarities were found, Figure 7 and Figure 8 shows these frequencies.
From Figure 4.6, it can be observed that some people think the services are nice, excellent, amazing and awesome. The helpful and friendly staff is one of the positive perceptions of users towards public transportation in Malaysia. Moreover, people are thrilled with the effectiveness of service providers in solving the cleanliness or service maintenance problems. For example, if the air-conditioner is not functioning, the technicians will come to solve the problem within a short space of time. In Figure 4.7, the negative users’ perceptions of public transportation can be divided into two categories — service and punctuality. Poor maintenance is one of the main factors. For example, at some stations, the malfunctioning escalators were left broken without any maintenance for a long period of time. This will cause inconvenience especially for those who are pressed for time. Furthermore, users are disappointed with bad or poor management and services. For instance, during peak hours, the frequency of public transport services
provided is insufficient to accommodate the number of passengers. Consequently, the passengers are forced to wait for the next round service.

The users are also dissatisfied with the cleanliness of the public transport, especially the dirty windscreen. During the ride, some users may feel unsafe when the drivers suddenly and frequently step on the emergency brake on the traffic jam road. In the punctuality context, the variation of public transport arrival and departure times from the schedules shows a low level of service punctuality. Most of the users experienced delays and waited for a long period of time for public transport. A sudden stop or stuck of train on the track may affect its punctuality. To make sure people can reach their destination on time, an announcement should be made clearly and properly to the users.

6. Conclusion

In this paper, we presented an analysis of computers’ perspective toward transportation network in Malaysia. The data was collected from Twitter for 12 weeks and then processed and annotated. Three lexicons (Afinn, Bing and syuzhet) with three supervised learning algorithms (SVM, RF and DT) were tested. The best was Afinn-SVM model with accuracy (76.77%), precision (76.38%) and F1-measure (76.65%). Afinn-SVM model used after that for sentiment classification. The frequency of bigrams in the positive and negative polarity was found. The negative bigrams indicated the commuters’ complaints while the positive ones indicated the satisfaction features.

7. Acknowledgment

This work was supported by the University of Malaya under Grant RK004-2017 and Sunway University under Grant CR-UM-SST-DCIS-2018-01.

References


