

## Using Topic Modeling to Identify Public Opinion on Public Transportation During the COVID-19 Pandemic



Prepared by:

Ali Rahim Taleqani  
Kenechukwu Ezekwem  
Antonio Molina  
Taraneh Askarzadeh  
Narendra Malalgoda  
Jill Hough, Ph.D.  
Jeremy Mattson, Ph.D.

Small Urban and Rural Center on Mobility  
Upper Great Plains Transportation Institute  
North Dakota State University

# **Using Topic Modeling to Identify Public Opinion on Public Transportation During the COVID-19 Pandemic**

Ali Rahim Taleqani  
Kenechukwu Ezekwem  
Antonio Molina  
Taraneh Askarzadeh  
Narendra Malalgoda  
Jill Hough  
Jeremy Mattson

Small Urban and Rural Center on Mobility  
Upper Great Plains Transportation Institute  
North Dakota State University  
Fargo, North Dakota

August 2021

## **Acknowledgements**

Funds for this study were provided by the Small Urban, Rural and Tribal Center on Mobility (SURTCOM), a partnership between the Western Transportation Institute at Montana State University and the Upper Great Plains Transportation Institute at North Dakota State University. The Center is funded through the U.S. Department of Transportation's Office of the Assistant Secretary of Research and Technology as a University Transportation Center. The Small Urban and Rural Center on Mobility (SURCOM) within the Upper Great Plains Transportation Institute at North Dakota State University conducted the research. This research was conducted by a group of SURCOM graduate students during the COVID-19 pandemic.

## **Disclaimer**

The contents presented in this report are the sole responsibility of the Upper Great Plains Transportation Institute and the authors.

NDSU does not discriminate in its programs and activities on the basis of age, color, gender expression/identity, genetic information, marital status, national origin, participation in lawful off-campus activity, physical or mental disability, pregnancy, public assistance status, race, religion, sex, sexual orientation, spousal relationship to current employee, or veteran status, as applicable. Direct inquiries to: Vice Provost, Title IX/ADA Coordinator, Old Main 201, 701-231-7708, [ndsueti@ndsueti.edu](mailto:ndsueti@ndsueti.edu).

## **ABSTRACT**

This study examines public opinion toward transit during the COVID-19 pandemic. This was accomplished through topic modeling using social media posts on Twitter. Results showed how discussion regarding public transit changed during the first several months of 2020. This study also shared lessons learned while developing the problem statement and methodology.

# TABLE OF CONTENTS

<b>1. Introduction .....</b>	<b>1</b>
<b>2. Public Transportation and COVID-19.....</b>	<b>2</b>
<b>3. Literature Review .....</b>	<b>3</b>
<b>4. Methods .....</b>	<b>5</b>
4.1 Topic Modeling.....	5
4.2 Data Preprocessing and Feature Engineering .....	5
<b>5. Results.....</b>	<b>8</b>
<b>6. Lessons Learned .....</b>	<b>10</b>
<b>7. Conclusions .....</b>	<b>12</b>
<b>References.....</b>	<b>13</b>

## **LIST OF TABLES**

**Table 4.1** Number of Collected Tweets by Month (Excluding Retweets) ..... 6

**Table 5.1** Topic and Top 10 Frequent Words by Month..... 9

## **LIST OF FIGURES**

**Figure 3.1** Topic Modeling Approach..... 5

# 1. INTRODUCTION

The COVID-19 pandemic has had significant impacts on travel behavior and created major challenges for public transit agencies. To respond to changes in public sentiment and to improve operational performance and better serve riders requires dynamic feedback. Feedback has been traditionally obtained through surveys of riders and the public, but the conventional survey method is costly and time-consuming. It cannot capture real-time issues associated with day-to-day transportation operations. With the growth of social platforms, there is an opportunity to better understand the users' opinions about transportation-related topics in real time. Social media provide an abundant source of data.

This study examines public opinion toward transit during the pandemic using topic modeling. Topic modeling is a type of statistical modeling for discovering the abstract topics that occur in a collection of documents. We used the topic modeling method to extract relevant topics from peoples' discussion about public transportation on social platforms. Social media usage has made a paradigm shift in the way people express feelings, opinions, or evaluations through social networks. One of the most used social media platforms for this kind of interaction is Twitter (Maghrebi et al. 2015). More feedback and ideas become available when a debatable topic happens. Everyone jumps on their smartphone and leaves their thoughts. Users now have all kinds of advantages to show their opinions on any topic they want. Extensive use of devices that have access to the internet creates a vast amount of data. The sum of these opinions can generate valuable information when analyzed (Luong and Houston 2015).

Sentiment analysis and topic modeling of users are the two primary machine learning tasks in public transportation. The former classifies users' opinions on a topic of interest into negative, neutral, or positive classes. The latter, as unsupervised, finds the latent topics in users' discussion. Topic modeling is a machine learning technique that identifies groups of words according to their statistical use in a set of documents. These documents can be any unstructured text. Every topic is the set of words that are more likely to appear together in texts; topic modeling allows the analysis to capture this insight into a mathematical framework. In this study, tweets regarding public transportation during January 2020 through July 2020 were collected and analyzed using topic modeling to examine public opinions during this period. This analysis covered a period from before the start of the pandemic through the first few months.

Section 2 provides background on how public transportation was affected by the pandemic. Previous studies that have used sentiment analysis and topic modeling are discussed in Section 3. The methodology is described in Section 4, and the results, lessons learned, and conclusions are presented in sections 5 – 7.

## 2. PUBLIC TRANSPORTATION AND COVID-19

The COVID-19 virus initiated in Wuhan, China, became a global pandemic in a few weeks. Its rapid spread can be attributed to our current lifestyle's hypermobility, connectivity caused by globalization, and access to different mobility options (Musselwhite, Avineri, and Susilo 2020). As a result of the pandemic, people worldwide face challenges. There have been significant changes in our lifestyle and global travel. For example, due to the pandemic, air travel decreased significantly, and there has been a massive increase in teleworking. Some of these impacts are governmental measures, whereas others are due to consumers' personal choices to stay safe (Tirachini and Cats 2020).

Furthermore, complete lockdowns of cities and countries had much more severe impacts on public mobility and considerable repercussions for work and fulfilling everyday duties (Musselwhite, Avineri, and Susilo 2020). Public transportation has been hit hard by COVID-19 worldwide. Operators are struggling with low ridership while maintaining all Centers for Disease Control and Prevention (CDC) and National Institutes for Health (NIH) guidelines, resulting in large operational costs. The situation, hence, created significant financial problems for transit agencies. Given the dominant car culture and telework options available to many American workers, public and private stakeholders must understand and evaluate the situation to make informed decisions for future planning.

Even during challenging times, such as during the Covid-19 pandemic, public transportation remains one of the top essential services. It sometimes plays a crucial role in providing access to health facilities. Therefore, one of the critical responsibilities of transport operators is to maintain operations despite the constraints brought by the spread of COVID-19. Thus, during these times, such providers should focus on their staff's safety and address challenges that come with absenteeism (UITP, 2020). Furthermore, public transit systems have become vital, especially to maintain economic stability and equity throughout the world (Musselwhite, Avineri, and Susilo 2020).

Although deemed essential, the pandemic poses a challenge for public transportation on a global scale, stemming from a drastic decline in both supply and demand (Tirachini and Cats 2020). Therefore, since the pandemic started, public transportation has been in crisis mode, with ridership in major cities in the United States, Europe, and China declining 50% to 90%. More specifically, there was a 74% decline in New York, 79% in Washington, D.C., 93% in Boston, and 87% in the Bay Area.

As cities reopen, transit agencies' next challenge is addressing massive budget shortfalls reclaiming previous riders (Clemens and Veuger 2020). Measures such as high-visibility cleaning and health messaging campaigns, together with universal mask-wearing mandates, can help reassure passengers that it is safe to use public transit.

During the pandemic, the perception that using public transit poses a higher risk than using personal vehicles led to loss of ridership and service reduction, respectively. As pointed out by Teixeira and Lopes (2020), major cities saw an increased use of personal bicycles, which may have served as an optional mode of transportation for transit users in these cities (Teixeira and Lopes 2020).



### 3. LITERATURE REVIEW

Recent studies have used sentiment analysis and topic modeling to analyze public opinions and customer satisfaction of transportation, and transit in particular. Haghghi et al. (2018) conducted a study in Salt Lake City to analyze and evaluate opinions about transit service quality. By using text mining techniques to analyze opinions on Twitter and a Latent Dirichlet Allocation (LDA) model, useful results were obtained for future investment in the decision-making process (Haghghi et al. 2018). Gal-Tzur et al. (2014) focused on defining the criteria for gathering transport-related knowledge from social media. The conclusions proved that these details would offer useful insight into creating or implementing transport policies and the difficulties that this implies (Gal-Tzur et al. 2014).

Korfiatis et al. (2019) used online feedback to display the knowledge gained from a consideration of variables found in the topic modeling of unstructured data and offered a flexible extension of numerical ratings to recognize customer satisfaction and service quality. Once the numerical and textual characteristics were mixed, the explained variance in total satisfaction increased dramatically. They extended their approach to airline passengers' online comments using the structural subject models (STM), a recent probabilistic expansion of the Latent Dirichlet Allocation (LDA) that allows for the inclusion of text-level covariates.

El-Diraby, et al. (2019) presented an approach for using social and semantic network analysis to support the modeling of customer sentiment and satisfaction of transit systems. They used data from the TransLink (Vancouver Transit Agency) Twitter account, monitored it for 11 months, and compared it with data from two other Canadian cities. Their findings were essential to discover a negative sentiment toward the transit system (resulting in areas that could be improved). They also realized that language used is critical to understand survey and Twitter data better (El-Diraby et al. 2019).

Ali et al. (2019) analyzed the opinions on Twitter, Facebook, and TripAdvisor. They went further by using ontologies to improve an LDA model and a word embedding system to get a more accurate sentiment classification. They studied traffic control and management systems of intelligent transportation systems. They found that this approach better identifies social media texts and enhances the analysis (Ali et al. 2019). Farrahi and Gatica-Perez (2011) proposed a method using an LDA model. Still, they chose to study 97 subjects, and the dataset was created from their cellphones' location for 16 months. They wanted to discover and analyze travel patterns and mobility conducts. Their approach showed that, depending on the scope of the study, an LDA model may be used in different ways to categorize the information efficiently (Farrahi and Gatica-Perez 2011).

Arafat et al. (2020) helped transportation researchers and the TRB group during the COVID-19 pandemic by reviewing the output of software models used to collect large-scale data from COVID-19-based Twitter sources. They expanded previous research efforts in the field of social media data mining by reviewing contemporary tools, including their computational sophistication and potential utility. They performed an open repository for the produced data frames to promote the rapid development of new research studies in transportation. They recommended that the TRB community use it when deciding to investigate further issues related to COVID-19 and tools for the mining of social media data (Arafat 2020).

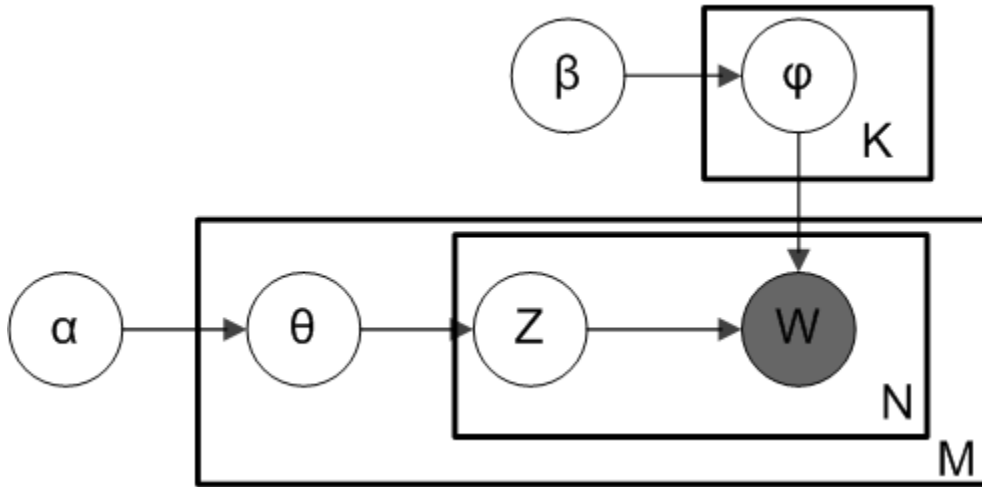
Qi et al. (2020) presented a comprehensive framework that facilitates high-efficiency public sentiment collection and analysis of transport services from Twitter. They chose the Miami-Dade County transport system as a case study to explain the framework production and validation process. Topical modeling and tokenization techniques were also used to extract the relevant semantic content required for data analysis. The developed framework can enhance the accuracy of Twitter data extraction and analysis (Qi et al. 2020).

Han et al. (2020) studied public opinion at the early stages of COVID-19 in China by evaluating Sina-Weibo (a Twitter-like microblogging system in China) texts in terms of space, time, and content. Harmonic distortion within one-hour intervals and the spatial distribution of the COVID-19 Weibo-related texts were evaluated. Based on the LDA model and the random forest algorithm, a concept extraction and classification model was established to hierarchically identify seven relevant COVID topics and 13 sub-topics from Weibo texts. The results showed that the number of Weibo texts changed over time with respect to different topics and sub-topics related to the developmental stages of a project. There is matching in between common daily discussions on Weibo and the pattern of the pandemic of COVID-19 in the actual world (Han et al. 2020).

## 4. METHODS

### 4.1 Topic Modeling

Blei et al. (2003) introduced the LDA method to explain a set of observations by a set of unobserved groups. In document analysis, LDA, as an unsupervised learning model, views documents as bags of words. We can assume a document was generated by picking a set of topics and then picking a set of words for each topic. Then we reverse engineer this process to extract relevant topics.



**Figure 3.1** Topic Modeling Approach

Figure note:  $\alpha$ : per-document topic distributions,  $\beta$ : per-topic word distribution,  $\theta$ : topic distribution for document  $m$ ,  $\phi$ : word distribution for topic  $k$ ,  $z$ : topic for the  $n$ -th word in document  $m$ ,  $w$ : specific word

First, we assume there are  $k$  topics across all of the documents. Then we distribute these  $k$  topics across document  $m$  (per-document topic distributions,  $\alpha$ , can be symmetric or asymmetric) by assigning each word a topic. For each word  $w$  in document  $m$ , we assume its topic is wrong, but every other word is assigned the correct topic. We probabilistically assign word  $w$  a topic based on two things: 1) what topics are in document  $m$ , and 2) how many times word  $w$  has been assigned a particular topic across all of the documents (or known as per-topic word distribution  $\beta$ ). We repeat this process several times for each document.

### 4.2 Data Preprocessing and Feature Engineering

The analysis covered seven months (from January 1, 2020, to July 31, 2020) and included tweets in English using Twitter API. A total of 2,661 unique tweets from 4,574 unique users with the hashtag (#) of “public transport” or “public transportation” were chosen as the sample in this analysis. The hashtag (#) is useful in categorizing tweets and simplifies the process of searching for tweets. Retweets (tweets that start with “RT”) were treated as duplicates and were removed from both sets of data. The initial analysis shows that #coronavirus, #COVID19, and #bayareatechtransit were the most repeated hashtags, which appeared over 50 times. Tweets were then cleaned by removing punctuation, special characters, digits, emoticons (such as emojis), and uniform resource locators (URLs) so that the dataset contained only words. The number of collected tweets by month are described in Table 2.1.

**Table 4.1** Number of Collected Tweets by Month (Excluding Retweets)

Month	Number of Tweets
January	400
February	408
March	419
April	330
May	332
June	267
July	324

Tokenization is the process of breaking up a sequence of strings into pieces called tokens. Tokenization aims to explore the words in a sentence and identify meaningful keywords. Punctuation was removed in the process of tokenization. Tokens can be made up of characters, numeric or alphanumeric. Following this, stop-words are removed from the tweets. Stop-words are words from the non-linguistic view that do not carry information. Prepositions, articles, and pronouns were treated as stop-words in our work. Eliminating stop-words helps to improve text processing performance. Next, word stemming was executed. Word stemming is a process of transforming words into their roots. Many words in English have different forms of the same words; for example, “stemming,” “stemmed,” and “stems” have the same root word: “stem.” Lastly, capital letters were converted to lower case. Tweets were then converted into a corpus. A corpus is a large, structured set of texts.

Texts have many distinct properties. There are several ways to extract the features within a text and present it properly for the classifier. The TF-IDF method is a statistical measure for evaluating how important a word is to a corpus document. The importance increases proportional to the frequency of a word’s appearance in a document and is offset by the word’s frequency in the corpus.

$tf(w, d) = f_a(w)$ : frequency of  $w$  in document  $d$

$$idf(w, D) = \log \frac{1 + |D|}{1 + df(d, w)}$$

$$tfidf(w, d, D) = tf(w, d) \times idf(w, D)$$

Unigrams, bigrams, and n-gram models, with their frequency counts, are considered as features. N-grams are all combinations of adjacent words or letters of length  $n$  in a continuous sequence from a given text or speech. Since tweets are usually very short in length, we considered only bigrams and used the Gensim library to train the LDA model. We also tuned the following hyperparameters to find the best coherence score. A set of statements or facts is said to be coherent if they support each other.

- **The number of topics:** Each topic is a set of keywords, each contributing a certain weight (i.e., importance) to the topic. Too few topics result in non-cohesive topics, containing a heterogeneous set of words that are difficult to distinguish. Too many topics do not provide any semantic meaning of any topic. Given that it was a big event, and tweets covered several days, the number of topics of all tweets is likely high. We set three as the number of topics. We also tried between two and 10 topics, and none showed a better result than three.

- **Alpha (Document-Topic Density):** A lower value for alpha indicates it is more likely that a tweet may contain a mixture of just a few of the topics. The default is  $1.0/\text{NUM\_TOPICS}$ . We used 0.01, as each tweet is relatively short and is very likely only to have one topic.
- **Eta (Word-Topic Density):** The lower the eta, the more likely a topic may contain a mixture of just a few of the words. Default is  $1.0/\text{NUM\_TOPICS}$ . We used 0.01.
- **The number of passes:** Number of training passes/iterations over all tweets. We set it to be 50. We tried bigger numbers, and they did not produce better results.

## 5. RESULTS

We applied the LDA algorithm to our corpus of tweets, and the results are presented in Table 2.2. We listed three important topics, each of which has 10 sub-topics. We analyzed the tweets by month to track if there has been any change throughout the pandemic. In the first two months of 2020, the topics were mostly about sustainability and general policy and operation. Starting in March, one of the main topics was about the COVID-19 outbreak and related safety issues.

Moreover, technologies that can be helpful were also mentioned, such as ZIG transit or contactless payment. Some keywords refer to problems that may describe an image embedded in the tweet. In April, with the increasing impacts of COVID-19 on the economy, money, work, and economy were part of Twitter discussion. As public transportation was under the pressure of low ridership and potential coronavirus exposure, a wave of support occurred in discussions in May. As the pandemic was getting worse in June and July, the discussions shifted to understanding new norms and grants to fund public transportation, infrastructure, and more safety issues.

As suggested in previous work, topic modeling could complement the conventional survey method to extend the target audience. Also, the locations could not always be captured because of privacy, so it is suggested to focus on certain public or private entities to extract topics. In this case, it is also possible to fetch more content from other social platforms like Facebook, in which data could be downloaded by page/user id. One of the potential applications of topic modeling in transportation could be event detection in real time.

**Table 5.1** Topic and Top 10 Frequent Words by Month

	January	February	March	April	May	June	July
<b>Topic 1</b>	Lighting	Sketchbook	Fare validation	Stay home	Work	MIT	Develop house
	Green building	Illustration	Train	Line	Avail	Payment	Transit
	Summary	Art	Contactless	COVID	Pandemic	Train	Mass transit
	Free	Sketch	COVID	Help	Heroes act	Avail	Brink
	Today	Artwork	Safe	Work	Support	Greenline	Business
	Work	Event	House	Transit	Mass transit	Day	Grant due
	Public transit	Street	Bus story	Travel	Commute	Import	Housing grant
	Next	Drawing	Coronavirus	Climate change	RAPT group	Use	transit-oriented development
	MIT	Travel	Work	Safe	System	Worker	Digital signage
	Time	Car	Transit	Mobility	Day	People	
<b>Topic 2</b>	Transit	Train	Today	Today	Today	Bus	Transit
	Use	Transit	Cold	Route	Support	Use	Drive efficient
	Commuter life	Part	Travel	COVID	Public transit	Travel	Efficient
	Fit	Portrait	Weather	Dayton local	Route	Avoid	System
	Ride	Job	Operation	Muncie in motion	SEPTA	Support	Video
	Bus	Work	Public transit	MIT	Frontline	Transport	Climate action
	Help/get	Good	Time	Food	Custom feel	San Francisco	Access
	Learn/use	Member Voice4 Transit	ZIG transit	Money	Train	Big	Return
	Commute	Pm_pt_pt	Place	Bus	Transit	Rail	Security
Look/way	Pm_pt	public transit	Economy	Help	Area	Post	
<b>Topic 3</b>	Train	Line	Smart solutions	Transit	Subway	Transit	Traffic
	Car	Today	Bus	Public transit	COVID	Wear mask	Mask
	Bus	Plan	Bike	Worker	Lighting	Train	Time
	Travel	Sure	e-mobility	Employee	Station	Cycling	Mind
	#bospoli	Work	BART chronicle	System	Train	Government	Infrastructure
	Station	New	e-roller	People	Transit	Urban mobility	New normal
	Badi	Budget	Country	COVID	GO Station	System	Time
	System	Transit	Ride	Operation	Summary	Public transit	Country
	Tri Rail	Public transit	Big	Open	Green building	Differ way	Disinfection
Good	Time	Input	Train	Mask	Agency	Law	

## 6. LESSONS LEARNED

We offer some recommendations on two main subjects: 1) conventional survey-based methods vs. machine learning models, and 2) social platforms and the difference between the two leading platforms of Facebook and Twitter. These recommendations are based on the current business model of social media and might change in the future.

**Noisy data:** If we assume noise as an unwanted, unpleasant, disruptive event to our analysis, data from social platforms either in graphical or textual format have a different type of noise. First, there is no guarantee that textual data are proofread, so it is impossible to check thousands or millions of texts' grammatical errors. Second, retweets on Twitter or sharing a Facebook post without adding any extract words do not represent the users' real intentions. It is too vague to conclude approval or disagreement. The third is mostly related to the Twitter platform. Using hashtags is a common way to find related tweets. For many reasons, users use hot hashtags on unrelated tweets to get more views. This made the entire dataset invalid. Finally, mostly as of 2020, these two leading platforms restrict/remove content if they violate their rules. While it is beneficial to remove harmful content, it raises the question about censorship.

**Bots:** A reliable source is a critical subject in the social network. Social network analysis is prone to social bots. Social bots are, to some extent, autonomous agents to influence users on social media. Social bot detection is complicated, requiring independent research, which is not the scope of this study. In addition to bots, the frequency of a user's tweets might also increase this approach's bias. Twitter, Facebook, and others delete a significant number of accounts every day. If a researcher collects tweets or posts from this group of users, their research could no longer be validated.

**Bias:** Online platforms are open to everyone. This level of flexibility could have an impact on someone's opinion toward a subject. The tweets and the user profile might influence the perception of others toward the user's tweets. Unlike traditional surveys, in which other participants are anonymous, online platforms to some extent have exposure to others.

**Context:** The availability of data is highly dependent on the context of research. Most textual data are around politics and entertainment, while there is less information on transportation, especially public transportation. These also vary among platforms mostly because of users' preference or the number of characters allowed.

**Archived vs. real-time collection:** When setting up the project and considering the budget, it was worth exploring the data collection options. Accessing archived data is usually a more expensive option than streaming ones. So, if the research question is well defined and you do not need to explore historical records, it is better to live-stream the data.

**Users or keyword:** Between two primary social platforms, Twitter and Facebook, there is a significant difference. Twitter allows for collecting tweets by either using hashtags or user-name. In contrast, because of different structures, you could only collect data on the user/entity. Hence, the research statement should clearly define the research scope if, for example, evaluating the difference between two entities or aggregated data from users.



**Location:** From a transportation planning perspective, every transportation study is, to some extent, limited to a geographic area. One might be broader than another. Social network analysis, including Twitter, Facebook, and others, fails to address this issue adequately. Because of users' privacy preferences, locations are either missing or not well represented. Users may choose, at their discretion, to enable/disable location tracking on their devices. For example, someone may post a tweet about an event without adding a location to the tweet—which happens in most cases—or someone may get off a plane and tweet an opinion (complaint) at their destination about the delay at the origin.

## 7. CONCLUSIONS

Topic modeling was used to analyze Twitter posts about public transportation during the first seven months of 2020. Results showed that in the first two months of 2020, the topics of tweets were mostly about sustainability and general policy and operation. Starting in March, one of the main topics was about the COVID-19 outbreak and safety issues. In April, with the increasing impacts of COVID-19 on the economy, money, work, and economy were part of Twitter discussion. As public transportation was under the pressure of low ridership and potential coronavirus exposure, a wave of support occurred in discussions in May. As the pandemic was getting worse in June and July, the discussions shifted to understanding new norms and grants to fund public transportation, infrastructure, and more safety issues.

A few challenges are encountered when conducting this type of analysis. This study described these issues and provided recommendations regarding noisy data, bots, bias, context, archived vs. real-time collection, users or keywords, and location.

## REFERENCES

- Ali, Farman, Daehan Kwak, Pervez Khan, Shaker El-Sappagh, Amjad Ali, Sana Ullah, Kye Hyun Kim, and Kyung Sup Kwak. 2019. "Transportation Sentiment Analysis Using Word Embedding and Ontology-Based Topic Modeling." *Knowledge-Based Systems* 174: 27–42. <https://doi.org/10.1016/j.knosys.2019.02.033>.
- Arafat, Mahmoud. 2020. "A Review of Models for Hydrating Large-Scale Twitter Data of COVID-19-Related Tweets for Transportation Research."
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.
- Clemens, Jeffrey, and Stan Veuger. 2020. "Implications of the COVID-19 Pandemic for State Government Tax Revenues." National Bureau of Economic Research.
- El-Diraby, Tamer, Amer Shalaby, and Moein Hosseini. 2019. "Linking Social, Semantic and Sentiment Analyses to Support Modeling Transit Customers' Satisfaction: Towards Formal Study of Opinion Dynamics." *Sustainable Cities and Society* 49 (April): 101578. <https://doi.org/10.1016/j.scs.2019.101578>.
- Farrahi, Katayoun, and Daniel Gatica-Perez. 2011. "Discovering Routines from Large-Scale Human Locations Using Probabilistic Topic Models." *ACM Transactions on Intelligent Systems and Technology* 2 (1): 1–27. <https://doi.org/10.1145/1889681.1889684>.
- Gal-Tzur, Ayelet, Susan M. Grant-Muller, Tsvi Kuflik, Einat Minkov, Silvio Nocera, and Itay Shoor. 2014. "The Potential of Social Media in Delivering Transport Policy Goals." *Transport Policy* 32: 115–23. <https://doi.org/10.1016/j.tranpol.2014.01.007>.
- Haghighi, N. Nima, Xiaoyue Cathy Liu, Ran Wei, Wenwen Li, and Hu Shao. 2018. "Using Twitter Data for Transit Performance Assessment: A Framework for Evaluating Transit Riders' Opinions about Quality of Service." *Public Transport* 10 (2): 363–77. <https://doi.org/10.1007/s12469-018-0184-4>.
- Han, Xuehua, Juanle Wang, Min Zhang, and Xiaojie Wang. 2020. "Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China." *International Journal of Environmental Research and Public Health* 17 (8): 2788.
- Korfiatis, Nikolaos, Panagiotis Stamolampros, Panos Kourouthanassis, and Vasileios Sagiadinos. 2019. "Measuring Service Quality from Unstructured Data: A Topic Modeling Application on Airline Passengers' Online Reviews." *Expert Systems with Applications* 116: 472–86. <https://doi.org/10.1016/j.eswa.2018.09.037>.
- Luong, Thuy T. B., and Douglas Houston. 2015. "Public Opinions of Light Rail Service in Los Angeles, an Analysis Using Twitter Data." *IConference 2015 Proceedings*.
- Maghrebi, Mojtaba, Alireza Abbasi, Taha Hossein Rashidi, and S. Travis Waller. 2015. "Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 208–13. IEEE.

- Musselwhite, Charles, Erel Avineri, and Yusak Susilo. 2020. "Editorial JTH 16–The Coronavirus Disease COVID-19 and Implications for Transport and Health." *Journal of Transport & Health* 16: 100853.
- Qi, Bing, Aaron Costin, and Mengda Jia. 2020. "A Framework with Efficient Extraction and Analysis of Twitter Data for Evaluating Public Opinions on Transportation Services." *Travel Behaviour and Society* 21: 10–23.
- Teixeira, João Filipe, and Miguel Lopes. 2020. "The Link between Bike Sharing and Subway Use during the COVID-19 Pandemic: The Case-Study of New York's Citi Bike." *Transportation Research Interdisciplinary Perspectives* 6: 100166.
- Tirachini, Alejandro, and Oded Cats. 2020. "COVID-19 and Public Transportation: Current Assessment, Prospects, and Research Needs." *Journal of Public Transportation* 22 (1): 1.
- UITP. 2020. Management of COVID-19: Guidelines for Public Transport Operators. February. [https://cms.uitp.org/wp/wp-content/uploads/2020/06/Corona-Virus\\_EN.pdf](https://cms.uitp.org/wp/wp-content/uploads/2020/06/Corona-Virus_EN.pdf)