3-20-2021

Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis

Syed Ahnaf Morshed
Department of Civil and Environmental Engineering, Florida International University

Raihanul Bari Tanvir
School of Computing and Information Sciences, Florida International University

Sifat Shahriar Khan
Department of Electrical Engineering, The University of Akron

Shafkath Nur
Department of Civil and Environmental Engineering, Islamic University of Technology

Journal of Urban Management 10 (2021) 155–165

Contents lists available at ScienceDirect

Journal of Urban Management

journal homepage: www.elsevier.com/locate/jum

Research Article

Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis

Syed Ahnaf Morshed [a,*], Sifat Shahriar Khan [b], Raihanul Bari Tanvir [c], Shafkath Nur [d]

[a] Department of Civil and Environmental Engineering, Florida International University, 10555, West Flagler Street, EC 3730, Miami, FL, 33174, USA
[b] Department of Electrical Engineering, The University of Akron, Leigh Hall, Room 515, Akron, OH, 44325-2101, USA
[c] School of Computing and Information Sciences, Florida International University, 10555 West Flagler Street, EC 3730, Miami, FL, 33174, USA
[d] Department of Civil and Environmental Engineering, Islamic University of Technology, Board Bazar, Gazipur, Bangladesh

ARTICLE INFO

Keywords:
Ride-hailing
TNC
Sentiment analysis
Emotion detection
Twitter data
COVID-19

ABSTRACT

Ride-hailing services have gained popularity in recent years due to attributes such as reduced travel costs, traffic congestion, and emissions. However, with the impact of COVID-19, the ride-hailing market is estimated to lose its fair share of an uprising as a transportation mode. During normal and critical circumstances, ride-hailing service users express their concerns, habits, and emotions through posting on social platforms such as Twitter. Hence, Twitter, as an emerging data source, is an effective and innovative digital platform to observe the rider's behavior in ride-hailing services. This study hydrates large-scale Twitter reactions related to shared mobility to perform comparative sentiment and emotion analysis to understand the impact of COVID-19 on transportation network services in pre-pandemic and during pandemic conditions. Amid pandemic, negative tweets (34%) associated with 'sad' (15%) and 'anger' (15%) emotions were most prevalent in the dataset.

1. Introduction

Shared mobility is the shared use of personal vehicles in exchange for monetary transactions. It fulfills the requisite temporal demand of consumers by providing essential transportation services. It is an innovative initiative that is based on the "access to transport" rather than "ownership of transport" (personal vehicles). This form of transportation provides mobility and access to destination places (Machado et al., 2018). Shared mobility can be used as an "umbrella" term that refers to a broad array of innovative transportation modes with different use cases, business models, and travel behavior (Shaheen et al., 2017).

Ride-hailing services is a branch of shared mobility which refers to carsharing. According to global economy reports, post-COVID-19, the global ride-hailing market size is projected to grow at a rate of 55.6% from 2020 to 2021. It is expected to reach USD 117.34 billion by 2021 from USD 75.39 billion in 2020. However, the projection for 2021 is estimated to be down by 2% compared to pre-COVID-19 estimation (Shaheen et al., 2015). Adopting strategies such as providing barriers between driver and passenger, equipping the vehicle with sanitizers, and installing digital thermometers to measure passengers' body temperature in order to eliminate the virus infection threat may rejuvenate the ridesharing market. However, the added precautionary measures can cause surge in fares. Therefore, transportation sectors observe a trend-inflicting change in travelers and driver's mode choice behavior due to irregular events such as

* Corresponding author.
E-mail addresses: smors005@fiu.edu (S.A. Morshed), sk310@zips.uakron.edu (S.S. Khan), rtanv003@fiu.edu (R.B. Tanvir), shafkathnur@iut-dhaka.edu (S. Nur).

https://doi.org/10.1016/j.jum.2021.03.002
Received 13 November 2020; Received in revised form 4 February 2021; Accepted 4 March 2021
Available online 20 March 2021

technology emergence, pandemic conditions etc. (Khoury, Amine, & Abi Saad, 2019).

Virtual social interaction platforms are collectively called social media, e.g., Facebook, Twitter, Instagram, etc. In other words, social media are a group of internet-based applications that allow the creation and exchange of user-generated content (Chi, 2008). The usage of social media data in various research studies, which requires public opinion, has created a paradigm shift in modern business models' research and development arena. Social media content includes personal statement or opinion, which can usually complement, enrich, or even replace traditional surveying data (Qi & Costin, 2019). Information collected from social media can become an important tool to illustrate public opinion on the socio-economic subject matter.

Globally 72% of the population use web services, whereas 89% of internet users are from the USA (Poushter, 2016). Interestingly, 76% of global internet traffic is on social media sites. This emergence of social media platforms and increasing engagement of people with social media have created a unique opportunity for transportation service providers to collect real-time information from social media users with minimum spending (Qi et al., 2020).

Twitter, a micro-blogging site, is one of the largest social media platforms where people express their opinions and feelings about various topics (Incredible and Interestin, 2020). Twitter has 330 million monthly active users with 1.3 billion accounts, where 500 million tweets (micro-blogs consist of 240 characters) are published every day. There are 50 million active monthly users in the US, which accounts for 23.64% of US adults. Recent studies have explored the potential of social media data for transportation planning, traffic prediction, real-time traffic management, and traffic information dissemination (Qian, 2016; Zhang et al., 2016). So, social media data creates an excellent opportunity to understand the public sentiment, emotional, and behavior patterns towards ride-hailing service. This study chose Twitter data for analysis over the other social media platforms such as Facebook or Instagram due to its flexible communicative structures. Twitter users searching for specific topics can easily find it through the rapid and ad hoc establishment of shared hashtags related to any particular event. Hashtags provide a mechanism for conversation, update threads with trending topics among users in the Twitter network (Bruns & Liang, 2012). Therefore, large-scale Twitter data can be used as a source of information to understand the ride-hailing service users' behavioral patterns by analyzing sentiment and emotions about a specific topic such as shared mobility or ridesharing.

To better understand the social and economic impact of COVID-19 on shared mobility, this research focused on analyzing large-scale twitter data during the global pandemic. This study aims to determine ride-hailing service users' behavioral pattern by analyzing sentiment and emotions from their perspectives. For better accuracy of Twitter data collection, the geolocation of the data extraction has been set to Florida, USA. Additionally, a small survey was conducted via Qualtrics in the Miami Beach area of Florida to understand the impact of COVID-19 on TNC usage.

The following sections of the paper include background, methodology, result analysis and the limitation of the study. The background section discusses about the past studies and existing conditions of ride-hailing service. This section also discusses about the novelty of this research. The methodology section discusses about development of the questionnaire survey, the emotion detection model and framework of the entire methodology including data collection and data pre-processing. The result section includes the detailed comparative sentimental analysis and in the final section limitations and future recommendations are discussed.

2. Background

Ride-hailing service providers or TNCs such as Uber and Lyft mainly provide access to commuters where individuals can track real-time availability and location of the vehicles and access them through commercial mobile application software. TNCs are expanding due to the increasing use of the internet, development in The Internet of Things (IoT) services and developing Global Positioning System (GPS).

In the USA, TNCs generate the lion share of their gross income from trip bookings in large metropolitan areas, including trips to and from airports. Uber, a global market leader in ride-hailing service, complete 14 million individual trips per day in the US (65% of the total market share). Uber has the archetypal disruptive business model and certainly played havoc with the taxi industry in major cities worldwide by easy accessibility and minimum fare (Iqbal, 2020). Studies show that 82% of US Uber customer has an age range from 16 to 44 years. 46% and 48% of their customer are from urban and suburban region respectively (McGrath, 2019). Interestingly the same study reported that commute related to business (12%) and tourism (77%) in Miami are Uber-based. For instance, in 2019, large metropolitan areas and tourists enriched areas such as Chicago, London, Los Angeles, New York City, San Francisco Bay Area, and Miami, Uber has earned 23% of its gross ride bookings (Uber Announces Results fo, 2019).

This research study is focused on Florida as it comprises of both commercial and tourist regions. Altogether, 10% of Uber trips in Florida are business-related. Overall, 14.5% of Uber trips in Florida are commuted by tourists, and 27.3% of tourists have reported an increase in spending during their trip because Uber enabled them to visit additional locations (Correoso, 2018).

The COVID-19 outbreak has caused operational disruption for the ride-hailing service providers because the hub of TNCs, i.e. the large metropolitan cities, has been in lockdown (Beck & Hensher, 2020). The number of trips (US) for Uber has fallen to 70% (Hawkins, 2020). During mid-March, Uber announced that rides were down 60–70% in Seattle, an area impacted early by COVID-19. Similarly, trip rate of Uber declined by 77% in two of its largest European market, i.e. London and Paris (SERAFIMOVA, 2020). COVID-19 has a profound effect on the global economy and transportation network. Service providers need a high degree of creativity in responding to this crisis (SERAFIMOVA, 2020). It is expected that the COVID-19 pandemic will cause a paradigm shift in travel behavior, psychological behavior due to hygiene concerns, and the financial capabilities of the customers. Hence, it is essential to understand and assess the behavioral change in customers' sentiment and emotion regarding management, security, monetization, and policy. As a result, the coping mechanism of TNCs in post-COVID-19 scenarios is of paramount importance to ensure a sustainable future for the ride-hailing service.

Sentiment analysis is a process that automates the mining of sentimental data like attitudes, opinions, views, and emotions from a text or speech. It involves understanding and classifying opinions in text into categories like “positive” or “negative” or “neutral”. Sentimental analysis can be categorized into four levels: (a) Word-level (b) Sentence level (c) Document-level (d) Feature level (A. and Sonawane, 2016). Sentiment analysis began as a document level classification task (Pang & Lee, 2004; Turney, 2002); it has been advanced to the sentence level (Hu & Liu, 2004; Kim & Hovy, 2004) and, more recently, at the word level (Agarwal et al., 2009; Wilson et al., 2005). Instances of sentiment analysis in Twitter include the use of distance learning to acquire sentiment data and analyzed based on emoticon (Go et al., 2009) and also classification of tweets through the implementation of supervised machine learning (Liang & Dai, 2014; Pak & Paroubek, 2010; Xia et al., 2011). Hieratical classification framework, machine learning models (Naïve Bayes), LDA model, topic modeling, lexicon-based modeling (Nafis et al., 2019; Qi et al., 2020; Ye et al., 2019; Zou et al., 2018) are recently used methods to extract sentiments and emotions from tweets.

The fundamental problem of sentiment analysis is that it only indicates whether the public reaction is positive or negative but fails to decipher the exact feelings or intensity of the users’ responses. Emotion analysis, on the other hand, provides a more profound analysis than sentimental analysis. Classification of tweets based on the Ekman model (Balabantaray et al., 2012), which is based on six basic emotions, focuses on mining emotions from texts. Past studies (Alm et al., 2005; Gaind et al., 2019; Morshed et al., 2020, pp. 427–435) on emotion detection mainly relied on the manual annotation of a small size dataset with a limited scope. However, lack of manually annotated data to train classifiers for labeling into different emotion categories and unavailability of a comprehensive bag of emotion-words and emoticons are notable barriers to emotion analysis via machine learning. Therefore, it is essential to annotate a reliable dataset for classifying and labeling tweets according to sentiment and emotion.

3. Objective of the study

The aim of this study is to conduct a comparative sentiment-emotion analysis of large-scale Twitter data related to the usage of shared mobility in the course of the COVID-19 pandemic. Additionally, this study proposes a state-of-the-art data pre-processing framework, a three-layered sentiment-emotion detection model. Finally, this study forms a large-scale manually annotated twitter dataset in pre-pandemic and during pandemic periods to understand the impact of COVID-19 on the usage of shared mobility.

4. Methodology

4.1. Survey

The utility of ride-hailing services depends on the vast array of socio-demographic variables. This study consists of a short state-preference survey to collect and document perceptions and preferences related to ride-hailing services in the Miami South Beach region. The survey was developed using the Qualtrics Research Core tool which follows the Institute of Transportation Engineers Manual on Transportation Engineering Studies guidelines (Hummer et al., 2011). The stated preferences surveys sought to get information about commuters’ usage of TNC along with detailed socio-demographic such as age, gender, education level etc. before and amid the COVID-19 pandemic. The data collection period was May 2020. After eliminating incomplete, duplicate, or irregular answers, 339 responses (21% residents and 79% tourists) from the Miami South Beach area were analyzed. The questionnaire survey consisted of questions such as trip purpose and frequency of ride-hailing service usage before and during the COVID-19 pandemic. The socio-demographic characteristics of the respondents are summarized in Table 1.

Table 1
Socio-demographic characteristics of the survey respondents.

Category	Range or Type	Percentages
Age	>50	13%
	30–49	32%
	18–29	50%
	<18	5%
Education Level	High School or less	16%
	College Graduate	38%
	Graduate School	21%
	Some college credit, no degree	26%
Gender	Female	60%
	Male	39%
	Others	1%
Annual Income	>75,000	27%
	$30,000-$74.99,000	45%
	< $30,000	29%
Race	White	53%
	Black or African American	29%
	Asian	6%
	American Indian or Alaska Native	3%
	Native Hawaiian or Pacific Islander	0%
	Others	8%

4.2. SENTIMENT-EMOTION detection model

Most emotion detection and analysis systems use Ekman's emotion model (Agarwal et al., 2019), which lists six basic human emotions – Anger, Disgust, Fear, Joy, Sadness, and Surprise. However, it is a challenge to extract emotions due to the presence of the emoticons and abstract structure of the tweets. In this study, a unique human feelings extraction model called Sentiment-Emotion Detection (SED) model (Fig. 1) has been proposed, which classifies tweets in three layers to extract a complete sequence of sentiment and emotion along with the statement verdict of the tweet. The model presents a decision-making procedure encapsulating eleven sub-components, forming a complete sequence of sentiment-emotion analysis for social media reactions or responses. Existing sentiment or emotion analyzing models are unidimensional, which often fails to capture the complete expression of tweets from users' perspective.

4.3. Framework

Fig. 2 represents the full framework developed, which covers data collection from Twitter to decision categorization of the collected tweets. The framework is based on four major sections; namely, (a) tweet collection through (Tweepy Application Programming Interface or API), (b) Temporal classification of tweets, (c) Pre-processing of tweets, (d) Manual annotation and labeling of tweets to fit proposed Sentiment-Emotion Detection Model. Labeled data obtained at the last phase of the process can be used as a feeder for the machine learning procedure for further analysis. Data storage is a crucial part of the framework as it is linked with each step of the process. This framework can be integrated with other social media application.

4.4. Data collection

To collect tweets, the Twitter premium package, Tweepy search full archive API was used. The package consisted of options for 100 requests per month and each request contained 500 tweets. To gather public opinions from tweets, parameters such as hashtags (e.g. #uber, #lyft, #ride, etc.) and keywords (Uber, Lyft, ride, etc.) related to ride-hailing services were used. For better accuracy of data, geolocation of the data collection hub has been set to the state of Florida, USA. There were about 25 requests put through for data collection. In total, 12,500 tweets were collected based on the assigned parameters through Tweepy API.

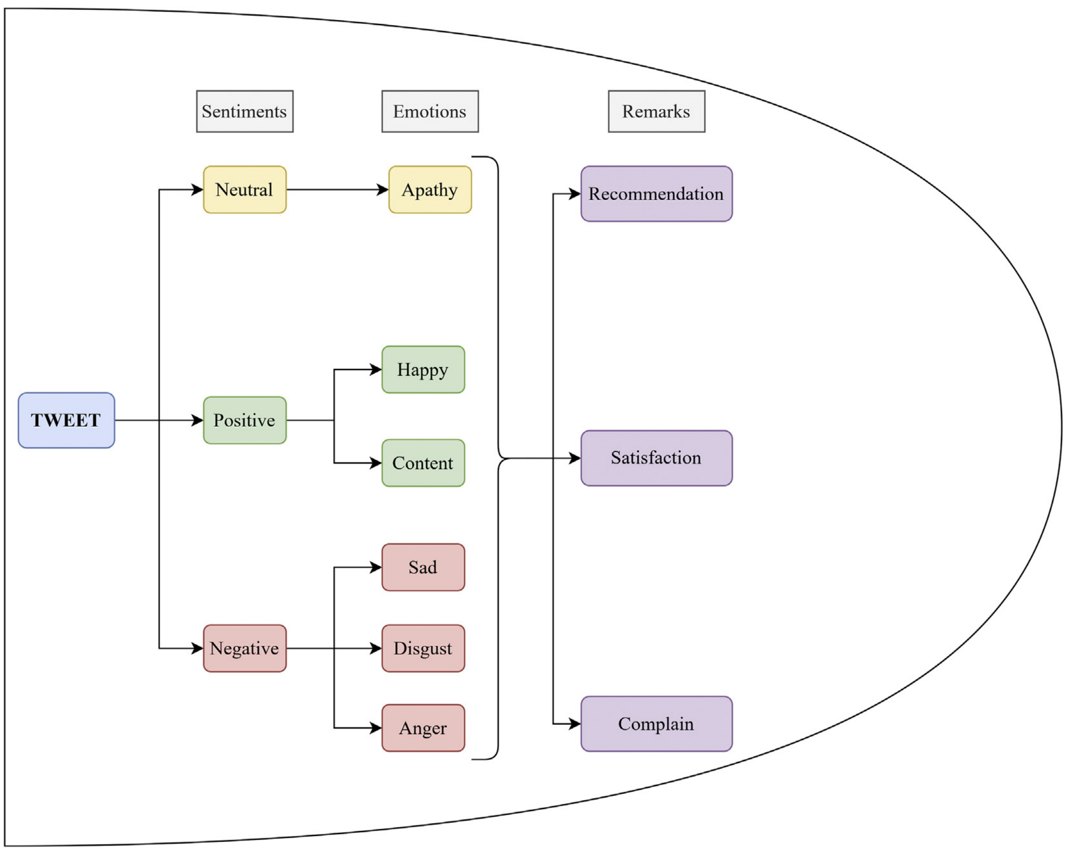

Fig. 1. Sentiment-emotion detection model.

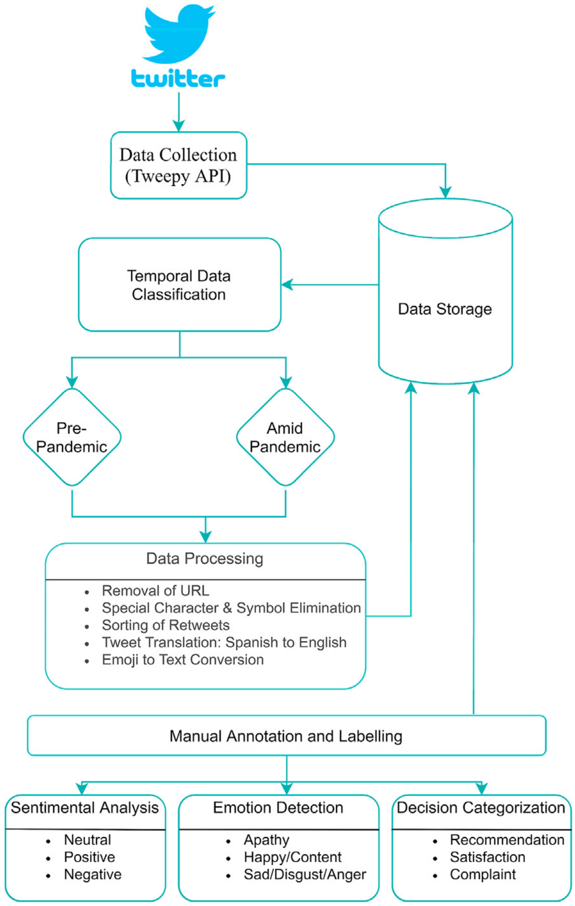

Fig. 2. Framework of tweet classification and sentiment-emotion detection and analysis.

4.5. *Temporal classification*

The data collection process was divided into two different timelines to represent pre-COVID-19 and during COVID-19 periods. Pre-COVID-19 data was collected every day from February 1st to April 30th in 2019 at 2 days interval. Similarly, data was collected during the same time for the year 2020, which was labeled as Amid-COVID-19 tweets. Table 2 shows instances of tweets related to ride-hailing services, TNCs collected through Twitter API in different timelines.

4.6. *Data pre-processing*

Twitter Data pre-processing is a stepwise function. The tweets collected required a pre-processing pipeline and as a result, the platform of python was used for data cleansing. Data pre-processing comprises of the following steps for further analysis:

1. Removal of the mentions of other Twitter IDs by removing the words started with '@', digits, and any URL links.
2. Removal of special symbols, characters, and words such as, "" …: '? !; # $ % & ( ) * + -/< > = [ ] n ^ _ { } | ~. etc.
3. Removal of URL and hyperlinks by using a regular expression finder tool, regex.

Table 2
Instance of tweets based on the temporal classification.

Date	Tweet Collected Using Hashtags and Keywords	Temporal Classification
April 13, 2019	While @Uber says it did perform a background check, Victores' last known felony conviction was in 2001. Currently, Florida law says rideshare drivers with felony convictions more than 5 years old can still be hired … #rideshare #uber #uberdriver	Pre-COVID-19
February 2, 2020	Some Uber drivers are refusing to pick up airport passengers because of coronavirus fears	During COVID-19

4. Conversion of emojis into a mapped word or words by using Python package emoji (Kim & Wurster, 2020) and demoji (Solomon, 2020).
5. Detection and translation of tweets from other languages by using google translator API wrapper package googletrans (Han, 2020) in Python. Tweets of other languages were removed.

4.7. Manual annotation and labeling

The pre-processing step reduced the size of the dataset to 6460 (Amid-COVID-19) and 6048 (Pre- COVID-19) relevant tweets. Pre-processed or 'cleaned' tweets were annotated manually and labeled in three different categories following the Sentiment-Emotion Detection Model. Initially, the tweets were evaluated to determine their relevance to shared mobility, ride-hailing services, and TNC. The 'On-Topic' relevant tweets were then subjected to sentiment analysis. The tweets were either labeled as positive, negative, or neutral based on the sentimental value of the tweet from the users' perspective. Sentiment analysis generally focuses on creating or mining a list of words or emojis associated with strongly positive or negative sentiment. The challenge for analyzing sentiments is that few negative words can indicate positive sentiment, while few positive words indicate negative sentiment based on the context. As a result, manual annotation is a time-consuming process, which required focus and contextual judgment for analyzing the tweets. In the next step, the sentimentally annotated tweets were further analyzed and labeled on the users' value of emotions expressed. All positive tweets were annotated with either 'happy or 'content' emotion. Similarly, all negative tweets were labeled as either 'sad', 'anger', or disgust. The remaining neutral tweets were labeled as 'apathy' tweets. Finally, after the sentiment and emotions were assigned to the tweets, each tweet was categorized based on a decision explained in the model. Manual annotation and labeling of large-scale twitter alleviate data sparsity for future sentiment analysis and emotion detection. A group of 20 English speakers from two different regions of the world (the United States and Bangladesh) possessing a morphological, syntactic, and semantic understanding of the language has annotated the tweets based on the Sentiment-Emotion Detection Model. The dataset obtained can be subjected to future analysis to determine the impact of human reaction and responses on shared mobility in standard and remarkable conditions by implementing machine learning or deep learning procedure (Morshed, 2021).

5. Results and discussion

5.1. Survey result

A stated preference survey conducted in the Miami Beach region questioned residents and tourists regarding their choice of a mode other than a personal vehicle in the pre-COVID-19 period. Fig. 3 shows that 67% of the respondents preferred ride-hailing services over other available modes such as public transit (6%) and organized ridesharing programs (11%).

Fig. 4 illustrates the trip purposes of the TNC users during weekdays and weekends. Since Miami is a tourism-prone area, weekend TNC trips are as popular as weekdays trips. The majority of the TNC trips are directed to restaurants and downtown for dining out and attending nightlife. This indicates that the TNC trips were destined for a social gathering in this region.

Fig. 5 shows the frequency of TNC usage before and during the pandemic period. It was evident that the use of TNC reduced after the pandemic has commenced. Amid the pandemic, the lockdown was not announced when the survey was conducted. Declination in the TNC usage indicated that COVID-19 hurt the TNC usage among the residents and tourists in the Miami Beach region.

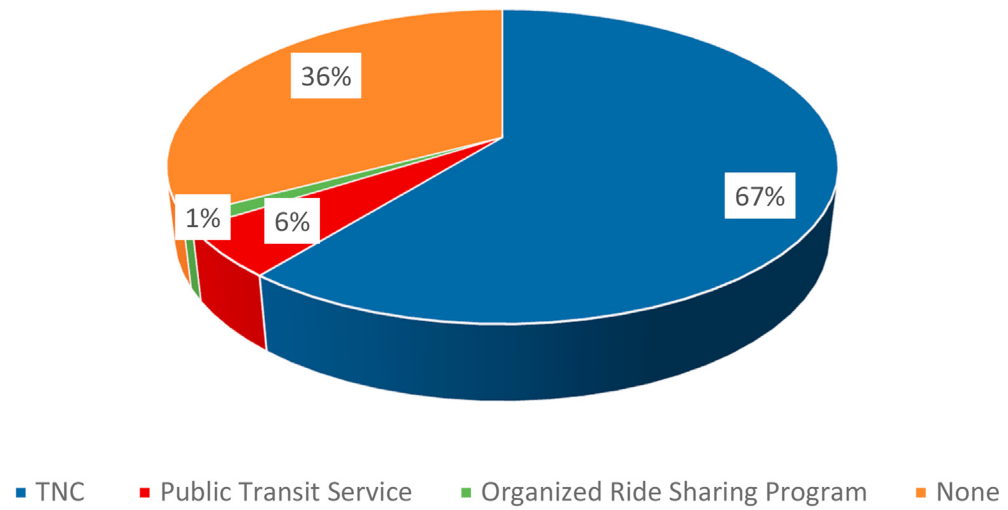

Fig. 3. Pre-COVID-19 mode choice.

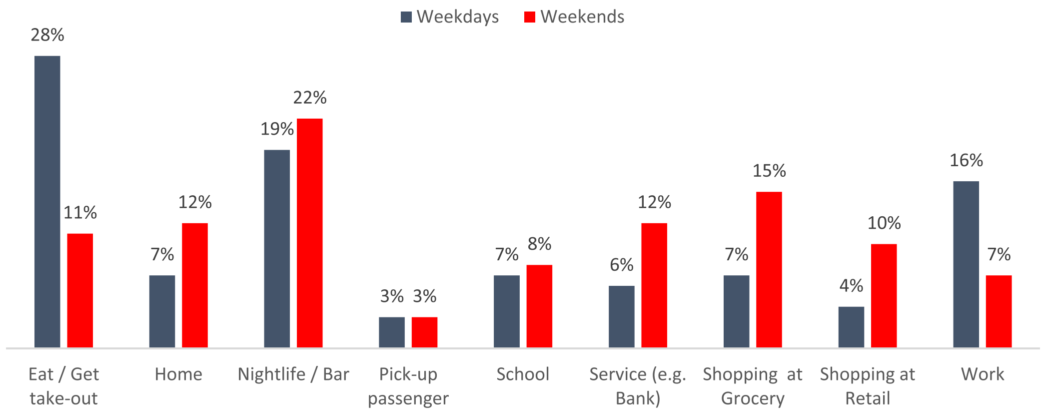

Fig. 4. Trip purpose of TNC trips.

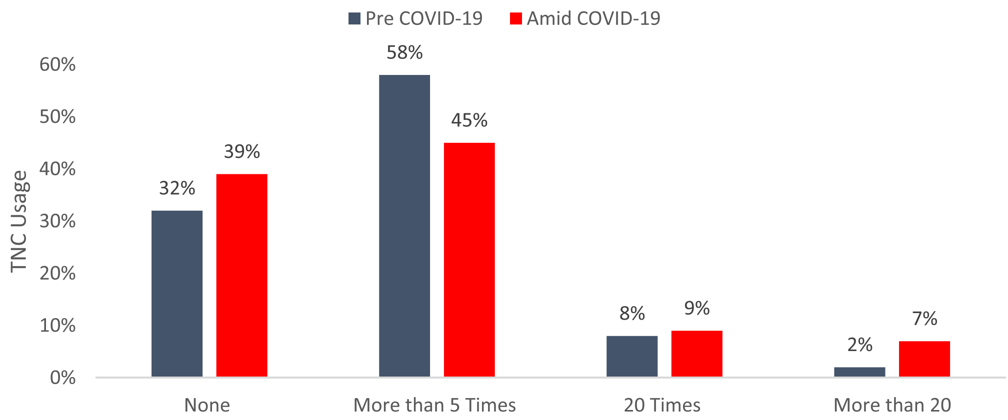

Fig. 5. Frequency of TNC trips.

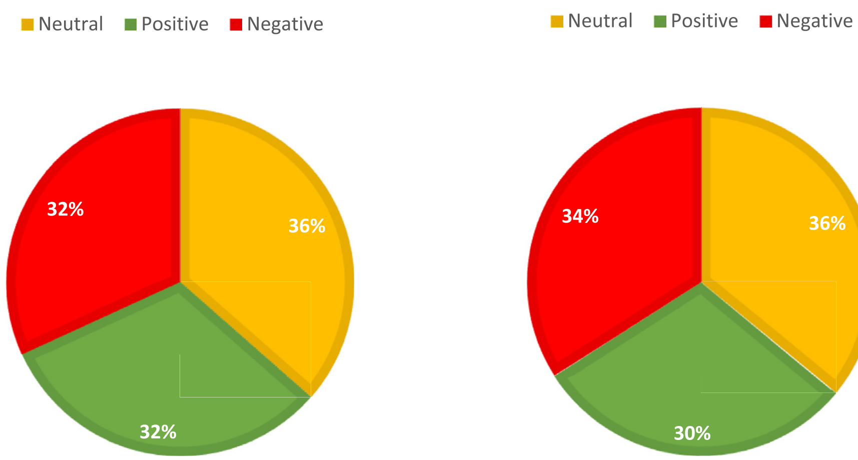

Fig. 6. (a) Pre-COVID-19 sentiment distribution, (b) amid-COVID-19 sentiment distribution.

5.2. SENTIMENT-EMOTION detection and analysis

The first layer of the sentiment-emotion detection model determines the sentiments within the tweet collected in both periods. Fig. 6 (a) shows that in the pre-COVID-19 period, the majority of the tweet sentiments (36%) were neutral. Interestingly, there was an equal share of (32%) of positive and negative tweets. Fig. 6 (b) illustrates that amid the COVID-19 period, more negative tweets (34%) were analyzed than positive tweets (30%). However, the percentage of neutral tweets remained unchanged (36%). Hence, a slight (2%) change of trend in sentiment from positive to negative on shared mobility topics during the early pandemic stage was observed. After completion of the first layer of analysis, the labeled tweets were subjected to the next layer (emotion) analysis.

Fig. 7 illustrates the six different emotions associated with the sentiment of the tweets in the pre-COVID-19 period. 59% of the positive tweets are labeled as happiness and 41% as content, respectively. Before the pandemic, negative tweets were mostly labeled as sad (44%) whereas 10% and 8% of the TNC users' expressed anger and disgust, respectively. Fig. 8 illustrates the emotion distributed

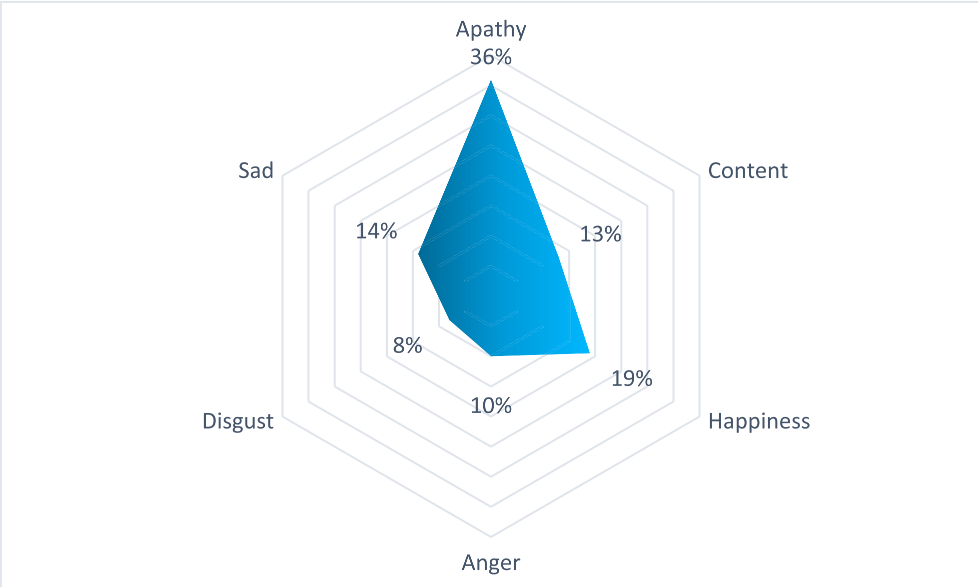

Fig. 7. Pre-COVID-19 emotion radar.

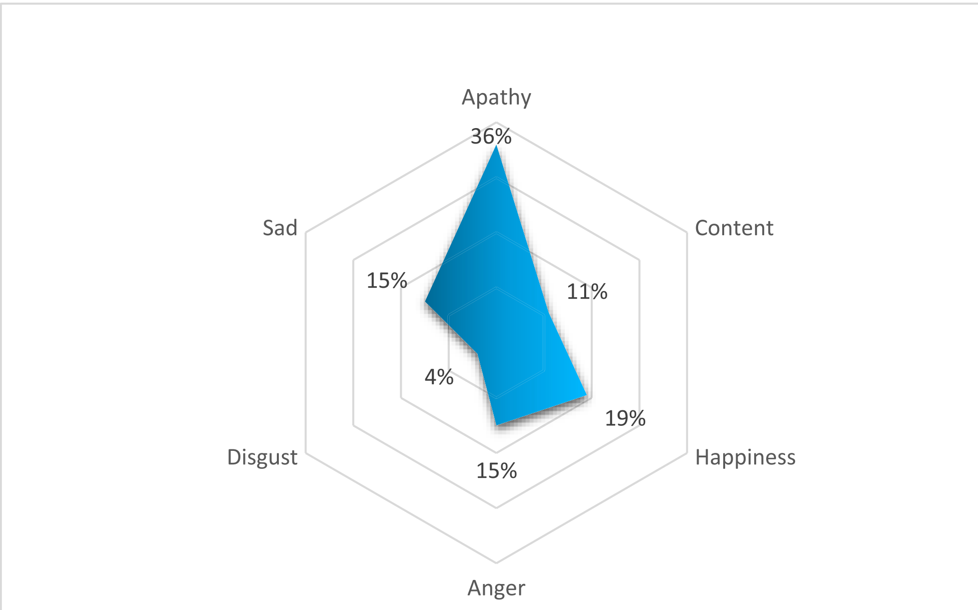

Fig. 8. Amid-COVID-19 emotion radar.

among the tweet collected amid the pandemic. From observation, it is clear that the change of trend in emotion among the Twitter users are located in disgust (4%) and anger (15%) sections. There is a slight change in the content-based emotion (11%) during the pandemic. However, apathy-based tweets in both periods remained the same (36%).

The shift in the TNC business model and consumer behavior passed through sentimental and emotional change. To better understand the trend of change regarding social media reaction and response, the large-scale Twitter data were further categorized into three different remarks as shown in Fig. 9. It can be observed from the bar chart diagram in Fig. 9 that complaints have increased during pandemic (40%) than the pre-pandemic period (34%). However, the satisfaction of TNC users has increased in the post-pandemic period (44%) than the pre-pandemic period (33%). The reason behind this change of behavior from the TNC users can be due to several reasons adopted by the ride-hailing service providers such as taking precautions against the coronavirus, more availability of the car, increased discount or decreased fare. Additionally, decreased congestion taking less travel time to reach destinations during the pandemic period can be another reason for posting satisfactory tweets from TNC users. Twitter users also expressed more concerns (33%) regarding the ride-hailing service and the drivers in the pre-pandemic period than amid pandemic (16%) by suggesting recommendations for improvement.

5.3. Word cloud

Fig. 10 shows an infographic that visualizes word cloud and word distribution frequency collected from the large-scale Twitter data. Top words such as “Uber”, “Rideshare”, “Drivers”, “Lyft” etc. were the most popular words generated within the tweets collected. The

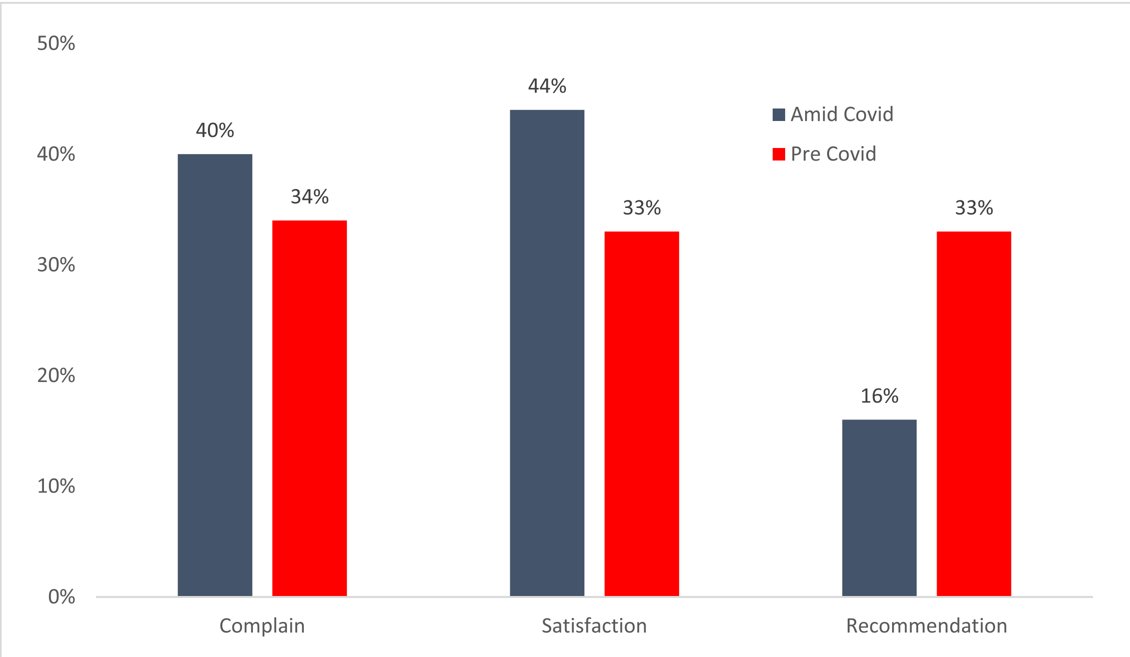

Fig. 9. Decision categorization in pre-COVID-19 and amid COVID-19 pandemic.

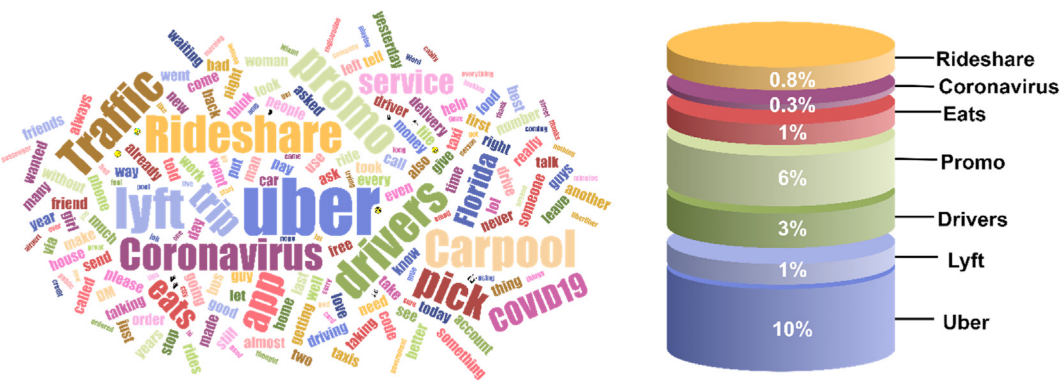

Fig. 10. Word cloud and frequency.

word “Coronavirus’ was found 0.3% in the dataset collected during the pandemic period. Common words associated with Uber and Lyft were ‘drivers’, ‘Uber-Eats’, ‘promo’, ‘service’ etc. Most of the tweets associated with these words were labeled as neutral tweets expressing an emotion of “apathy”. Tweets related to coronavirus such as ‘*Uber suspends user accounts of Mexicans after drivers possibly exposed to coronavirus …* ’ were of negative sentiment with either ‘sad’ or ‘anger’ emotion. Also, the remarks for all those tweets were labeled as either ‘dissatisfaction’ or ‘complaint’.

6. Limitation and future work

The collection of large-scale Twitter data requires Twitter API access, which is an expensive procedure. As a result, this study was only able to collect data till April 2020. Data beyond the month of April would have created a deeper pool of datasets.

An alternate neural network model using a Multilayered Perceptron to perform multi-class classification task is the next step of this research. The authors have already started developing the neural network model. It has eight hidden layers including five dense layers and three drop-out layers. A publicly available ‘crowdflower’ dataset (Data World) consisting 13 labels of sentiments and emotions has been used to train the model. Implementing an unsupervised algorithm, the training model achieved an accuracy of 94–96% for sentimental analysis. However, in the model’s testing, accuracy has been dropped to 37%, which indicates that the model has significant overfitting. Currently, the authors are working to resolve the overfitting issue of the model by improving the size and quality of the training dataset.

7. Conclusion

COVID-19 pandemic possesses a significant impact on the global economy and transportation network. Due to the COVID-19 crisis, consumers’ reaction has become an essential requirement for any product or service. Ride-hailing service providers are at stake for ongoing lockdown in major cities globally and massive change in worldwide travel behavior. Collecting real-time information and studying the behavioral pattern of customers from social media platforms, creates a unique opportunity for TNCs to establish global engagement and communication, which helps to improve their existing business model. Manual annotation of large-scale Twitter data related to shared mobility and ride-hailing services developed in this study builds a profound data-library for future usage. The proposed ‘Sentiment-Emotion Detection’ model adopts a simplified yet robust and synergistic framework to label and categorize tweets based on consumers’ basic feelings. During the COVID-19 pandemic, opinions and expressions of TNC users on Twitter were inclined towards negative sentiment and emotions. However, positive remarks of satisfaction regarding ride-hailing service during the COVID-19 pandemic suggest that commuters and TNC users still possess a positive mindset towards ride-hailing services. Therefore, for sustainability, constant revenue generation, and policy development in the ride-hailing service domain, it is important to understand public sentiment and reactions.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design—Syed Ahnaf Morshed, Sifat Shahriar Khan; data collection—Syed Ahnaf Morshed, Raihanul Bari Tanvir, Sifat Shahriar Khan; analysis and interpretation of results— Syed Ahnaf Morshed, Sifat Shahriar Khan, Shafkath Nur; draft manuscript preparation— Syed Ahnaf Morshed, Shafkath Nur. All authors reviewed the results and approved the final version of the manuscript.

Declaration of competing interest

We know of no conflicts of interests associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome. As Corresponding Author, I confirm that the manuscript has been read and approved for submission by all the named authors.

References

60 Incredible and Interesting Twitter Stats and Statistics. (2020). *Brandwatch*. https://www.brandwatch.com/blog/twitter-stats-and-statistics/#:%7E:text=Twitter%20user%20statistics,users%20write%2080%25%20of%20tweets.

A, V., & Sonawane, S. S. (2016). Sentiment analysis of twitter data: A survey of techniques. *International Journal of Computer Application, 139*(11), 5–15. https://doi.org/10.5120/ijca2016908625

Agarwal, A., Biadsy, F., & Mckeown, K. (2009, March). Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In *Proceedings of the 12th conference of the European chapter of the ACL (EACL 2009)* (pp. 24–32).

Agarwal, A., Xie, B., Vovsha, I., & Owen Rambow, R. P. (2019). Sentiment analysis of twitter data. *International Journal of Healthcare Information Systems and Informatics, 14*(2), 1–16. https://doi.org/10.4018/ijhisi.2019040101

Alm, C. O., Roth, D., & Sproat, R. (2005). October). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 579–586).

Balabantaray, R. C., Mohammad, M., & Sharma, N. (2012). Multi-class twitter emotion classification: A new approach. *International Journal of Applied Information Systems, 4*(1), 48–53. ISSN: 2249-0868.

Beck, M. J., & Hensher, D. A. (2020). Insights into the impact of COVID-19 on household travel and activities in Australia–The early days of easing restrictions. *Transport Policy, 99*, 95–119. https://doi.org/10.1016/j.tranpol.2020.08.004

Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing twitter data during natural disasters. *First Monday, 17*(4). https://doi.org/10.5210/fm.v17i4.3937

Chi, E. H. (2008). The social web: Research and opportunities. *Computer, 41*(9), 88–91. Collins, C., Hasan, S., Ukkusuri, S.V., (2013). A novel transit rider satisfaction metric: rider sentiments measured from online social media data. J. Public Transp. 16 (2). http://doi.org/10.5038/2375-0901.16.2.2.

Correoso, J. (2018, July 26). *Uber's economic impacts in Florida. Uber newsroom*. https://www.uber.com/newsroom/florida-economic-impact/.
CrowdFlower. (2016). *CrowdFlower*. https://data.world/crowdflower/sentiment-analysis-in-text.
Gaind, B., Syal, V., & Padgalwar, S. (2019). *Emotion detection and analysis on social media*. arXiv preprint arXiv:1901.08458.
Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report. *Stanford, 1*(12), 2009.
Han, S. (2020, January 14). *Googletrans 3.0.0. pypi.org*. https://pypi.org/project/googletrans/.
Hawkins, A. J. (2020, March 19). *Uber is doing 70 percent fewer trips in cities hit hard by coronavirus. The Verge*. https://www.theverge.com/2020/3/19/21186865/uber-rides-decline-coronavirus-seattle-sf-la-nyc.
Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177). https://doi.org/10.1145/1014052.1014073
Hummer, J. E., Robertson, H. D., & Nelson, D. C. (2011). *Manual of transportation engineering studies. Institute of transportation engineering*. Englewood Cliffs, New Jersey: Prentice-Hall, inc.
Iqbal, M. (2020, June 24). *Uber revenue and usage statistics*, 2020. Business of Apps https://www.businessofapps.com/data/uber-statistics/.
Khoury, J., Amine, K., & Abi Saad, R. (2019). An initial investigation of the effects of a fully automated vehicle fleet on geometric design. *Journal of Advanced Transportation, 2019*, Article 6126408. https://doi.org/10.1155/2019/6126408.
Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. In *Coling 2004: Proceedings of the 20th international conference on computational linguistics* (pp. 1367–1373).
Kim, T., & Wurster, K. (2020, August 2). *Emoji 0.6.0. pypi.org*. https://pypi.org/project/emoji/.
Liang, P.-W., & Dai, B.-R. (2014). Opinion mining on social MediaData. In *IEEE 14th international conference on mobile data Management,Milan, Italy, june 3 - 6, 2013* (pp. 91–96). https://doi.org/10.1109/MDM.2013. ISBN: 978-1-494673-6068-5.
Machado, C. A. S., Hue, N. P. M., de, S., Berssaneti, F. T., & Quintanilha, J. A. (2018). An overview of shared mobility. *Sustainability, 10*(12), 1–21. https://doi.org/10.3390/su10124342
McGrath, F. (2019, November 27). *The demographics of uber's US users: GlobalWebIndex. GWI*. https://blog.globalwebindex.com/chart-of-the-day/uber-demographics/.
Morshed, S. A., Xuan, L. V., & Tanvir, R. B. (2020). *Network-based information extraction from IFC files to support intelligent BIM companion (IBcom) technology, ASCE construction research congress 2020*. Computer Applications.
Morshed, Syed, et al. (2021). 8R Resilience Model: A stakeholder-centered approach of disaster resilience for transportation infrastructure and network. *Transportation Engineering, 4*. https://doi.org/10.1016/j.treng.2021.100058.
Nafis, S. R., Alluri, P., Jung, R., Ennemoser, R., & Gan, A. (2019). *A comprehensive review of states' existing practices in using communication technologies to increase public involvement for transportation projects*. Transportation Research Board 98th Annual MeetingTransportation Research Board, 19-01674.
Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. *LREc, 10*(2010), 1320–1326.
Pang, B., & Lee, L. (2004). *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*. arXiv preprint cs/0409058.
Poushter, J. (2016, February 22). Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center's Global Attitudes Project*. https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-Internet-usage-continues-to-climb-in-emerging-economies/.
Qian, Z. (2016). *Real-time incident detection using social media data*. Www.Penndot.Gov. https://www.penndot.gov/ProjectAndPrograms/Planning/Research-And-Implementation/Documents/Real%20Time%20Incident%20Detection%20Using%20Social%20Media%20Data.pdf.
Qi, B., & Costin, A. M. (2019). Investigation of the influence of Twitter user habits on sentiment of their opinions towards transportation services. In *Computing in civil engineering 2019: Smart cities, sustainability, and resilience* (pp. 314–321). Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784482445.040.
Qi, B., Costin, A., & Jia, M. (2020). A framework with efficient extraction and analysis of Twitter data for evaluating public opinions on transportation services. *Travel Behavior and Society, 21*, 10–23. https://doi.org/10.1016/j.tbs.2020.05.005. December 2019.
Serafimova, T. (2020). *COVID-19: An opportunity to redesign mobility towards greater sustainability and resilience?*. https://doi.org/10.2870/695530
Shaheen, S., Bell, C., Cohen, A., & Yelchuru, B. (2017). *Travel behavior: Shared mobility and transportation equity (No. PL-18-007)*.
Shaheen, S., Chan, N., Bansal, A., & Cohen, A. (2015). *Shared mobility: A sustainability & technologies workshop: Definitions, industry developments, and early understanding*.
Solomon, B. (2020, August 30). *Demoji 0.3.0. pypi.org*. https://pypi.org/project/demoji/.
Turney, P. D. (2002). *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. arXiv preprint cs/0212032.
Uber Announces results for fourth quarter and full year 2019.(2019). Retrieved November 04, 2020, from https://investor.uber.com/news-events/news/press-release-details/2020/Uber-Announces-Results-for-Fourth-Quarter-and-Full-Year-2019/.
Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347–354).
Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*(6), 1138–1152. https://doi.org/10.1016/j.ins.2010.11.023
Ye, Q., Chen, X., Zhang, H., Ozbay, K., & Zuo, F. (2019). Public concerns and response pattern toward shared mobility security using social media data. In *2019 IEEE intelligent transportation systems conference, ITSC 2019* (pp. 619–624). https://doi.org/10.1109/ITSC.2019.8917010
Zhang, Z., Ni, M., Gao, J., & He, Q. (2016). *Mining transportation information from social media for planned and unplanned events*.
Zou, L., Lam, N. S. N., Cai, H., & Qiang, Y. (2018). Mining twitter data for improved understanding of disaster resilience. *Annals of the Association of American Geographers, 108*(5), 1422–1441. https://doi.org/10.1080/24694452.2017.1421897