2023 IEEE 26th International Conference on
Intelligent Transportation Systems (ITSC)
24-28 September 2023. Bilbao, Bizkaia, Spain

Whose Attitudes Toward Transit Are Most Affected by Rising Subway Crimes in New York City?

Xu Chen, Yishi Wang, Xuan Di *Member, IEEE*

2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC) | 979-8-3503-9946-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITSC57777.2023.10421971

***Abstract*— This paper aims to understand the public attitude toward crimes in the subway system using social media data. In particular, we focus on the users who tweeted about New York City (NYC)'s subway crimes in the past two years. The association between the sentiment of the tweets and the demographics of the users is analyzed to better help us understand which population are most affected by the rising subway crimes. We analyze textual contents of Tweets and perform sentiment analysis with a transformer-based tool to gain insight into public attitudes and concerns regarding crimes in the NYC subway system. Negative sentiments, which arise from crimes in subway system, indicate a sense of fear among the public and can consequently result in a decrease in subway ridership. We also accompany such analysis by regression between Twitter users' sentiment and their demographics to better understand whose attitudes toward taking public transit are most affected by these crimes. Such understanding could identify the population who has the most severe concerns about transit safety, and how these concerns affect their willingness to take subway. We find that age, gender and ethnicity are important factors influencing people's sentiments toward crimes. Compared to others, women have more concerns about their safety while riding public transportation in NYC. Such analysis would help better understand the association between rising crimes and people's attitudes toward taking public transit, so as to assist transit planners and policymakers to urgently create a safer environment to recover low transit affected substantially by the pandemic.**

Key-words: Social Media, Crime, Subway

I. Introduction

This paper aims to understand the public attitude toward crimes in the subway system using social media data. In particular, we focus on the users who tweeted about New York City (NYC)'s subway crimes in the past two years. The association between the sentiment of the Tweets and the demographics of the users is analyzed to better help us understand which population is most affected by the rising subway crimes.

Violent crimes in NYC subway system have more than doubled since 2019 [1]. A mass shooting that resulted in 29 injuries took place on the New York City (NYC) subway in Brooklyn on April 12, 2022, sparking heated conversations about gun violence. Police statistics show that subway crime had increased by 68% as of April 2022 compared to that in the same period of 2021 [2]. The general public has accessible information about reported crimes through social media. However, it is worth noting that this information is predominantly presented in a negative manner, particularly in the aftermath of such crimes.

Textual contents of tweets are used to perform sentiment analysis in order to gain insight into public attitudes and concerns regarding crimes in the NYC subway system. We also accompany such analysis by Twitter users' demographics to better understand whose attitudes toward taking public transit are most affected by these crimes. Such understanding could identify the population who has the most severe concerns about transit safety, and how these concerns affect their willingness to take subway. Such analysis would further help inform the police and policymakers to tailor public communication strategies and to prioritize resources to the neighborhoods where the most affected populations reside.

A. Related Work

A majority of studies utilize crime data from police statistics to investigate how transportation system design [3]–[6] can impact crimes. Ridership data is also used to study relationships between neighborhood crimes and public transit usage [7], [8]. [9] conducts a survey to investigate the relationship between users' travel behavior and fear of crime. However, these studies primarily focus on the aggregate crime rate in urban systems. To investigate the impact of crime events, we employ the social media data to gather opinions from individual users. This approach is advantageous, as the majority of people obtain news and post their comments on the social media platforms [10]. Collecting social media data is also less costly compared to conducting surveys, allowing us to reach a larger and more diverse demographic of people. In recent years, there has been a growing trend of using social media data [11] to investigate people's attitudes toward incidents in transportation systems, such as car crashes [12], [13] and murder [14]. In this research, we leverage sentiment and demographic analysis to target groups with more negative opinions regarding crimes in the subway system.

B. Contributions of this paper

This work focuses on understanding public attitudes and concerns regarding crimes in the NYC subway system through the analysis of social media data. The ultimate

(Corresponding author: Xuan Di.)

‡This work is sponsored by NSF SCC-2218809.

Xu Chen is with the Department of Civil Engineering and Engineering Mechanics, Columbia University, New York City, NY 10027 USA (e-mail: xc2412@columbia.edu).

Yishi Wang is with the Department of Civil Engineering and Engineering Mechanics, Columbia University, New York City, NY 10027 USA (e-mail: yw3751@columbia.edu).

Xuan Di is with the Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, 10027 USA, and also with the Data Science Institute, Columbia University, New York, NY, 10027 USA (e-mail: sharon.di@columbia.edu).

Id	Time	Summary	Source	Number of tweets
1	Feb 05, 2021	A man was slashed across the face during a dispute on a subway train in Manhattan.	https://abc7ny.com/subway-slashing-l-train-man-slashed-on-1st-avenue-station/10312535/	501
2	Feb 13, 2021	A man stabbed four people who were sleeping in the subway, killing two of them.	https://www.nytimes.com/2021/02/14/nyregion/nyc-subway-stabbing-arrest.html	578
3	Apr 17, 2021	A man with AK-47 and gas mask arrested in Times Square subway station.	https://www.kiro7.com/news/trending/ohio-man-arrested-times-square-subway-station-with-ak-47-rifle-police-say/5A4JPC7LO5CNXM7S4E3ODF4X7Q/	528
4	May 31, 2021	A 64-year-old man was slashed, beaten, and robbed inside a subway station.	https://www.wftv.com/news/trending/thieves-cut-beat-rob-64-year-old-man-new-york-city-subway-station/GIQJKEGCMNHBPMP7FQSCKZQVUA/	508
5	Oct 25, 2021	A man was shot at a subway station.	https://www.nbcnewyork.com/news/local/man-shot-at-union-square-subway-station-police/3350441/	512
6	Nov 21, 2021	A soccer player was randomly stabbed to death when sleeping on the subway.	https://www.nbcnewyork.com/news/local/crime-and-courts/soccer-player-was-sleeping-on-subway-when-randomly-stabbed-to-death-ny-only/3413943/	500
7	Jan 15, 2022	A woman died after she was pushed onto subway tracks.	https://www.cnn.com/2022/01/15/us/nypd-make-arrest-connected-to-asian-woman-death-subway/index.html	3578
8	Apr 12, 2022	A mass shooting on the Brooklyn subway left 29 people injured.	https://www.nytimes.com/live/2022/04/12/nyregion/brooklyn-subway-shooting	20022
9	May 22, 2022	A man was fatally shot on NYC subway's Q train.	https://www.nytimes.com/2022/05/22/nyregion/nyc-subway-shooting.html	2920

TABLE I: Crimes in NYC subway system

goal is to understand the association between rising crimes and people's attitudes toward taking public transit, so as to assist transit planners and policymakers to create a safer environment to restore transit ridership, which suffered from a substantial drop during the pandemic. The study leverages sentiment analysis and demographic analysis to gain insights into people's sentiments and the factors that influence their attitudes towards nine subway crimes. The specific crimes of interest are crime 7 (a woman pushed onto the subway tracks) and crime 8 (a mass shooting on the subway). The contribution of this work lies in providing valuable insights into public attitudes regarding crimes in the NYC subway system and highlighting the factors that influence people's attitudes towards subway crimes.

The rest of the paper is organized as follows. Section II describes the data sample. Section III presents sentiment results of individual tweets corresponding to each crime event. Section IV conducts demographic analysis. Section V concludes.

II. Data Collection

In this section, we present the data sample collected from social media and the NYC open data source, focusing on crime events in the NYC subway system. Table I displays nine specific crimes that occurred in the NYC subway system between January 2021 and June 2022. Each crime is assigned a unique event ID, listed in chronological order. For simplicity and easy reference, we will use the event ID to refer to each crime in the rest of the paper. More detailed information regarding each crime can be found in the provided source links.

A. Social media data

Textual data: In order to obtain textual data related to each crime, we utilized the Twitter API to fetch tweets associated with the corresponding crime hashtags. The retrieved tweets contain valuable information, including the user ID, username, tweet ID, tweet text, and timestamp. The number of collected tweets for each event is summarized in Table I. Note that the size of data sample regarding crime 1-6 is relatively small, indicating that these events did not lead to a hot discussion on Twitter. It means public interest in different crime events may vary based on the severity and scope of crimes. More severe and high-impact incidents tend to attract greater attention from the public, as they are more concerning. In this study, we aim to analyze the sentiment results following each crime and delve into the details of crimes 7 and 8, which have gained broader attention.

User demographic data: In this study, we investigate the potential impact of user demographics on sentiment results for crime 7, which involved a woman being pushed onto the subway tracks, and crime 8, which involved a mass shooting incident. To explore this relationship, we leverage available demographic information provided by Twitter users, assuming there are no privacy restrictions on their profiles. For each crime event, we collect demographic information from 2000 users. Specifically, we utilize user IDs to manually collect demographic details such as gender, ethnicity, age, and occupation (as outlined in Section IV).

B. Subway ridership data

To better understand the impact of crimes in subway system, we analyze mobility usage data [15] from the NYC

open data source. We collect NYC subway turnstile data, which is available at `http://web.mta.info/developers/turnstile.html`.

III. Sentiment Analysis

In this section, we use sentiment analysis to investigate people's attitudes toward each crime in NYC subway system. The goals of sentiment analysis are to understand:

1) How do people's attitudes change post each crime?
2) To which crime events are people more sensitive?
3) How do subway ridership change pre, during, and post crime events?

A. Sentiment classifier

In this subsection, we provide a brief introduction to BERT (Bidirectional Encoder Representation from Transformers), a transformer-based natural language processing (NLP) model [16]. BERT has gained popularity for its ability to understand the context and meaning of words in a sentence by considering the surrounding words. We utilize BERT to develop a sentiment classifier that can categorize the sentiment of each tweet into either "positive" or "negative" categories. By leveraging BERT's powerful language understanding capabilities, we determine the sentiment expressed in tweets.

To train the sentiment classifier [17], we first follow a labeling process in which each tweet in a data sample is assigned sentiment categories. The labeling task is performed by graduate students. To measure the agreement between labellers, we calculate the proportion of tweets labeled with the same category for each pair of students, which measures the consistency of their labeling. If the concordance score falls below 0.9, indicating that more than 10% of the tweets are marked with different categories, the data sample is then reevaluated by another pair of labellers. This process ensures that the sentiment labels assigned to the tweets are reliable and consistent, and it helps mitigate potential biases.

We use the labeled data to train the classifier and apply the trained model to determine the sentiment category. The labeled data is split into training and validation sets. Each sentence in the training data is tokenized and passed to a pre-trained BERT encoding system with 12 encoder blocks. Each BERT encoder block applies attention methods to the input and a feed-forward network to pass the result to the next block. After encoding, we only use the output from the last BERT encoder block as the input for the sentiment classifier. Here, we choose to build a neural network with one input layer, three hidden layers, and one output layer with the softmax function as the activation function. We adopt weighted cross-entropy as our loss function due to the imbalanced categories in the data sample. After fine-tuning the parameters in the neural network, the sentiment classifier achieves an 82.7% accuracy rate with our validation data. We then apply the trained classifier to determine the sentiment categories for the remaining data.

B. Results

Fig. 1 illustrates the sentiment results following each crime event. The x-axis represents the event ID in Table. I. The left and right y-axes indicate the proportions of negative and positive tweets, respectively. The findings reveal that the proportion of negative tweets associated with each crime exceeds 80%. This observation suggests that crimes have a significant impact on public discussions and debates regarding subway safety.

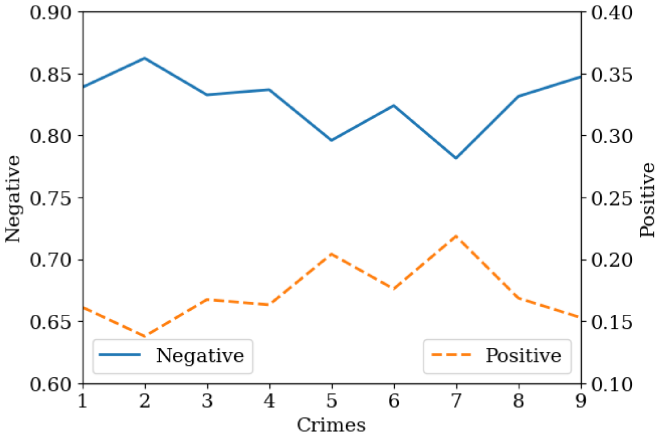
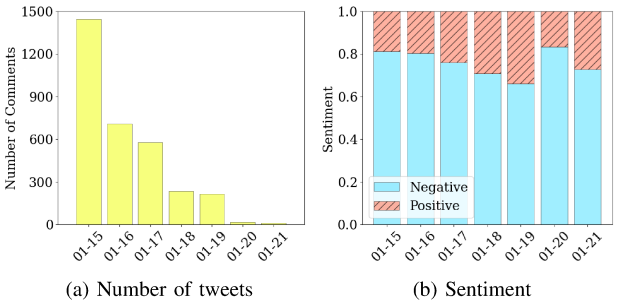


Fig. 1: Sentiment results of 9 crimes

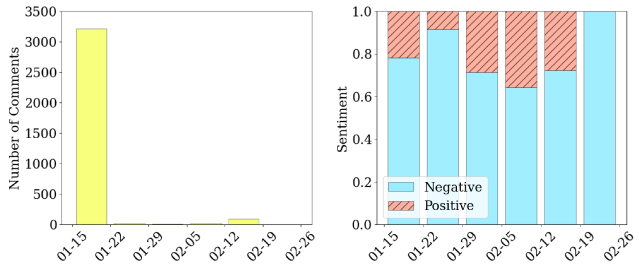
Fig. 2(a) and Fig. 2(b) present the number of tweets and sentiment results over a seven-day period following the crimes. The figures reveal that the number of tweets related to crime 7 decreased after January 15, indicating a diminishing discussion among users. However, the proportion of negative tweets remained consistently high. This suggests that users did not alter their negative attitude toward the crime despite the decrease in overall discussion volume. Fig. 3(a) and Fig. 3(b) present the number of tweets and sentiment results over a six-week period following the crimes. It is evident that after the first week when the crime occurred, there was a significant decrease in user discussions regarding the event on social media. This trend can also be observed in the sentiment results of crime 8, as depicted in Fig. 4 and Fig. 5.



(a) Number of tweets (b) Sentiment

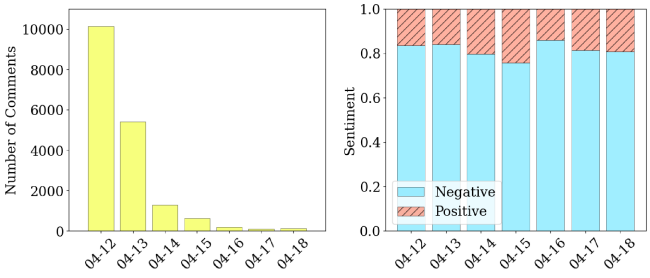
Fig. 2: Crime 7 - daily performance

Fig. 6 and Fig. 7 visualize the subway ridership in stations where crimes 7 and 8 occurred. The turnstile data records the number of exits and entries at each subway station every 4 hours. We first look into crime 7 on January 15,



(a) Number of tweets (b) Sentiment

Fig. 3: Crime 7 - weekly performance



(a) Number of tweets (b) Sentiment

Fig. 4: Crime 8 - daily performance

2022. The x-axis is the timeline and the y-axis denotes the entries. The yellow line illustrates the trend of entries on the day when the crime took place. The blue and green lines represent the trends of entries one week prior to and after the crime, respectively. The subway ridership first experienced a significant decrease and then bounced back on January 15. We now look into crime 8 where a mass shooting occured on the subway starting from the station at 59th Street to the station at 36th Street. In Fig. 7(b) and Fig. 7(a), we present the plotted trends of ridership at the stations where the shooting originated and ended, respectively. It was observed that on April 12, the ridership at the 36th Street station decreased to 0 after 13:00 due to a police lockdown. In contrast, the ridership at the 59th Street station rebounded to 600 after 16:00, which was 60% lower than the ridership recorded one week prior to and after the crime. This indicates that the mass shooting caused fear among the public, leading to a decrease in ridership at the affected stations.

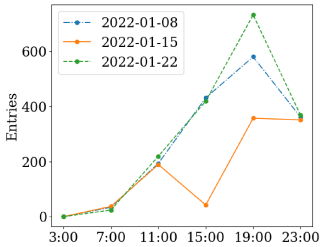
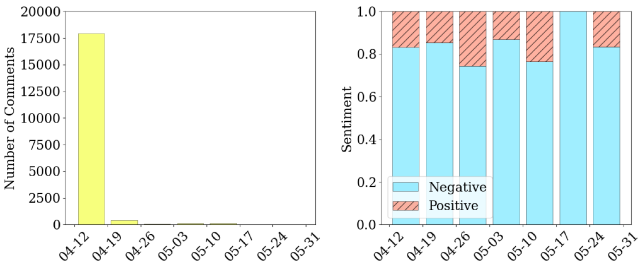
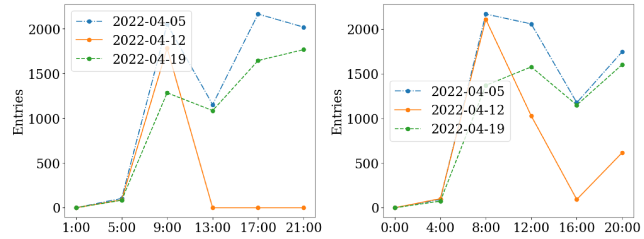


Fig. 6: Pushed to track - Crime 7



(a) Number of tweets (b) Sentiment

Fig. 5: Crime 8 - weekly performance



(a) 36th ST (b) 59th ST

Fig. 7: Mass shooting - Crime 8

We delve into individual tweets post these two crimes. In the aftermath of crime 7, where a woman was pushed onto the subway tracks, there was a heated discussion on anti-Asian hate crimes (See Table II). Many tweets exemplify this sentiment. These tweets highlight the individual's personal experience and acknowledge the existence of racism, specifically in relation to anti-Asian sentiments. Similarly, following the mass shooting on the subway (crime 8), there was a debate surrounding gun violence. Some tweets expressed the safety concern for the entire transportation system. These tweets reflect a broader apprehension about the safety within various modes of public transportation.

	Tweets	Event
Negative	100% of violent racism. I've experienced enough of it to agree that Black on Asian racism has been a thing since I've been alive.	7
	I'm just curious what would happen if we actually allowed the CDC to study the data surrounding gun violence, instead of actively forbidding them from doing so.	8
	All public transportation is a cesspool and an absolute hazard to Asians. I've read so many incidents in subway trains, at subway stations, in public buses, at bus stops, airplanes, and sidewalks.	8
Positive	I bet he would not escape NYC.	7
	It would be nice if the richest city in the richest nation on earth could afford barriers on their subway stations.	8
	I have lived in New York for 6 years now. I have never felt unsafe on the subways.	8

TABLE II: Sample tweets

Note that there exists a small proportion of positive tweets pertaining to each crime. Some tweets reflected a positive outlook on the authorities' ability. Additionally, there were

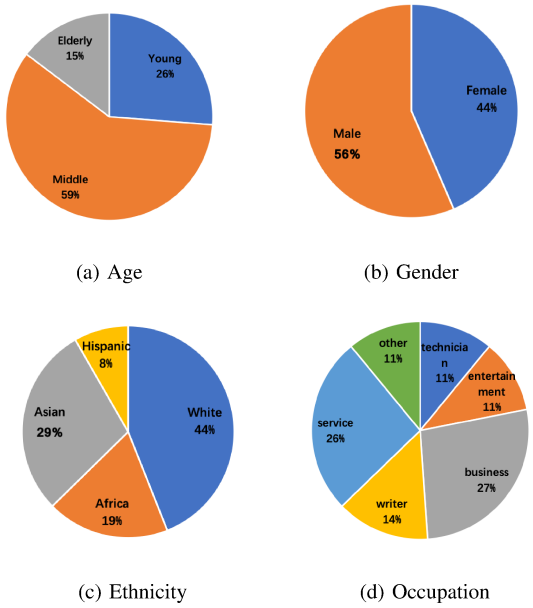
tweets that conveyed a positive attitude toward the public transit system. These tweets showed a positive sentiment towards improving the safety measures of the subway system, emphasizing the desire for better infrastructure. Some users maintain a positive attitude toward subway safety. They have consistently felt secure when taking public transits.

IV. Demographic Analysis

In this section, we study the relationship between user demographics and their sentiments toward crimes in NYC subway system. In particular, we implement regression analysis on the sentiment results regarding two highly discussed crimes: pushed to track (i.e., crime 7) and mass shooting (i.e., crime 8). By performing the regression analysis, we aim to gain insights into how user demographics may influence the sentiments regarding specific crimes.

1. Pushed to track - Crime 7

Fig. 8 presents a summary of demographic information obtained from 2076 users who posted tweets or comments regarding crime 7. Among these users, 56% are male, while 44% are female. The users are further categorized into three age groups: the young (age $\leqslant$ 30), middle-age (age 30-65), and the elderly (age $\geqslant$ 65). The proportions of users in these groups are 26%, 59%, and 15%, respectively. The distribution of ethnicity among the users is depicted in Fig. 8(c). Furthermore, in Fig. 8(d), we provide a generalization of several occupational categories among the users.



(a) Age (b) Gender

(c) Ethnicity (d) Occupation

Fig. 8: User demographic - Crime 7

Regression analysis is conducted using three different approaches, which are briefly introduced as follows:

1) Logistic Regression: Logistic Regression is suitable when the dependent variable is binary, such as positive or negative sentiment. By applying this method, we can examine the relationship between user demographics and the likelihood of expressing a particular sentiment category. We use the softmax function to normalize all features and choose cross entropy as the loss function.
2) Random Forest: Random Forest is a machine learning algorithm that constructs multiple decision trees and combines their predictions to make a final prediction. In this case, Random Forest can be applied to determine how user demographics contribute to predicting sentiment categories.
3) XGBoost: XGBoost is an ensemble learning algorithm that is commonly used for regression tasks. It builds a strong predictive model by iteratively adding decision trees to minimize the prediction error. We apply XGBoost to analyze the relationship between user demographics and sentiment categories.

We divide the data into training and test sets using an 8:2 ratio. The data is split in a stratified manner to ensure that both the training and test sets have balanced classes, including positive and negative sentiments. To mitigate overfitting, we employ the 10-fold cross-validation method during the training process. The package "scikit-learn" in Python3 is used to implement logistic regression and random forest. The package "XGBoost" in Python3 is used to implement XGBoost. The accuracy results of XGBoost, logistic regression and random forest are shown in Table. III. XGBoost outperforms the other two.

Model	Accuracy
XGBoost	75%
Logistic Regression	65%
Random Forest	73%

TABLE III: Accuracy

Fig. 9 illustrates the feature importance from XGBoost results. It indicates that among the different demographic groups, the middle-age group exhibits the highest sensitivity to crime 7. Additionally, gender and ethnicity are identified as important factors influencing people's sentiments toward crimes. Note that among all the ethnicity categories, the Asian group demonstrates the highest sensitivity to the crime. This sensitivity can be attributed to the fact that the woman who tragically lost her life after being pushed onto the subway tracks was of Asian descent. Consequently, the fear and concern among the Asian community led to more discussions and negative sentiments regarding anti-Asian hate crimes on Twitter.

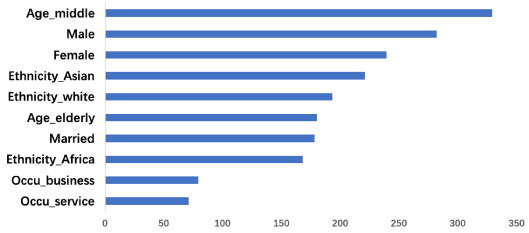
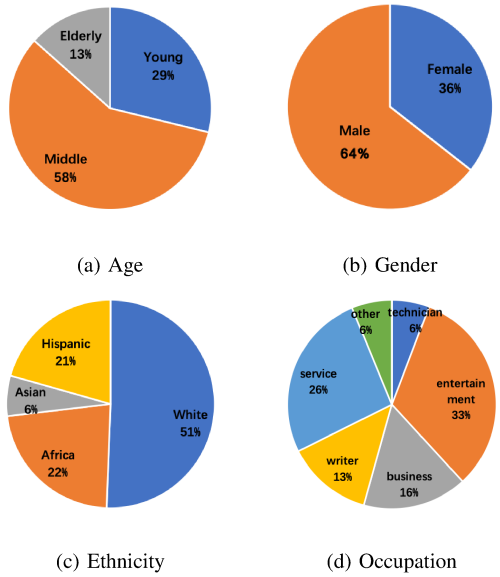


Fig. 9: Feature Importance ranking - Crime 7

2. Mass shooting - Crime 8

We now look into the demographic analysis of the mass shooting crime. Fig. 10 presents a summary of demographic information obtained from 1967 users who posted tweets or

comments regarding crime 8. The accuracy of XGBoost, logistic regression and random forest are 78%, 60% and 72%, respectively. Fig. 11 illustrates the feature importance results. It reveals that among the various demographic groups, females exhibit the highest sensitivity to the mass shooting event. This finding aligns with the reality that many women have concerns about their safety while using public transportation in NYC [1]. Age and ethnicity are also identified as important factors that influence people's sentiments toward crimes. The demographic analysis of both crimes demonstrates that age, gender, and ethnicity have a significant impact on people's attitudes toward crimes.



(a) Age (b) Gender

(c) Ethnicity (d) Occupation

Fig. 10: User demographic - Crime 8

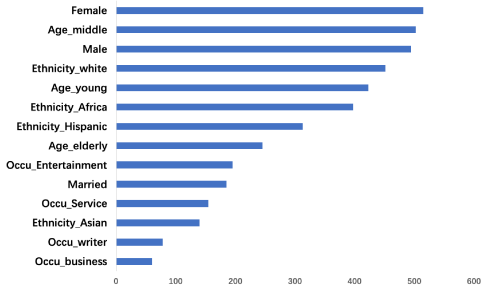


Fig. 11: Feature Importance ranking - Crime 8

V. CONCLUSION

In this work, we analyze social media data to understand the public opinion on rising subway crimes, particularly in NYC. We focus on a series of 9 crimes, which happened between Feb 2021 and May 2022 in NYC subway systems, and extract 29647 tweets tweeted around these crimes. BERT model is performed to analyze the sentiment of Tweets toward the subway crimes. In particular, we study Asian hate crimes and mass shooting, which have been rising all over the country, especially post the pandemic [1]. Demographics of 4043 twitter users are extracted, including age, gender, ethnicity, and occupation. We find that (1) Negative attitudes towards crimes indicate fear among the public and lead to a decrease in subway ridership. (2) Age, gender and ethnicity are important factors influencing people's sentiments toward crimes. (3) The fear and concern among the Asian community lead to negative sentiments regarding anti-Asian hate crimes. (4) Compared to others, women have more concerns about their safety while using public transportation in NYC.

In the future, we will investigate the relationship between crimes and other mode choices (e.g., bus, bike, taxi) and look into populations who are most affected by these crimes and may give up transit and switch to other travel modes in a multimodal transportation system.

REFERENCES

[1] A. Ley, "As subway ridership rebounds, some women are reluctant to return," https://www.nytimes.com/2023/02/13/nyregion/nyc-subway-women-crime.html, 2023.

[2] "We will not allow new yorkers to be terrorized': What we know about brooklyn subway shooting," https://www.usatoday.com/story/news/nation/2022/04/12/brooklyn-new-york-subway-shooting-what-we-know/7290322001/, 04 2022.

[3] Y. Irvin-Erickson and N. La Vigne, "A spatio-temporal analysis of crime at washington, dc metro rail: Stations'crime-generating and crime-attracting characteristics as transportation nodes and places," *Crime Science*, vol. 4, p. 14, 2015.

[4] O. Zervina, "The environment of extremist textual content threatening transportation systems," *Transportation Science and Technology*, pp. 541–551, 2020.

[5] P. Anciaes and P. Jones, "Transport policy for liveability- valuing the impacts on movement, place, and society," *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 157–173, 2020.

[6] N. Su, W. Li, and W. Qiu, "Measuring the associations between eye-level urban design quality and on-street crime density around new york subway entrances," *Habitat International*, vol. 131, p. 102728, 2023.

[7] E. J. Traut and A. Steinfeld, "Identifying commonly used and potentially unsafe transit transfers with crowdsourcing," *Transportation Research Part A: Policy and Practice*, vol. 122, pp. 99–111, 2019.

[8] N. Li and Y.-A. Kim, "Subway station and neighborhood crime: An egohood analysis using subway ridership and crime data in new york city," *Crime & Delinquency*, 2022.

[9] J. Soto, M. Orozco-Fontalvo, and S. A. Useche, "Public transportation and fear of crime at brt systems: Approaching to the case of barranquilla (colombia) through integrated choice and latent variable models," *Transportation Research Part A: Policy and Practice*, vol. 155, pp. 142–160, 2022.

[10] E. Shearer and A. Mitchell, *News Use Across Social Media Platforms in 2020*. [Online]. Available: https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/

[11] A. Wong, S. Ho, O. Olusanya, M. V. Antonini, and D. Lyness, "The use of social media and online communications in times of pandemic covid-19," *Journal of the Intensive Care Society*, p. 1751143720966280, 2020.

[12] W. Yao and S. Qian, "From twitter to traffic predictor: Next-day morning traffic prediction using social media data," *Transportation Research Part C: Emerging Technologies*, vol. 124, p. 102938, 2021.

[13] X. Chen, H. Zeng, H. Xu, and X. Di, "Sentiment analysis of autonomous vehicles after extreme events using social media data," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1211–1216.

[14] Q. Ye, X. Chen, H. Zhang, K. Ozbay, and F. Zuo, "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.

[15] X. Chen and X. Di, "How the covid-19 pandemic influences human mobility? similarity analysis leveraging social media data," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 2955–2960.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, 2019.

[17] X. Chen, Z. Wang, and X. Di, "Sentiment analysis on multimodal transportation during the covid-19 using social media data," *Information*, vol. 14, no. 2, 2023.