

Article

Investigating the Impacting Factors on the Public's Attitudes towards Autonomous Vehicles Using Sentiment Analysis from Social Media Data

Shengzhao Wang ¹, Meitang Li ², Bo Yu ^{1,*} , Shan Bao ^{2,3} and Yuren Chen ¹

¹ Key Laboratory of Road and Traffic Engineering of the Ministry of Education, College of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China

² Human Factors Group, University of Michigan Transportation Research Institute, 2901 Baxter Rd., Ann Arbor, MI 48109, USA

³ Industrial and Manufacturing Systems Engineering Department, University of Michigan—Dearborn, 4901 Evergreen Rd., Dearborn, MI 48128, USA

* Correspondence: boyu@tongji.edu.cn

Abstract: The attitudes of the public play a critical role in the acceptance, purchase, utilization, and research and development of autonomous vehicles (AVs). Currently, the attitudes of the public toward AVs have been mostly estimated through traditional survey data, which bears a low quantity of samples with high labor costs. It is probably also one of the reasons why the critical factors on the attitudes of the public toward AVs have not been studied from a comprehensive perspective yet. To address the issue, this study aims to propose a method by using large-scale social media data to investigate key factors that affect the attitudes of the public toward AVs. A total of 954,151 Twitter data related to AVs and 53 candidate independent variables from seven categories were extracted using the web scraping method. Then, sentiment analysis was used to measure the public attitudes towards AVs by calculating sentiment scores. Random forests algorithm was employed to preliminarily select candidate independent variables according to their importance and a linear mixed model was utilized to explore the impacting factors, considering the unobserved heterogeneities caused by the subjectivity level of tweets. The results showed that the attitudes of the public toward AVs were slightly optimistic. Factors, such as “drunk”, “blind spot”, and “mobility”, had the largest impacts on public attitudes. In addition, people were more likely to express positive feelings when talking about words, such as “lidar” and “Tesla”, related to high technologies. Conversely, factors, such as “COVID-19”, “pedestrian”, “sleepy”, and “highway”, were found to have significantly negative effects on the attitudes of the public. The findings of this study are beneficial for the development of AV technologies, the guidelines for AV-related policy formulation, and the understanding and acceptance of the public toward AVs.

Keywords: autonomous vehicles; social media data; public attitudes; sentiment analysis; linear mixed model



Citation: Wang, S.; Li, M.; Yu, B.; Bao, S.; Chen, Y. Investigating the Impacting Factors on the Public's Attitudes towards Autonomous Vehicles Using Sentiment Analysis from Social Media Data. *Sustainability* **2022**, *14*, 12186. <https://doi.org/10.3390/su141912186>

Academic Editor: Peng Jing

Received: 30 August 2022

Accepted: 22 September 2022

Published: 26 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advent of autonomous vehicles (AVs) represents a new technological revolution in the transportation sphere. AVs have a great potential in reducing gas emissions, fuel consumption, and traffic accidents associated with human-related errors and increasing mobility, especially for the elderly and the disabled [1–3]. However, the popularity of AVs is still at a low level due to legal, liability, privacy, and safety issues [4]. It was shown in a survey study that 63% of the respondents were not likely to use AVs [5]. The public's attitudes play a critical role in the acceptance, purchase, use, and research and development of AVs [6–8].

Although many studies have tried to explore factors affecting peoples' acceptance of AVs, most of them used traditional survey methods, such as online questionnaires [9,10] and

field investigations [3,11]. For example, a national online survey with open-ended questions was administered to a sample of 1624 Australians and it showed that the negative attitudes towards AVs may be derived from a combination of cognitive and emotional factors [9]. However, the sample size of the traditional survey method is relatively small, in a range of 300–5000 copies, which may be not able to effectively represent the public's attitudes [12]. Additionally, labor cost consumption and limitations of time and space are also additional weaknesses of the questionnaire survey regarding transportation services [13].

These survey studies on the impacting factors of the attitudes towards AVs found that: (1) Males, younger people, urban residents, and those with higher income and higher levels of education were likely to hold positive attitudes towards AVs [14–16]. (2) Better fuel and energy efficiency and lower vehicle emissions were expected to be brought from AVs by those who pursued environmental protection [17,18]. (3) The unwillingness to pay any additional fees for AVs was expressed by a large proportion of respondents [19]. (4) Safety of AVs, security issues, traffic congestion, and automatic parking were also reported to be impacting factors in public acceptance towards AVs [1,20].

Unlike traditional survey methods, social media provides a new source of timely data at a large scale and low cost. As of the first quarter of 2021, Twitter (a representative platform of social media) had 199 million monetizable daily active users worldwide. It is entirely possible to obtain wide-range information from the public promptly through social media [21]. People have become both producers and disseminators of news and reviews with the flourishing of social media [22]. It can be found that a large number of comments related to AVs have been released on social media and the public's close interest is essential for the promotion and application of AV technologies. In addition, social media data are widely used in areas other than social studies. For instance, Facebook groups were used to explore the transnational online communities generated by the Zimbabwean diaspora [23] and Twitter can be used as a source of journalism [24]. Meanwhile, TikTok had a significant influence on the tourists' choices of destination [25]. There have been some studies using social media data to perform further analysis in the field of transportation. For example, the study of Ding et al. [26] provided a new framework to analyze public opinions on AVs using Twitter feeds, but the detailed impacting factors were not further explored. The 1-year over 3 million Twitter contents in Northern Virginia and New York City were extracted to detect traffic accidents [21]. The results showed that nearly 66% of the accident-related tweets can be located by the actual accident on highways, indicating the possibility to regard tweets as an accident detection tool due to their effectiveness regarding time and location. In addition, due to the large-scale and real-time nature of social media data, these kinds of new data have also been introduced in studies of traffic flow prediction [27,28], traffic strategy designation [29], route planning [30], etc.

To analyze the big data from social media, natural language processing (NLP) is a potentially powerful method. NLP is an important direction in the field of computer science, which provides an opportunity to understand people's opinions regarding hot topics by enabling computers to obtain the meaning of human language from text documents [31,32]. In the field of transportation, various objectives, such as traffic investigation, safety, and management, can be achieved in the NLP methods [33]. Sentiment analysis is one of the most commonly used methods in NLP for analyzing attitudes and feelings [34]. In a study by Liu et al. [13], a sentiment analysis method was proposed to calculate people's satisfaction with different transit facilities based on the data extracted from the website. In another study, sentiment analysis was performed for analyzing comments on the 15 most-viewed AV-related videos [7].

Many different statistical analyses and machine learning methods have been used to investigate the impacting factors on public attitudes towards AVs, such as *t*-tests and ANOVA [9], support vector machines algorithm [35], the univariate model [10], and so on. Compared to the above methods, the mixed model can consider unobserved heterogeneities to obtain a more accurate relationship between variables [36]. It was employed in many studies to explore significant factors. For example, mixed model analyses were used

to assess and test the impacting factors on drivers' compliance with the speed choice suggestions [37]. Driver behavior in response to the warning was modeled by mixed-effects Poisson regression models [38].

Additionally, since social media data can provide abundant variables, variable selection is needed at the beginning of analyses. The random forests (RF) algorithm is widely employed to screen variables in the early stage because it can provide the variables' importance and handle a large number of variables with higher dimensionality at the same time [39]. For instance, the most important explanatory variables associated with severe crash occurrence were selected by RF [40]. RF was also used for the selection of significant variables related to the injury severity of secondary incidents [41].

Given the above, most current studies explored the public's attitudes towards AVs through questionnaires. However, the quantity of data collected in the traditional survey methods was very limited and the influencing factors on the public's attitudes and acceptance of AVs have not been studied from a comprehensive perspective yet. This study aims to provide a new method to investigate the public's attitudes towards AVs and their impacting factors by using larger and more timely social media data. In this study, a total of 954,151 tweets and 53 candidate independent variables were extracted using the web scraping method, and the sentiment analysis method was used to measure public attitudes on AVs. Then, the random forests algorithm and a linear mixed model were employed to analyze the impacting factors. The results of this study are beneficial for the development of AV technologies, the guidelines for AV-related policy formulation, and the public's understanding and acceptance of AVs.

2. Methodology

The overall analysis framework is outlined in Figure 1 and the details of the methodology are explained below.

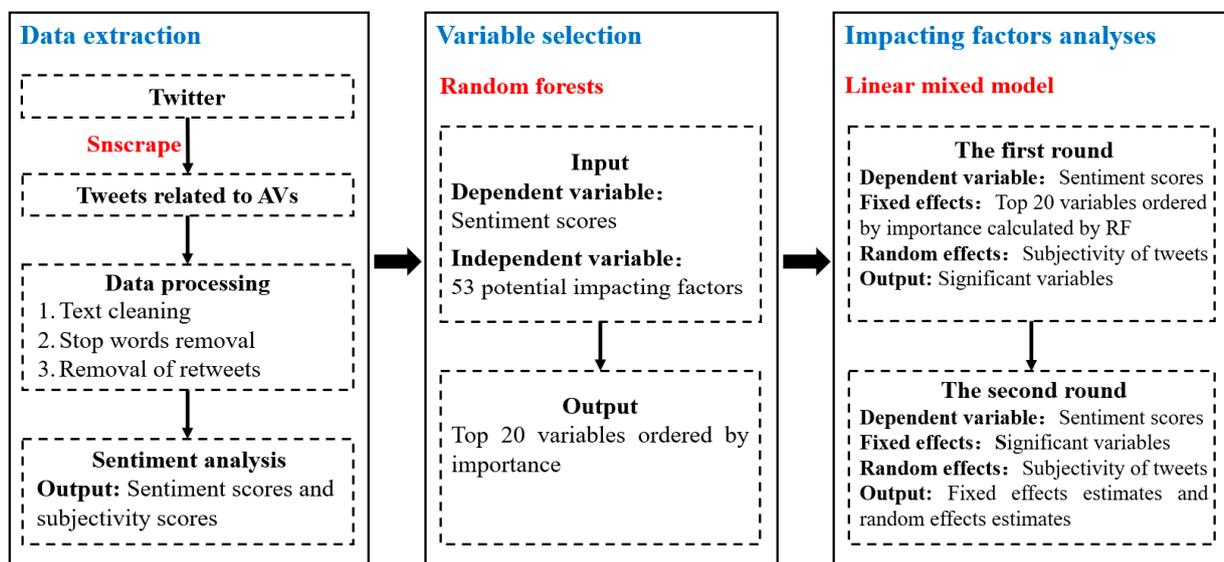


Figure 1. The overall analysis framework.

2.1. Data Extraction

In this study, a sentiment analysis method was applied to investigate the public's attitudes towards AVs and their impacting factors based on large-scale Twitter data. Sentiment analysis is a ramification in the natural language processing (NLP) domain, which focuses on detecting, processing, and analyzing the feelings and emotions expressed by human beings in their language. Text sentiment analysis is one of the most important parts of sentiment analysis, which can deal with sentimental subjective texts using NLP and text mining techniques. Sentiment analysis tasks can be divided into the chapter level, sentence

level, and word or phrase level according to the granularity of analysis. This study was conducted from the word or phrase level.

The first step of this study was to obtain the Twitter data relevant to people's opinions on AVs. As an efficiency scraper for social networking services, "snsraper" (<https://github.com/JustAnotherArchivist/snsraper> (accessed on 12 December 2020)) was used to extract valid data based on several AV-related keywords: "Autonomous vehicle", "Autonomous driving", "Autonomous car", "Driverless", and "Self driving". Each keyword was associated with a large number of search results and the resulting corpus contained a total of 954,151 tweets from 1 January 2019 to 30 November 2020 for further pre-processing. This period witnessed the global outbreak of COVID-19 and the further development of AV technologies, which contributed to a large number of relevant tweets during this period. Hence, this period was chosen for investigation in this paper.

The following detailed steps were carried out on the corpus:

- (1) Text cleaning: The text part of the tweets contained irrelevant information such as numbers, URLs, special symbols, mentions, emojis, HTML tags, etc., which were filtered out. For hashtags, only the "#" symbol was deleted and the text content was retained for further analysis. In addition, empty and NaN entries were removed from the dataset.
- (2) Stop words removal: Stop words included irrelevant filler words such as for, at, on, which, is, etc., which were removed from the corpus since these words could not accurately reflect the user's comment attitudes.
- (3) Removal of retweets: Retweets were the tweets' replies. These replies were supposed to be filtered out because most of them were consistent in attitude with the existing tweet.

The next step was to perform the sentiment analysis to identify emotional attitude (i.e., positive or negative) and the subjective degree (i.e., subjective or objective) in each given tweet text. Two text processing tools in Python, TextBlob (<https://textblob.readthedocs.io/en/dev/> (accessed on 19 January 2021)), and IBM Watson (<https://ibm.com/watson> (accessed on 20 January 2021)) were adopted to compute the sentiment and subjectivity for the preprocessed text, respectively. Sentiment scores ranged from -1 (most negative) to 1 (most positive), among which 0 indicates a neutral attitude. Subjectivity scores ranged from 0 (most objective) to 1 (most subjective) and were further classified into four categories, including "very objective", "objective", "subjective", and "very subjective", as illustrated in Table 1.

Table 1. Classification and distribution of the subjective score.

Category	Very Objective	Objective	Subjective	Very Subjective
subjective score	0~0.25	0.25~0.5	0.5~0.75	0.75~1
proportion	27.9%	31.7%	33.9%	7.5%

Then, 53 potential impacting factors that may affect the public's attitudes towards AVs were considered, deriving from seven categories: "event", "people", "vehicles", "roads", "environment", "autonomous driving-related companies", and "autonomous driving-related characteristics". Each category included a number of subdivided potential influencing factors and then each factor needed to have detailed words for further interpretation, which are listed in Table 2. These seven categories were selected considering the role of AVs in the seven elements of transportation.

Table 2. AV-related topics, variables, and detailed words from Twitter data.

Topics	Variables	Detailed Words
events	policy testing COVID-19	policy; policy publication; testing; test scenario coronavirus; COVID-19; COVID19; epidemic; lockdown; quarantine
people	pedestrian stress passenger car following brake accelerate drunk driving fatigue driving sleepy male female young old income	pedestrian; passerby stress; easy; relaxed; convenience; nervous; tension; anxious passenger; chauffeur car following; follow the car brake; slow down accelerate; speed up; acceleration drunk; zonked; stoned; drink-driving; intoxicated fatigue driving; tired sleep; sleepy; drowsy male driver female driver; woman driver; chauffeuse young man; stripling; teenager old; elderly income; earning; salary; afford; price; expensive
vehicles	truck bike speed lidar blind spot	truck; wagon; van; lorry bike; bicycle speed; velocity lidar; radar blind spot; vision blind area
roads	highway roadway urban road roundabout toll station ramp intersection	highway; expressway roadway urban road roundabout toll station; toll gate; services station ramp intersection
environment	weather morning noon afternoon evening traffic signal	weather; sunny; rainy morning noon afternoon evening traffic light; signal
AV-related companies	Baidu Uber Volvo Zoox Voyage Waymo Argo AI Tesla	Baidu Uber Volvo Zoox Voyage Waymo Argo AI Tesla
AV-related characteristics	mobility parking energy congestion safety legal issues privacy public transportation cyber issues ethical issues	mobility; convenient; parking energy conservation; emission reductions; environmentally friendly; fuel economy; emission fuel efficiency; energy efficiency congestion safe; risk; security; crash; accident; collision legal issues; legal liability; liability issues privacy; personal data bus; public trans; metro; subway cyber security; network security; internet security; hack ethical; moral

Note: Detailed word searches are not case sensitive.

For Topic 1 (events), the public's opinions relevant to the measures imposed by the governments and COVID-19 were collected. With reference to Topic 2 (people), due care was taken to mainly discuss the human attributes and driver behavior. Detailed words such as "brake", "accelerate", and "drunk driving" constituted the major part of the driver's status and behavior. As for Topic 3 (vehicles), detailed words such as "truck", "speed", and "lidar" showed attention which was drawn to obtaining comments relative to the vehicle types, driving characteristics, and vehicle equipment. As for Topic 4 (roads), different road types and road nodes were considered to explore the public views on the applicability of AVs in different road conditions. Detailed words such as "weather", "morning", "afternoon", "traffic signal", etc., were added in Topic 5 (environment) to discuss the potential relationships between environment-related factors and AVs. Topic 6 mainly covered AV-related companies including "Tesla", "Baidu", "Waymo", etc., which may largely reflect the public's attitude towards AVs. Topic 7 (AV-related characteristics) focused on capturing viewpoints relative to the possible problems caused by AVs, such as safety concerns, legal issues, privacy issues, mobility, congestion, etc.

The detailed words were used to mine the massive tweets as the basic data for the subsequent mathematical–statistical analysis model. If a piece of comment data contained the corresponding word, it was marked as 1, while if it did not contain the corresponding word, it was marked as 0. Thus, every independent variable in the model was a binary variable.

2.2. Random Forests

In this study, the random forests (RF) algorithm was used to choose variables by calculating their importance. RF is an ensemble learning method for classification or regression by constructing a multitude of decision trees [42]. One of the advantages of RF is the power of handling large data sets with hundreds of input variables and mitigating the multicollinearity problem [37]. RF can automatically balance data sets when a class is more infrequent than other classes in the data. The method also handles variables fast, making it suitable for complicated tasks. Among all the available regression methods, random forests provides high accuracy. In addition, RF can output the importance of variables [43]. Thus, it is regarded as one of the handiest dimensionality reduction methods for variable selection.

Since the dependent variable in this study (i.e., sentiment scores) was a continuous variable, the decision trees constructed in RF were regression trees. Three methods are combined in the RF algorithm, including bootstrapping, boosting, and bagging [44]. Bootstrapping refers to random sampling with replacement. Bagging is a method that can calculate multiple models at the same time, which can realize parallel calculation and improve the robustness of the model. Boosting can help to reduce bias and avoid overfitting.

The basic process of Random forests consists of the following three steps [45]. Firstly, n_{tree} sets of bootstrap databases are formed according to the original database, where n_{tree} is the number of decision trees in the forest. Each bootstrap database set randomly selects around two-thirds of samples from the original database with replacement and bootstrap databases have the same dimension as the original one. Those leaf-out samples in each bootstrap database are called "out-of-bag" (OOB) data. Secondly, each bootstrap database set is used to grow an unpruned regression tree and each split within each tree is created by trying m_{try} candidate variables, where m_{try} denotes the number of different independent variables tested at each node. Lastly, the accuracy of each tree will be calculated by OOB data for each bootstrap database set and the whole accuracy of RF is the average accuracy of all the trees.

During the modeling process, RF can calculate the importance of variables by the mean decrease in Gini (i.e., Gini importance), which is a crucial feature for further selecting variables [41]. Gini importance measures how much a variable can decrease the impurity of the final model, which is calculated by the weighted average value of the impurity

reduction (i.e., variance reduction for regression) at all the splits with this variable across all the trees, as shown in the following Equation [46]:

$$IMP_{Gini}(x_l) = \frac{1}{n_{tree}} \sum_{n=1}^{n_{tree}} \sum_{t \in n: v(s_t)=x_l} p_t \Delta i(s_t, t) \quad (1)$$

where $IMP_{Gini}(x_l)$ denotes the Gini importance of the independent variable x_l ; p_t is the weight, namely, the proportion of samples reaching the t th node; $\Delta i(s_t, t)$ is the impurity decrease at the split s_t of the t th node; n_{tree} is the number of trees in the forest; $n : v(s_t) = x_l$ denotes that the variable x_l is used at the split s_t in the n th tree.

In this study, the “RandomForestRegressor” function in the “scikit-learn” (<https://scikit-learn.org/stable/> (accessed on 6 March 2021)) package of Python (version 3.7) was used to build the RF model. For training the model, 70% of the samples was randomly selected and used, while the remaining 30% was used to assess the quality of the model. All the parameters were tuned based on accuracy. When the accuracy reached stable and maximum values, the parameters were determined: the number of trees was 50, the number of variables tested at each split was 7, and nodes were expanded until being pure or containing less than 2 samples.

2.3. Linear Mixed Model

This study employed the linear mixed model to investigate the impacting factors on the public’s attitude towards AVs, considering the unobserved heterogeneities caused by the subjectivity in tweets. Considering the fact that simple linear regression ignores possible correlations between observations in the data, a linear mixed model was used to mitigate this problem (i.e., unobserved heterogeneities). The linear mixed model is a statistical technique that accounts for within- and between-subject variance for repeated measures data by providing both unbiased estimates of fixed effects and unbiased predictions of random effects [38]. This approach is a flexible and widely used tool to calculate maximum likelihood estimations for hierarchical, longitudinal, or correlated data analyses [36,37,47].

A linear model was employed to model the continuous outcome (e.g., sentiment scores) using the “lme4” (<https://cran.r-project.org/> (accessed on 16 April 2021)) package in the statistical software R (version 4.1.0), where odds ratios of fixed effects were analyzed to show relative likelihoods. The process of the general linear mixed model is calculated as below:

$$y = X_i \alpha + Z_i b_i + e_i \quad (2)$$

where y donates the sentiment scores of tweets related to AVs, which is the dependent variable; X_i denotes the fixed effects; α denotes the coefficient of the fixed effects; Z_i denotes the subjectivity in tweets related to AVs, which is selected as the random effect; b_i denotes the coefficient of the random effect; e_i denotes the observation level error terms.

Firstly, the top 20 variables in the RF regression algorithm with variable importance were used to be the fixed effects of the linear mixed model. Then, according to the results of the first model run, each variable was tested for statistical significance at the $p < 0.05$ level. All the less significant variables ($p > 0.05$) were filtered out and the remaining variables were selected as the input of the next model. Ultimately, the final model was checked on the goodness of fit as suggested by the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). AIC represents the relative amount of information loss caused by a given model and BIC estimates a posterior probability function of a model being true [48]. The AIC and BIC can be calculated as below:

$$AIC = 2k - 2 \ln \hat{L} \quad (3)$$

$$BIC = k \ln(n) - 2 \ln \hat{L} \quad (4)$$

where k donates the number of estimated parameters in the model; \hat{L} donates the maximum value of the likelihood function for the model; n donates the number of observations.

3. Results

3.1. The Distribution of Attitudes towards AVs

The average of all tweets' sentiment scores in each month is plotted in Figure 2, which shows the distribution of the public's general attitudes and acceptance towards AVs during the selected period (from January 2019 to November 2020). The maximum value of the average sentiment scores was 0.138 (January 2020) and the minimum value was 0.096 (March 2020). As time changes, the overall trend of the public's attitudes towards AVs was relatively stable. The most dramatic fluctuation occurred in March 2020, which was a 24.6% decrease compared to February 2020, perhaps affected by the outbreak of COVID-19. Overall, the distribution of average sentiment scores (>0) reflected that the public was slightly optimistic about the prospects of AV technologies. In addition, the number of tweets related to AVs each month was more than 30,000, which showed that AV technologies maintain a place as a hot topic for public discussion. Such large sample data also contributed to improving the effectiveness of our subsequent analysis results.

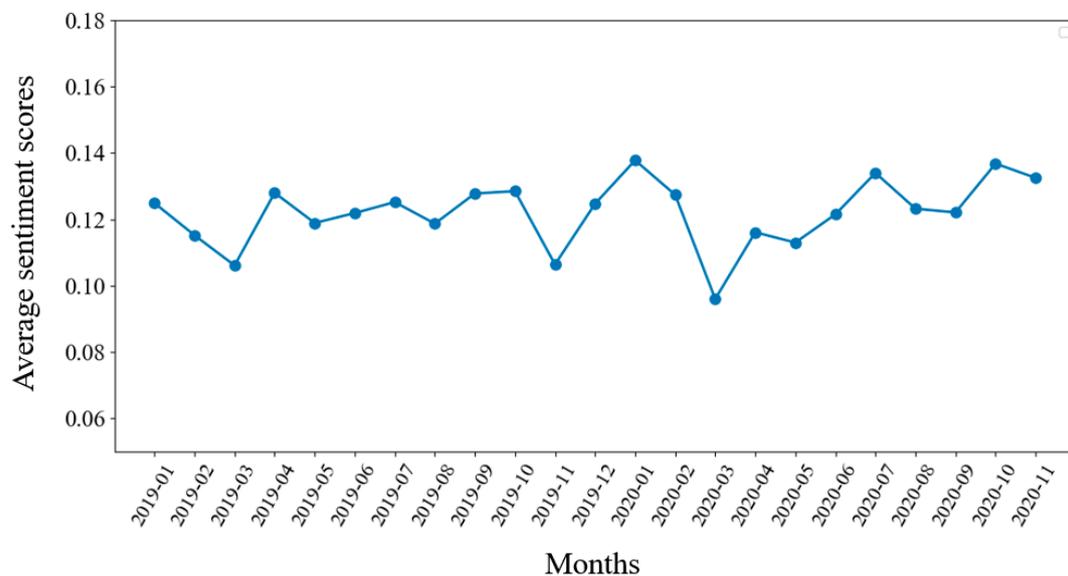


Figure 2. Distribution of public attitudes towards AVs.

3.2. Variable Selection by Random Forests

The RF algorithm was employed to select the most important candidate independent variables for further analyses. The dependent variable in the RF model was the sentiment scores of tweets related to AVs, which was a continuous variable. The candidate independent variables were 53 potential impacting factors, which are listed in Table 2. In this RF model, the importance of all these variables that may affect the public's attitudes towards AV was calculated and ranked.

After calculating the variable importance, the word-cloud function in "wordcloud2" (a visualization toolkit in python) was used to generate the word-cloud representation in Figure 3. This visual representation provided a view of all the candidate impacting factors, in which the importance of each variable was defined by its size (the larger, the more significant). The importance of some variables was relatively prominent, such as "mobility", "blind spot", and "Tesla", etc. The correlation between these variables and sentiment scores was further analyzed through a linear mixed model.

In addition, Figure 4 illustrates the top 20 variables ordered by importance. It is obvious that "mobility" was the top-ranked one, with variable importance of 0.202. The advent of AV technologies can further enhance public mobility [49], which greatly affected people's attitudes towards it. Then, "drunk driving", "blind spot", and "Tesla" were ranked second to fourth, with variable importance of 0.121, 0.105, and 0.084. The importance of the remaining 16 variables ranged from 0.011 to 0.037.

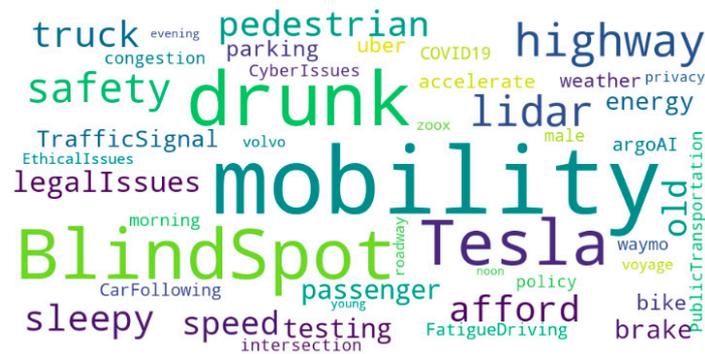


Figure 3. Word cloud of potential impacting factors.

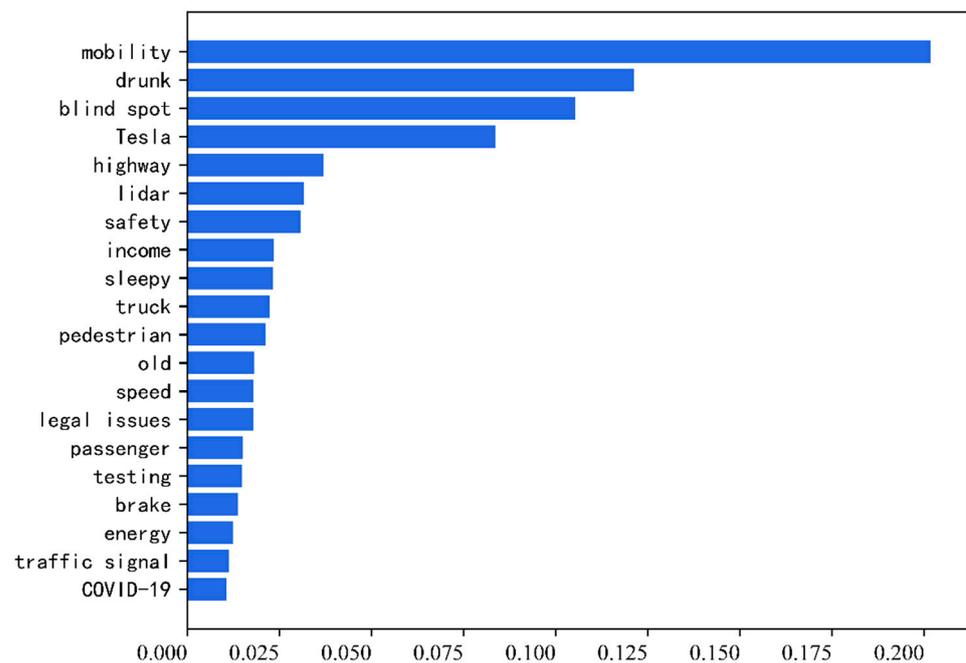


Figure 4. Variable importance in random forests (top 20).

From another perspective, 8 of these 20 variables belonged to Topic 2 (people), indicating that human attributes and driver behavior were the hot spots of public attention to AVs. Topic 7 (AV-related characteristics) contained many important factors, including “mobility”, “safety”, “legal issues”, and “energy”, indicating that while AV technologies bring benefits to people, they also bring some legal and environmental conflicts. In addition, “lidar”, “truck”, “blind spot”, and “speed” were three important factors that belonged to Topic 3 (vehicles). Variables, such as “COVID-19”, “testing”, “highway”, “traffic signal”, and “Tesla”, also showed strong importance.

3.3. Impacting Factors on the Public’s Attitude towards AVs Using a Linear Mixed Model

Figure 5 shows the distribution of the sentiment scores grouped by the subjectivity in tweets. The boxplot displays variation in samples of a statistical population and the results showed that there was a large number of outliers in the “very objective” group, without obvious skewness. In the “very objective” group, 86.1% (310,026 in 360,148) of the sentiment scores was 0, representing neutral attitudes. The distribution of the “very subjective” group had fewer outliers and the median was about 0.2. In addition, the medians of the “objective” group and the “subjective” group were 0.14 and 0.35, respectively. These three groups showed a slightly positive attitude towards AVs. Additionally, Kruskal–Wallis tests by ranks (the nonparametric equivalent of one-way analysis of variance (ANOVA)) were

employed to examine whether there were significant differences among the distribution of sentiment scores in these four subjectivity groups [50]. The results showed that the distributions of the sentiment scores in four groups were significantly inconsistent ($p < 0.01$). Thus, it is necessary to treat subjectivity as a random effect in a linear mixed model.

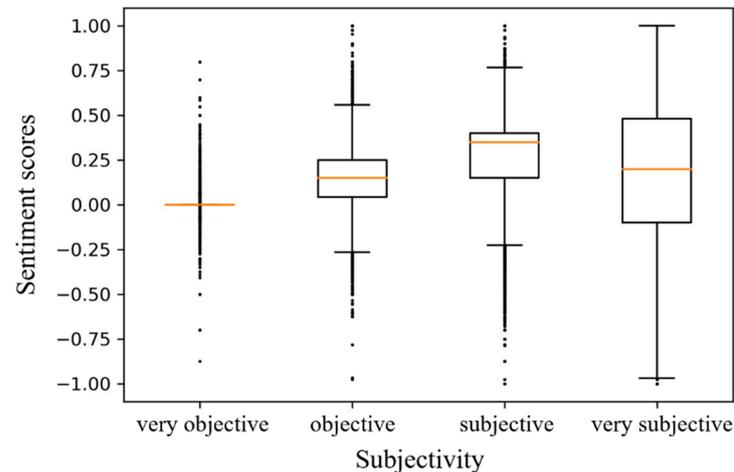


Figure 5. Distribution of the sentiment scores grouped by the subjectivity of tweets.

To further investigate the significant impacting factors on sentiment scores, the linear mixed model was run twice. The objective of the first linear mixed model was to judge the valid variables. The top 20 variables ranked in descending order of variable importance calculated by RF were used as the inputs of the first linear mixed model. After eliminating the insignificant variables ($p > 0.05$), ten significant variables were chosen as fixed effects for the second linear mixed model, including an AV-related company (“Tesla”), an AV-related characteristic (“mobility”), a road-related variable (“highway”), people-related variables (“drunk”, “sleepy”, and “pedestrian”), an event-related variable (“COVID-19”), and vehicle-related variables (“lidar”, “blind spot”, and “speed”), while the subjectivity in tweets was selected as a random effect.

A violin plot is a method of plotting numeric data, which is more informative than a plain boxplot by showing the full distribution of the data with the addition of a density plot on each side. Figure 6 shows the distribution of the sentiment scores of tweets containing each fixed effect. n is the number of tweets with each fixed effect. Among the 18,667 tweets relevant to “mobility”, the sentiment scores were mostly larger than 0, which had a maximum median value of 0.27. Negative attitudes were expressed more frequently than the other emotions towards the variable “drunk” ($n = 2723$). In terms of the “blind spot” ($n = 659$), the majority of Twitter users were skeptical and worried about it. “Tesla” gained most people’s attention, due to the number of tweets mentioning Tesla being the largest, 79,678 in total. The median values of sentiment scores of the tweets related to “highway” ($n = 4842$), “lidar” ($n = 17,126$), and “speed” ($n = 8536$) were all larger than 0 (0.07, 0.15, and 0.1, respectively). As for the tweets related to “sleepy” ($n = 3973$), “pedestrian” ($n = 6882$), and “COVID-19” ($n = 7009$), the median values of sentiment scores of these tweets were all 0.

The results of the final linear mixed model are illustrated in Table 3, which shows the estimated coefficients of fixed effects and random effects. The “mobility” was significantly positively associated with sentiment score ($t = 48.12$, $p < 0.01$). Most people hold the same view that AV technologies contributed to improving daily commuting ability, especially for people with disabilities [3]. Drunk ($t = -79.42$, $p < 0.01$) and blind spot ($t = -72.29$, $p < 0.01$) were significantly negatively associated with sentiment scores. It might indicate that the public had a low tolerance for drunk driving, which was illegal even if on AVs occasion. As for the variable “sleepy” ($t = -23.84$, $p < 0.01$), it was also negatively correlated with the dependent variable. This might show that people were worried that the AV system would

lead to drivers' excessive trust so they completely abandoned the control of the vehicle. There were some tweets expressing strong dissatisfaction about the situation where the driver fell asleep when the autopilot function was activated on the highway.

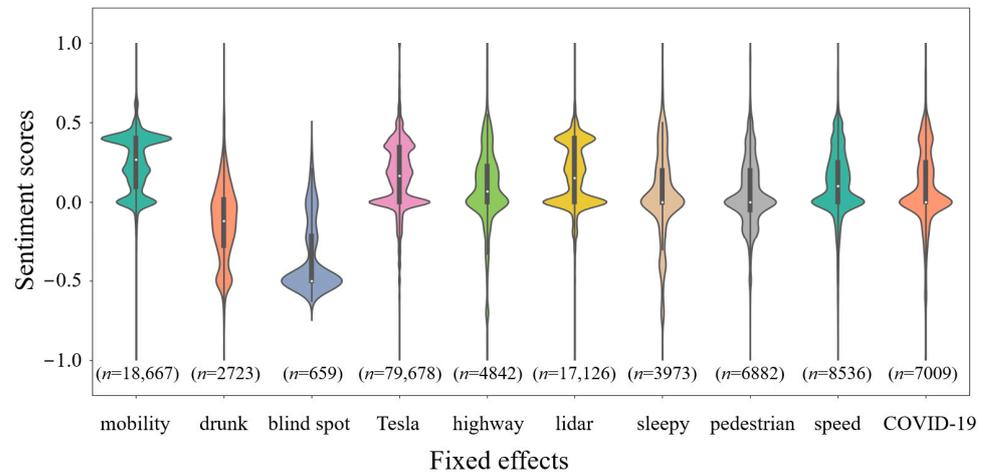


Figure 6. Distribution of sentiment scores of fixed effects related to AVs from Twitter data.

Table 3. Linear mixed model results.

Fixed Effects Estimates					
Fixed Effects	Estimate	Standard Error	DF	t-Value	Pr > t
mobility	7.294×10^{-2}	1.516×10^{-3}	9.541×10^5	48.12	<0.01
drunk	-3.123×10^{-1}	3.932×10^{-3}	9.541×10^5	-79.42	<0.01
blind spot	-5.759×10^{-1}	7.966×10^{-3}	9.541×10^5	-72.29	<0.01
Tesla	7.884×10^{-3}	7.608×10^{-4}	9.541×10^5	10.36	<0.01
highway	-4.450×10^{-2}	2.948×10^{-3}	9.541×10^5	-15.10	<0.01
lidar	3.057×10^{-2}	1.576×10^{-3}	9.541×10^5	19.40	<0.01
sleepy	-7.760×10^{-2}	3.256×10^{-3}	9.541×10^5	-23.84	<0.01
pedestrian	-6.316×10^{-2}	2.472×10^{-3}	9.541×10^5	-25.55	<0.01
speed	-1.215×10^{-2}	2.224×10^{-3}	9.541×10^5	-5.46	<0.01
COVID-19	-3.365×10^{-2}	2.451×10^{-3}	9.541×10^5	-13.73	<0.01
Random effect estimates					
Random effect	Variance		Standard Deviation		
Subjectivity	0.009897		0.09948		
Residual	0.041745		0.20431		

In addition, due to the great progress in the technology of vehicle-mounted sensor equipment, such as millimeter-wave radar and high-definition cameras, the variable “lidar” ($t = 19.40$, $p < 0.01$) was positively correlated with sentiment scores. However, the “blind spot” remained a public concern ($t = -72.29$, $p < 0.01$). In most AV solutions, due to the limitations of the lidar vertical field-of-view range and the overhead installation method, there will be a perception blind spot that is difficult to cover by lidar in the near-field area around the body. Potential low obstacles, such as pets and children, could be extremely risky.

As the vulnerable group in the road environment, “pedestrians” ($t = -25.55$, $p < 0.01$) generally had low trust in AV from the perspective of their safety. Although vehicles were equipped with increasingly sophisticated safety and crash-avoidance technology, pedestrian fatalities have risen slightly [2]. As for the AV-related companies, relevant tweets further demonstrated that “Tesla” ($t = 10.36$, $p < 0.01$) contributed to increasing the public’s favorability and trust in AVs. Further, “highway” ($t = -15.10$, $p < 0.01$) was negatively associated with sentiment scores. Despite extraordinary efforts from many pioneers in tech and automaking, fully AV technologies on the highway are still out of

reach except in special trial programs. In terms of speed ($t = -5.46$, $p < 0.01$), the public was worried about the high speed advertised by AVs.

Restrictions on public activities have been in place because of “COVID-19” ($t = -13.73$, $p < 0.01$), resulting in significant changes in mobility patterns and the stagnation of autonomous vehicles tests. It could be the reason making people generate negative attitudes towards AVs during the pandemic period.

The public also used Twitter to report accidents or casualties caused by AV, due to which artificial intelligence and AVs have always faced tough questioning. This study set the variable “safety” which included detailed words, such as “crash”, “accident”, “injured”, etc. However, the influence of “safety” on the public’s attitudes towards AVs was not significant ($p > 0.05$), so it was not contained in the final model. The public may have mixed feelings about the safety benefits of AVs. On the one hand, AV technologies have potential in improving traffic safety. On the other hand, the public still worries about the safety of current AV technologies.

In addition, a linear regression model (without considering random effects) was employed by using the “lm” (<https://www.r-project.org/> (accessed on 16 April 2021)) function in the statistical software R (version 4.1.0) to be compared with the linear mixed model. As demonstrated in Table 4, the AIC and BIC of the linear mixed model were $-322,623$ and $-322,459$, which were much lower than those in the linear regression model ($-124,912$ and $-124,759$ for AIC and BIC, respectively). This indicated the superiority of the linear mixed model.

Table 4. Comparison of the linear mixed model and linear regression model.

	AIC	BIC
Linear mixed model	$-322,623$	$-322,459$
Linear regression model	$-124,912$	$-124,759$

4. Discussion

Among all the significant impacting factors explored by the linear mixed model on the public’s attitudes towards AVs, “mobility” was the most important one according to the results of RF and it was positively correlated with sentiment scores. A study found that people’s awareness of mobility-related developments can increase the acceptance of driverless shuttles [8]. Increased mobility, especially for vulnerable people, was one of the main perceived advantages of AVs. The elderly and the disabled will be able to better obtain medical services and participate in society with improved mobility [51]. Moreover, the higher mobility brought about by AVs in the future will enhance people’s ability to travel from one place to another, which may lead to people’s willingness to live farther away from the city. Faced with these changes, it may cause a chain effect and be a challenge for population distribution, education structure, the transportation system, and so on. It is worth thinking about for policymakers.

This study also found that the public was more inclined to express a negative attitude towards AVs when the public expressed themselves with the following words: “drunk”, “sleepy”, “pedestrian”, “speed”, and “blind spot”. Thousands of fatalities are attributable to drunk driving every year despite the attempts made to warn and educate drivers by the government in the methods of propaganda and even criminal punishment. Although negative attitudes were expressed by the public towards drunk driving regarding AVs, it remained a question whether we should punish someone who is under alcohol’s influence riding in the driver’s seat of an autonomous car in the same way that we would punish someone who is controlling a vehicle [52].

Drivers are also strictly forbidden to sleep while driving in the current legal system. It can be found in this study that the public is opposed to this behavior when AV technologies have not been fully popularized. However, drunk or sleepy drivers may be allowed to drive autonomous vehicles when full driving automation is realized. In addition, the efforts of legislatures are needed to decrease the public’s concerns. The redefining of the word

“operate” is supposed to be considered by state legislatures [53]. Traditional laws mean that the operator of a motor vehicle actively controls the vehicle, which is not the case with AVs.

The public consideration of pedestrian safety issues and the speed of AVs has also affected their attitudes towards AVs. The pedestrian fatalities have risen slightly [2], despite the high technologies and new theories [54], which have been widely applied in AVs. Policymakers are supposed to take more considerations related to pedestrian safety when formulating traffic policies regarding AVs. There are still many people who refuse to sit in high-speed vehicles controlled by software. However, AVs contribute to a reduction in speed variance in mixed traffic conditions with both AVs and human-driven vehicles, which help to decrease the probability of collisions [55]. Therefore, people’s understanding of AVs needs to be guided, since there is an amount of uncertainty before the actual integration of full AV technology occurs.

The variable “lidar” was found to be positively correlated with sentiment scores in this study, while “blind spot” showed a negative relationship with the public’s attitudes towards AVs, which seemed to be a contradiction. Vehicle equipment, such as 3D lidar, stereo vision cameras and thermal cameras, can be helpful in pedestrian recognition and tracking [56,57]. However, the hidden danger in blind spots has not been eliminated yet. The limitations of the lidar and relevant equipment can lead to blind spots, such as low obstacles (e.g., pets and children). Lidar manufacturers should increase investment in the research and development of lidar products, aiming for better performance, smaller size, and lower cost, which performed as an important guarantee for the safety of AVs.

Tesla is one of the pioneers to push AVs into the market and has recognized the need to shift its innovation from the mechanical parts of the car to its electronics and software [58]. Positive attitudes were found in 65% of tweets mentioning Tesla in this study due to its contribution to the development of AVs. However, sufficient supervision is still needed for electric vehicles companies, such as Tesla, since they have been an important market segment. Private companies always tend to exaggerate the benefits of their products; the autopilot functionality of Tesla is claimed to meet full self-driving capabilities in the future with software updates designed to improve functionality over time. An objective understanding of AVs is needed for consumers that current AV products still need further testing by the market.

In terms of the factor “COVID-19”, it created a spreading and ever-higher health threat to people and the manufacturing system, which incurred severe disruptions and complex issues to industrial networks [59], thus, making people generate negative attitudes towards AVs during the pandemic period. The most dramatic fluctuation in average sentiment scores shown in Figure 2 occurred in March 2020, which was seemingly affected by the outbreak of COVID-19. However, after more detailed analysis, this fluctuation was most likely influenced by a report released by the National Transportation Safety Board (NTSB) on 25 February 2020 [60], which determined the probable cause for the fatal 23 March 2018 crash of a Tesla Model X in Mountain View, California. This tragic crash clearly demonstrates the limitations of advanced driver assistance systems available to consumers today, which sparked public concern on social media about the safety of AVs. In addition, this period of the COVID-19 outbreak also witnessed the development of autonomous delivery vehicles that had the potential to radically change the way groceries are delivered to customers’ homes [61]. A higher acceptance towards AVs can be predicted as more needs in people’s life are met by AVs.

5. Conclusions

This study aims to propose a method by using large-scale social media data to investigate key factors that affect the public’s attitudes and acceptance towards AVs. A total of 945,151 Twitter data related to AVs and 53 candidate variables from seven categories were extracted using the web scraping method. Then, sentiment analysis was used to measure the public’s attitudes on AVs by calculating sentiment scores. Random forests algorithm was employed to preliminarily select candidate independent variables according to their

importance, while a linear mixed model was utilized to explore the significant impacting factors on public attitudes towards AVs considering the unobserved heterogeneities caused by the subjectivity level in tweets.

Through random forests algorithm and linear mixed model analyses, several factors were found significantly correlated with sentiment scores. To be specific, “mobility” had the largest impact on public attitudes toward AVs and it was positively correlated with sentiment scores. In addition, people were also likely to express positive feelings when talking about words, such as “lidar” and “Tesla”, related to high technologies. Conversely, factors, such as “drunk”, “blind spot”, “COVID-19”, “pedestrian”, “sleepy”, and “highway”, were found to have significantly negative effects on the attitudes of the public. In addition, compared with the linear regression model, which was without considering subjectivity as random effects, the linear mixed model had much lower AIC and BIC, which indicated the superiority of the linear mixed model.

In this study, the use of social media data provided an opportunity for the collection of comprehensive information that might be representative enough for the public’s attitudes towards AVs. Traditional survey methods, such as field investigations and online questionnaires, were limited in understanding the opinions of the masses due to the lack of comprehensive and effective information retrieval and collection. This study showed that extracting data from social media platforms can be considered an alternative method, for its large quantity, timeliness, and effectiveness.

The results of this study are beneficial to policymakers, automotive industries, technology companies, and general consumers. It can help policymakers and legislators create plans and designate new laws based on public opinions. As shown by the top negative topics, drunk driving is an essential factor, which shows the strong concern of the public. It is worth considering by policymakers and legislators to define the harmfulness of drunk drivers in AVs. Developers and manufacturers of AVs can rethink their commercial strategy and product positioning according to the voice of customers. As mentioned previously, many social media users hold a suspicious attitude towards AVs for safety issues. Developers and manufacturers of AVs need to take note of this and formulate better solutions for user safety through integrated improvements in hardware and software. In addition, this study also proposed a complete and efficient technical route from social media platform data collection, data processing to modeling analysis at a low cost, which can be applied to other transportation-related themes.

There are some limitations in this study that should be addressed in future work. Although 53 potential impacting factors that may affect the public’s attitudes towards AVs were considered, there are definitely other related factors, such as personal information (age, education level, gender, etc.), locations, and so on. Additionally, it should be noted that in this study, only Twitter data were extracted and analyzed; richer and more effective results could be obtained if data from more popular social media are extracted. With the maturity and commercialization of AVs throughout the world, social media data containing public attitudes are constantly updated, so we will continue to add new potential impacting factors and extract data from more social media platforms to improve the accuracy and effectiveness of our models in the future.

Author Contributions: S.W.: Conceptualization, Data curation, Methodology, Writing—original draft. M.L.: Data curation, Methodology, Software, Validation. B.Y.: Conceptualization, Data curation, Methodology, Writing—original draft. S.B.: Conceptualization, Data curation, Methodology, Writing—original. Y.C.: Conceptualization, Funding acquisition, Writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This project was jointly supported by the National Natural Science Foundation of China (52102416), the Natural Science Foundation of Shanghai (22ZR1466000), and the Fundamental Research Funds for the Central Universities (22120220126).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Payre, W.; Cestac, J.; Delhomme, P. Intention to use a fully automated car: Attitudes and a priori acceptability. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *27*, 252–263. [\[CrossRef\]](#)
2. Combs, T.S.; Sandt, L.S.; Clamann, M.P.; McDonald, N.C. Automated vehicles and pedestrian safety: Exploring the promise and limits of pedestrian detection. *Am. J. Prev. Med.* **2019**, *56*, 1–7. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bennett, R.; Vijaygopal, R.; Kottasz, R. Attitudes towards autonomous vehicles among people with physical disabilities. *Transp. Res. Part A Policy Pract.* **2019**, *127*, 1–17. [\[CrossRef\]](#)
4. Fagnant, D.J.; Kockelman, K. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transp. Res. Part A Policy Pract.* **2015**, *77*, 167–181. [\[CrossRef\]](#)
5. Abraham, H.; Lee, C.; Brady, S.; Fitzgerald, C.; Mehler, B.; Reimer, B.; Coughlin, J.F. Autonomous Vehicles and Alternatives to Driving: Trust, Preferences, and Effects of Age. In Proceedings of the Transportation Research Board 96th Annual Meeting (TRB'17), Washington, DC, USA, 8–12 January 2017.
6. Menon, N.; Pinjari, A.; Zhang, Y.; Zou, L. Consumer Perception and Intended Adoption of Autonomous-Vehicle Technology: Findings from a University Population Survey (No. 16-5998). In Proceedings of the Transportation Research Board 95th Annual Meeting, Washington, DC, USA, 10–14 January 2016.
7. Das, S.; Dutta, A.; Lindheimer, T.; Jalayer, M.; Elgart, Z. YouTube as a source of information in understanding autonomous vehicle consumers: Natural language processing study. *Transp. Res. Rec.* **2019**, *2673*, 242–253. [\[CrossRef\]](#)
8. Nordhoff, S.; De Winter, J.; Kyriakidis, M.; Van Arem, B.; Happee, R. Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *J. Adv. Transp.* **2018**, *2018*, 5382192. [\[CrossRef\]](#)
9. Pettigrew, S.; Worrall, C.; Talati, Z.; Fritschi, L.; Norman, R. Dimensions of attitudes to autonomous vehicles. *Urban Plan. Transp. Res.* **2019**, *7*, 19–33. [\[CrossRef\]](#)
10. Bansal, P.; Kockelman, K.M.; Singh, A. Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 1–14. [\[CrossRef\]](#)
11. Kassens-Noor, E.; Kotval-Karamchandani, Z.; Cai, M. Willingness to ride and perceptions of autonomous public transit. *Transp. Res. Part A Policy Pract.* **2020**, *138*, 92–104. [\[CrossRef\]](#)
12. Gkartzonikas, C.; Gkritza, K. What have we learned? A review of stated preference and choice studies on autonomous vehicles. *Transp. Res. Part C Emerg. Technol.* **2019**, *98*, 323–337. [\[CrossRef\]](#)
13. Liu, Y.; Li, Y.; Li, W. Natural language processing approach for appraisal of passenger satisfaction and service quality of public transportation. *IET Intell. Transp. Syst.* **2019**, *13*, 1701–1707. [\[CrossRef\]](#)
14. Nielsen, T.A.S.; Hausteijn, S. On sceptics and enthusiasts: What are the expectations towards self-driving cars? *Transp. Policy* **2018**, *66*, 49–55. [\[CrossRef\]](#)
15. Hulse, L.M.; Xie, H.; Galea, E.R. Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Saf. Sci.* **2018**, *102*, 1–13. [\[CrossRef\]](#)
16. Hardman, S.; Berliner, R.; Tal, G. Who will be the early adopters of automated vehicles? Insights from a survey of electric vehicle owners in the United States. *Transp. Res. Part D Transp. Environ.* **2019**, *71*, 248–264. [\[CrossRef\]](#)
17. Begg, D. *A 2050 Vision for London: What Are the Implications of Driverless Transport?* Transport Times: London, UK, 2014.
18. Casley, S.V.; Quartulli, A.M.; Jardim, A.S. *A Study of Public Acceptance of Autonomous Cars*; Worcester Polytechnic Institute: Worcester, MA, USA, 2013.
19. Daziano, R.A.; Sarrias, M.; Leard, B. Are consumers willing to pay to let cars drive for them? Analyzing response to autonomous vehicles. *Transp. Res. Part C Emerg. Technol.* **2017**, *78*, 150–164. [\[CrossRef\]](#)
20. Schoettle, B.; Sivak, M. *A Survey of Public Opinion about Autonomous and Self-Driving Vehicles in the US, the UK, and Australia*; University of Michigan, Transportation Research Institute: Ann Arbor, MI, USA, 2014.
21. Zhang, Z.; He, Q.; Gao, J.; Ni, M. A deep learning approach for detecting traffic accidents from social media data. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 580–596. [\[CrossRef\]](#)
22. Rajendran, L.; Thesinghraj, P. The impact of new media on traditional media. *Middle-East J. Sci. Res.* **2014**, *22*, 609–616.
23. Mpofu, P.; Asak, M.O.; Salawu, A. Facebook groups as transnational counter public sphere for diasporic communities. *Cogent Arts Humanit.* **2022**, *9*, 2027598. [\[CrossRef\]](#)
24. Kapidzic, S.; Neuberger, C.; Frey, F.; Stieglitz, S.; Mirbabaie, M. How News Websites Refer to Twitter: A Content Analysis of Twitter Sources in Journalism. *J. Stud.* **2022**, *23*, 1247–1268. [\[CrossRef\]](#)
25. Wengel, Y.; Ma, L.; Ma, Y.; Apollo, M.; Maciuk, K.; Ashton, A.S. The TikTok effect on destination development: Famous overnight, now what? *J. Outdoor Recreat. Tour.* **2022**, *37*, 100458. [\[CrossRef\]](#)
26. Ding, Y.; Korolov, R.; Wallace, W.A.; Wang, X.C. How are sentiments on autonomous vehicles influenced? An analysis using Twitter feeds. *Transp. Res. Part C Emerg. Technol.* **2021**, *131*, 103356. [\[CrossRef\]](#)

27. Lin, L.; Ni, M.; He, Q.; Gao, J.; Sadek, A.W. Modeling the impacts of inclement weather on freeway traffic speed: Exploratory study with social media data. *Transp. Res. Rec.* **2015**, *2482*, 82–89. [[CrossRef](#)]
28. Ni, M.; He, Q.; Gao, J. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1623–1632. [[CrossRef](#)]
29. Cottrill, C.; Gault, P.; Yeboah, G.; Nelson, J.D.; Anable, J.; Budd, T. Tweeting Transit: An examination of social media strategies for transport information management during a large event. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 421–432. [[CrossRef](#)]
30. Huang, A.; Gallegos, L.; Lerman, K. Travel analytics: Understanding how destination choice and business clusters are connected based on social media data. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 245–256. [[CrossRef](#)]
31. Anta, A.F.; Chiroque, L.N.; Morere, P.; Santos, A. Sentiment analysis and topic detection of Spanish tweets: A comparative study of NLP techniques. *Proces. Leng. Nat.* **2013**, *50*, 45–52.
32. Jelodar, H.; Wang, Y.; Rabbani, M.; Ahmadi, S.B.B.; Boukela, L.; Zhao, R.; Larik, R.S.A. A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments. *Multimed. Tools Appl.* **2021**, *80*, 4155–4181. [[CrossRef](#)]
33. Stambaugh, C.L. Social media and primary commercial service airports. *Transp. Res. Rec.* **2013**, *2325*, 76–86. [[CrossRef](#)]
34. Liu, B. Sentiment analysis and opinion mining. In *Synthesis Lectures on Human Language Technologies*; Morgan and Claypool Publishers: San Rafael, CA, USA, 2012; Volume 5, pp. 1–167.
35. Kohl, C.; Mostafa, D.; Böhm, M.; Krčmar, H. Disruption of individual mobility ahead? A longitudinal study of risk and benefit perceptions of self-driving cars on twitter. In Proceedings of the 13th International Conference on Wirtschaftsinformatik, St. Gallen, Switzerland, 12–15 February 2017.
36. Wang, Y.; Bao, S.; Du, W.; Ye, Z.; Sayer, J.R. A spectral power analysis of driving behavior changes during the transition from nondistracted to distracted. *Traffic Inj. Prev.* **2017**, *18*, 826–831. [[CrossRef](#)]
37. Yu, B.; Bao, S.; Feng, F.; Sayer, J. Examination and prediction of drivers' reaction when provided with V2I communication-based intersection maneuver strategies. *Transp. Res. Part C Emerg. Technol.* **2019**, *106*, 17–28. [[CrossRef](#)]
38. Jermakian, J.S.; Bao, S.; Buonarosa, M.L.; Sayer, J.R.; Farmer, C.M. Effects of an integrated collision warning system on teenage driver behavior. *J. Saf. Res.* **2017**, *61*, 65–75. [[CrossRef](#)] [[PubMed](#)]
39. Wright, M.N.; Ziegler, A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [[CrossRef](#)]
40. Yu, R.; Abdel-Aty, M. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Saf. Sci.* **2014**, *63*, 50–56. [[CrossRef](#)]
41. Li, J.; Guo, J.; Wijnands, J.S.; Yu, R.; Xu, C.; Stevenson, M. Assessing injury severity of secondary incidents using support vector machines. *J. Transp. Saf. Secur.* **2020**, *14*, 197–216. [[CrossRef](#)]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
43. Xu, Y.; Bao, S.; Pradhan, A. Modeling drivers' reaction when being tailgated: A Random Forests method. *J. Saf. Res.* **2021**, *78*, 28–35. [[CrossRef](#)]
44. Yu, B.; Chen, Y.; Bao, S. Quantifying visual road environment to establish a speeding prediction model: An examination using naturalistic driving data. *Accid. Anal. Prev.* **2019**, *129*, 289–298. [[CrossRef](#)]
45. Li, Y.; Yu, B.; Wang, B.; Lee, T.H.; Banu, M. Online quality inspection of ultrasonic composite welding by combining artificial intelligence technologies with welding process signatures. *Mater. Des.* **2020**, *194*, 108912. [[CrossRef](#)]
46. Louppe, G.; Wehenkel, L.; Sutter, A.; Geurts, P. Understanding variable importances in forests of randomized trees. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 431–439.
47. Yu, B.; Bao, S.; Zhang, Y.; Sullivan, J.; Flannagan, M. Measurement and prediction of driver trust in automated vehicle technologies: An application of hand position transition probability matrix. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102957. [[CrossRef](#)]
48. Burnham, K.P.; Anderson, D.R. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **2004**, *33*, 261–304. [[CrossRef](#)]
49. Zhao, L.; Malikopoulos, A.A. Enhanced mobility with connectivity and automation: A review of shared autonomous vehicle systems. *arXiv* **2019**, arXiv:1905.12602. [[CrossRef](#)]
50. Feng, F.; Bao, S.; Hampshire, R.C.; Delp, M. Drivers overtaking bicyclists—An examination using naturalistic driving data. *Accid. Anal. Prev.* **2018**, *115*, 98–109. [[CrossRef](#)] [[PubMed](#)]
51. Yang, J.; Coughlin, J.F. In-vehicle technology for self-driving cars: Advantages and challenges for aging drivers. *Int. J. Automot. Technol.* **2014**, *15*, 333–340. [[CrossRef](#)]
52. Hanna, K.L. Old Laws, New Tricks: Drunk Driving and Autonomous Vehicles. *Jurimetrics* **2014**, *55*, 275.
53. Douma, F.; Palodichuk, S.A. Criminal liability issues created by autonomous vehicles. *St. Clara L. Rev.* **2012**, *52*, 1157.
54. Mahadevan, K.; Somanath, S.; Sharlin, E. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.
55. Khondaker, B.; Kattan, L. Variable speed limit: An overview. *Transp. Lett.* **2015**, *7*, 264–278. [[CrossRef](#)]
56. Wang, H.; Wang, B.; Liu, B.; Meng, X.; Yang, G. Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle. *Robot. Auton. Syst.* **2017**, *88*, 71–78. [[CrossRef](#)]
57. Chen, Z.; Huang, X. Pedestrian detection for autonomous vehicle using multi-spectral cameras. *IEEE Trans. Intell. Veh.* **2019**, *4*, 211–219. [[CrossRef](#)]

58. Mallozzi, P.; Pelliccione, P.; Knauss, A.; Berger, C.; Mohammadiha, N. Autonomous Vehicles: State of the art, future trends, and challenges. *Automot. Syst. Softw. Eng.* **2019**, *347–367*. [[CrossRef](#)]
59. Li, X.; Wang, B.; Liu, C.; Freiheit, T.; Epureanu, B.I. Intelligent manufacturing systems in COVID-19 pandemic and beyond: Framework and impact assessment. *Chin. J. Mech. Eng.* **2020**, *33*, 58. [[CrossRef](#)]
60. Tesla Crash Investigation Yields 9 NTSB Safety Recommendations. Available online: <https://www.nts.gov/news/press-releases/Pages/NR20200225.aspx> (accessed on 12 September 2022).
61. Kapser, S.; Abdelrahman, M.; Bernecker, T. Autonomous delivery vehicles to fight the spread of COVID-19—How do men and women differ in their acceptance? *Transp. Res. Part A Policy Pract.* **2021**, *148*, 183–198. [[CrossRef](#)] [[PubMed](#)]