Public Perceptions Towards Autonomous Vehicle technology: A sentiment analysis using Twitter data and natural language processing models.

Md Tanvir Ashraf1, Md Younus Ahamed2, Donald Adjeroh2, Brian Powell2, Kakan Dey1

1Department of Civil and Environmental Engineering, West Virginia University, Morgantown, West Virginia.

2 Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, West Virginia.

**Abstract.** Autonomous Vehicles (AVs) are expected to shape the future of the transportation system by eliminating human driver error-related traffic crashes and enhancing traffic safety. The public’s attitudes toward AVs are an important factor in adopting this technology on a wide scale. Previously several studies analyzed the AV sentiments using survey-based methods which are costly to obtain, and often sample size of these surveys is small. Few previous studies explored time series AV perception using social media data and did not consider spatio-temporal variations in public attitudes. Also, they didn’t explore specific features of AVs such as AVs a shared mode, or the impact of COVID-19 on public attitudes. To bridge this gap, this study attempts to explore spatio-temporal variations in public perceptions of autonomous vehicles (AVs) using multiple years of Twitter data within the US and using specific AV related tweets such as shared AVs and driverless taxis. For the sentiment analysis, the state-of-the-art natural language processing (NLP) technique called Bidirectional Encoder Representations from Transformers (BERT) was used, and the results were compared with those of using the Support Vector Machine (SVM) model. The BERT model performed better in terms of the accuracy value in predicting the sentiments as positive, neutral, and negative AV tweets. The results show that public opinion on Twitter, at least, seems to be positive over the study period. The authors also attempt to model public sentiment on shared AVs (SAVs) showing that the public is more hesitant about their use after the COVID-19 pandemic.

**Keywords:** Autonomous Vehicle, Sentiment Analysis, NLP, Twitter, social media, Traffic Safety.

1. **Introduction**

Autonomous Vehicle (AV) technology has the potential to revolutionize the future of transportation systems by reducing the amount and severity of traffic crashes and increasing mobility for people with disabilities and the elderly. AVs can eliminate human-related errors such as speeding, misjudging other drivers, alcohol impairment, distraction, and fatigue from the driving scenarios and even perform better than human drivers in certain driving scenarios (e.g., no blind spots, faster and more precise control of steering, brakes, and acceleration). Despite the enormous safety and mobility benefits of AVs, they can pose new and serious crash risks (e.g., cyber-attacks, driving in inclement weather, etc).

Few recent incidents such as Uber-pedestrian crash and Tesla’s Model X crash in 2018 impacted the public perception and credibility of the Autonomous Vehicle technology. A question was raised that whether or not people would want to surrender control of their cars. Would people prefer to spend their commute reading, or watching videos, or do they prefer to control their own vehicle? These questions are almost as important as developing the technology itself. The negative perceptions about AV technology can hamper its deployment. Thus, analyzing trends in AV perceptions could help policy makers to identify key factors causing negative attitudes and propose new or modified policies to overcome those setbacks.

While many have attempted to explore AV perceptions through the use of survey data the authors decided to consider them through the vein of social media. Survey-based sentiment analysis has several limitations such as (i) surveys are costly and resource intensive, (ii) low response rate, (iii) social desirability bias, and (iv) assumptions that participants can understand the scenarios [1]. Social media platforms such as Twitter have become the major platform for public to share their opinions [1]**.** Previous studies that analyzed AV perceptions using survey-based methods expressed that exposure or familiarity with AV technology is a significant factor in people’s attitudes toward the technology. Although few studies explored the AV perceptions using Twitter data, none of them explored the spatial-temporal patterns of AV sentiments among the places that introduced AV testing on public roads. The objective of this paper is to analyze the spatial-temporal pattern of AV perceptions using geotagged Twitter data to see the effect of AV testing as well as a few recent incidents (e.g., AV crash, COVID-19) that could potentially impact AV perception.

1. **Literature Review**

In recent years, numerous research studies are being conducted to better understand different aspects of AV technology. Previous and current studies mainly focused on AV characteristics such as motion planning, safety benefits, environmental aspects, and adaptation process. The widescale adoption of AVs not only depend on its technology but also public perception or views towards this technology.

Social media has become a promising data source for studies related to public [1,2]. A growing number of studies are utilizing the Twitter data as an alternative to survey-based methods to analyze the public opinions or sentiments on trending topics [1]. Hu et al. [3] explored the public opinion towards COVID-19 vaccines using Twitter data in the United States. The study found that the positive sentiment towards vaccines increased over the time and negative sentiments decreased in most of the states.

Previously several studies have explored the public perception of AV technology using survey-based and social media data [4-7]. A study done by Pyrialakou et al. [4] used a stated preference survey data which was conducted in Phoenix, Arizona, a metropolitan area to understand the difference between various population groups concerning the perceived safety of driving, cycling, and walking near AVs. In this study, the authors found that cycling was considered a less safe activity compared to walking and driving. Results of this study also suggest that people who had previous exposure to AVs or AV testing had a more positive attitude towards them. Rahman et al. [5] performed a qualitative analysis of the perceptions of pedestrians and bicyclists in terms of sharing the road with an autonomous vehicle. For the analysis, this study used survey responses from pedestrians and bicyclists in Pittsburgh, Pennsylvania area. The authors found that pedestrians showed more positive attitudes towards AVs compared to negative attitudes. Public perceptions were significantly affected by their concerns regarding AV safety, familiarity with the technology, and their household automobile ownership. Penmetsa et al. [6] performed another study using Pittsburgh, Pennsylvania data to see the public perceptions in terms of vulnerable road users. This study also concluded that public attitudes towards AVs are more likely to be positive if they had previous interactions with AVs. The authors of this study recommended that policymakers should allow the public to interact with AVs and make AV deployment easier [6].

Several studies in the past have also used social media data such as Twitter data to analyze the public perceptions towards autonomous vehicles [1,8,9].Ding et al. [1] used Twitter data to understand the public perceptions towards AVs using sentiment analysis method. This study also performed topic modeling and time series analysis. For the sentiment analysis this study used Support Vector Machine (SVM) method and for the topic modeling they used LDA method. This study found positive attitude towards autonomous vehicles with sentiment biases towards different AV terms [1]. On the other hand, Penmetsa et al. [8] used Twitter data before and after two AV involved crashes (March 2018 Uber-pedestrian crash and Tesla Model X’s crash while it was operating in self-driving mode) to see the shift in public perceptions towards AVs. The study found that after the crash events negative tweets about AVs increased by 32% [8].

Sentiment analysis using Twitter data followed Naïve Bayes classifier (NBC) [10,11], ensemble learning classification method [12], and Support Vector Machine (SVM) methods have their limitations as they are not able to consider the context of the sentences and they struggle to deal with long patterns. The limitations of these modeling techniques can be overcome by the Bidirectional Encoder Representation from Transformers (BERT) model. The BERT model can process longer sentences and can also include the context of the sentences in the analysis process.

Furthermore, previous studies did not consider specific characteristics of AVs in the analysis process. For example, AVs will shape the future of the shared modes such as driverless taxis or driverless for-hire vehicles (i.e Uber, Lyft, etc.). Especially after the COVID-19 outbreak, it will be beneficial to know how people are thinking about SAVs and how their concerns can be alleviated, or positive attitudes can be further improved. Thus, this study contributes to the existing AV literature in the following three aspects- this study analyzed public perception towards autonomous vehicles using readily available social media data in terms of AV use as a shared mode and transportation mode as a whole. This study applied transformer-based model BERT to handle the context of the tweets which presents huge performance improvements compared to Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN). Finally, this study explored the positive and negative attitudes toward AVs, facilitating their deployment and increasing overall traffic safety.

**3. Methodology**

**Timeline

Description automatically generated**The overall framework for data collection, preprocessing, and analysis is outlined in Figure 1. For the sentiment analysis, this study utilizes machine learning models such as Support Vector Machines (SVM) and Bidirectional Encoder Representations from Transformers (BERT) models. The details of the methodology are discussed below:

**3.1 Data Collection and Preprocessing**

Collection of Twitter data involves getting access to Twitter API (Application Programming Interfaces) through academic license in the Twitter developer account. Thus, the authors created a developer account with academic license to collect the Tweets data. Python library called “Tweepy” was used for the searching and filtering of the tweets based on the time, language, and keywords. Multiple keywords such as “Autonomous Vehicle”, “Driverless Car”, “Self-Driving car/vehicle”, “Driverless Taxi”, and “Autonomous Driving” were used for the data collection. The tweets were limited to English language only. Moreover, filtering was done so that the collected data do not list any retweets. Using the above-mentioned filters almost sixteen thousand tweets were collected for the period of January 2019 to December 2021 that were generated within the US. The collected information about the tweets contains user information, user location, text data, time of the tweet, and user engagements with the tweets. However, only the text, user location, and time of the tweets were used for the analysis. As tweets often have non-standard spelling, inconsistent punctuation, and special characters, the raw tweets need to be pre-processed before applying sentiment analysis model on it. The stop words were removed, along with the punctuation, repeating characters, emojis, hashtags, and numerals. After that tokenization, lemmatization and stemming process was performed on the tweets to make them suitable for analysis.

**Fig. 1.** Overview of the proposed data collection, preprocessing, and spatio-temporal analysis

## **3.2 Modeling technique**

In this study, a state-of-the-art natural language processing technique called Bidirectional Encoder Representations from Transformers (BERT) was used for sentiment classification. For comparison, another machine learning algorithm called Support Vector Machine (SVM) was also utilized to predict public sentiments toward AVs. For the training of these models, 700 randomly selected tweets were manually labeled in three categories. The tweets were labeled based on the perception expressed in the tweets, subject knowledge, previous literature, and relevance to the transportation system [1]. To remove subjectivity from the annotated tweets authors classified the tweets separately and compared the labels to finalize the training dataset.

## **Support Vector Machine (SVM) model**

Support Vector Machine (SVM) model was used to perform the sentiment analysis which also serves as a basis of comparison. SVM is a supervised classification model and proved to be a useful machine learning tool to classify the text data in previous studies [13,14,15]. SVM is a discriminative classifier which works by defining an optimal hyperplane based on the labeled training data [1]. Manually labeled 700 tweets were used to train and test the SVM classifier. The SVM classifier was fed with cleaned Twitter data after removing the stop words, unnecessary words and tokenization. The best performing SVM classifier was selected after tuning the hyper-parameters.   The accuracy of the SVM classifier in terms of predicting the sentiments was compared with the BERT model.

## **Bidirectional Encoder Representations from Transformers (BERT) model**

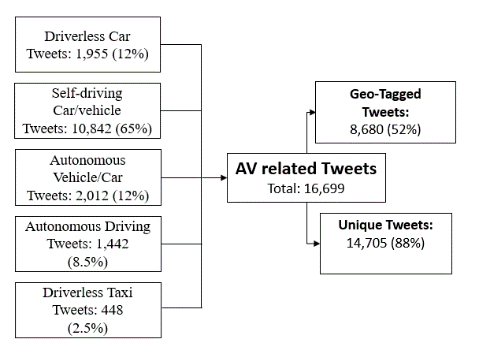
Previously applied machine learning models such as SVM and LSTM models are sensitive to long-term patterns. Furthermore, previous models are unidirectional (left to right architecture) where every token can only attend to previous tokens in the self-attention layers [16]. The BERT model alleviates the unidirectional and long-term pattern problem by using a “masked language model” (MLM) pre-training objective. The objective of the pre-training is to predict the original vocabulary id from the randomly masked tokens from MLM [17]. The MLM models enable the bidirectional training of the transformers where the model learns the context of each word from the words appearing before and after it which was not possible in the left-to-right language model pre-training In the BERT modeling framework, there are two steps- the first one is the pre-training and the second one is the fine-tuning [17].  In the pre-training phase, the BERT model was trained on large amounts of unlabeled data from Wikipedia (2,500 million words) and Book Corpus (800 million words) over two different pre-training tasks. In the fine-tuning task, the BERT model is initialized with the pre-trained parameters and all of the parameters are fine-tuned using labeled data specific to the task. For example, in this study, the BERT model was fine-tuned using the AV tweet data to perform the sentiment prediction task. The token representation vectors of the collected AV tweets are used as input for the network. The token representation vectors are the sum of three representation vectors called: a word embedding vector, a position vector, and a sentence vector [18]. The position vector is related to the position of a token within a sentence and a sentence vector-only is used when there are multiple sentences in a text. In the developed BERT model, each input sentence of the AV tweets is attached with an initial artificial token donated by [CLS]. The output representation from the artificial token is used to feed the classification layer during fine-tuning the model on the AV sentiment classification task. The classification layer in the BERT model is a typical softmax layer [18]. Explainable artificial intelligence (XAI) is a set of procedures and techniques that permits users to understand and rely on the outcomes and results produced by machine learning algorithms [19]. In this study, Shapley Additive explanations (SHAP) which were proposed by [20] been used to interpret the outputs of the sentiment classification models.

**4. Data**

This section presents a brief overview of the collected AV tweets based on their spatio-temporal distribution and word distribution within the text of the tweets.

**4.1 Characteristics of the Tweets**

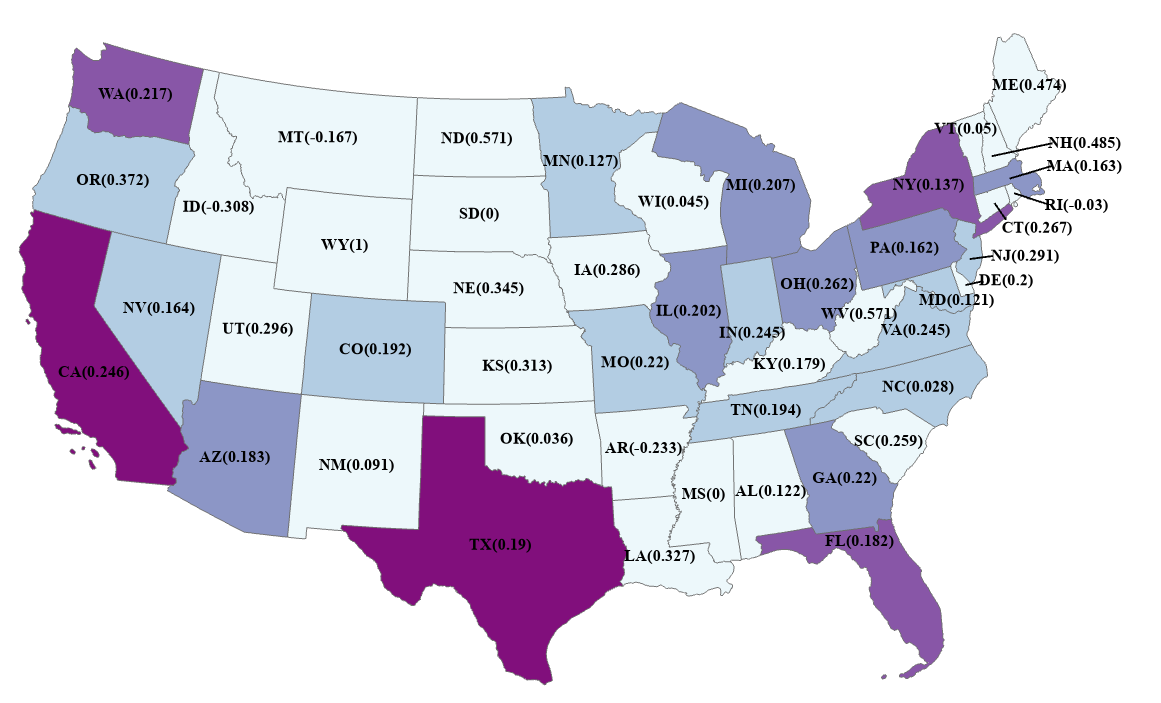
AV related tweets were collected from the US that were posted between January 2019 to December 2021. To filter the tweets within the US, tweet geo-location was used which is the boundary of the location from where the user posted the tweet. Figure 2 presents the distribution of the collected tweets based on the keywords. Using the keywords and location filter 16,699 tweets were retrieved. Among these tweets 65% was related to keyword “self-driving car or vehicle”. Other keywords such as “driverless car”, “autonomous vehicle/car”, “autonomous driving”, and “driverless taxi” yielded 11%, 12%, 9%, and 2.5% of the total tweets, respectively. There were about 14,705 non-repeated or unique tweets in the collected dataset which was used for the analysis. Moreover, only 52% of the tweets contained the US states info that is required for the analysis (Figure 2).

**4.2 Spatio-Temporal Distribution of the tweets**

Spatio-Temporal distribution of the collected tweets was analyzed to conduct the validity and representativeness of the collected dataset. Figure 3 represents the spatial distribution of the tweets based on the US states. The number of tweets originating from different states had similar distribution over the years. States such as California, Florida, New York, Texas, and Washington had a higher frequency of tweets compared to other states.

On the other hand, states such as Iowa, Wisconsin, Kansas, Missouri, New Mexico, Delaware, and Arkansas generated low number of tweets when compared to the above-mentioned states. The distribution is in line with the general understanding that highly populated states could generate more tweets compared to low population-density states.

**Fig. 2.** Distribution of the collected tweets



**Fig. 3.** Spatial Distribution of the AV tweets and sentiments

**4.3 Word Distribution in the Tweets**

After the preprocessing of the tweets, keyword trend analysis was performed to see the most frequently mentioned words in the tweets. Word clouds (not shown) for (a) positive tweets, and (b) for negative labeled tweets generated using top 25 words within the manually labeled tweets. The word cloud gives a view on which words were used for positively labeled tweets and which words were used in negatively used words. As expected, in both positive and negative labeled tweets terms such as autonomous vehicle, self-driving cars, driverless car was most frequent. Apart from that, for positively labeled tweets, words like ‘good things’, ‘future’, ‘safety’ were prominent. The negative tweets were made up of terms such as ‘crash,’ ‘attack,’ ‘accident,’ and ‘robot’ implying not only a darker view, but perhaps a less rational one.

**4.4 Manually Labeled Tweets**

In total, 700 randomly selected tweets were manually labeled by three different persons. After that, the labels were combined. The manually labelled tweets contains 260 tweets in positive attitude, 240 tweets in negative attitude, and 200 tweets in neutral attitude. After the labelling was completed, each tweet was preprocessed following the text processing steps such as tokenization, stop word removal, lemmatization, and stemming. The dataset was then divided into training and test dataset to develop the SVM and BERT model. Based on the performance of the models, the best performed classifier was used to identify sentiments using whole dataset.

**5. Result and Discussion**

This section presents a comparison of the SVM and BERT models in predicting AV sentiments as well as an analysis of the predicted sentiments to address the research questions proposed in this study.

**5.1 Model Results**

To develop the SVM and BERT model, 80% of the manually labeled tweets was used as training dataset and 20% was used as validation dataset. Accuracy, recall, precision, and f1-score were used as performance measures to select the best performing model in predicting the AV sentiments. The selected performance measures can be presented as:

* Recall or Sensitivity=
* Precision= = class agreement of the data labels with the positive labels
* Accuracy== overall accuracy
* weighted average of the recall and precision

where TP is the true positive values, TN is the true negative values, FP is the false positive values, and FN is the false negative values.

The BERT model was implemented in PyTorch and the pre-trained uncased base model of BERT was fine-tuned to fit the AV tweets data. The hyper-parameters that were used to develop the BERT model is presented in Table 1. The models were developed in a single intel-Core i7 CPU with 16 GB of RAM and the runtime for the model was three hours.

|  |  |
| --- | --- |
| **Table 1.** Hyperparameters used in the BERT model. | |
| **Parameter** | **Value** |
| Batch Size | 32 |
| Learning Rate | 3e-05 |
| Epsilon | 1e-08 |
| Max epochs | 2 |
| Max Sequence length | 128 |
| Optimizer | Adam |

After the training, the prediction of sentiments was done using the test dataset. The prediction result of the BERT and SVM models are provided in Table 2. From the result presented in Table 3, it can be seen from that the BERT model performed better in terms of overall model accuracy. The overall accuracy of the BERT model was 90% whereas the accuracy of the SVM model was 81%. The precision, recall, and f1-score for positive class data was 87%, 91%, 89% for BERT model, respectively. And for SVM model the precision, recall, and f1-score was 81%, 83%, and 82%, respectively. The result shows an improvement in score for BERT model in terms of precision, recall, and f1 values over the SVM model. Similarly, for negative and neutral tweets BERT model performed better than the SVM model based on the performance measure values of precision, recall, and f1-score. As BERT model performed better in training and test dataset, it was applied on the rest of the dataset to measure the sentiment scores. The following sections presents discussion on the AV sentiments based on different measures.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 2.** Performance measures of the BERT and SVM model | | | | | |
| **Model** | **Class** | **Precision** | **Recall** | **F1-Score** | **Validation Accuracy** |
| **BERT** | Positive | 0.87 | 0.91 | 0.89 | 0.90 |
| Negative | 0.80 | 0.80 | 0.80 |
| Neutral | 0.81 | 0.733 | 0.77 |
| **SVM** | Positive | 0.81 | 0.83 | 0.82 | 0.81 |
| Negative | 0.75 | 0.60 | 0.67 |
| Neutral | 0.50 | 0.60 | 0.55 |

**5.2 Overall AV perceptions over time**

To evaluate the overall public attitude towards AVs, the monthly count of positive, negative, and neutral tweets was plotted from January 2019 to December 2021. Figure 4 shows the tweet count by month and sentiments. As seen from this figure, the number of monthly positive tweets was higher than the number of neutral and negative tweets during this period. This result indicates that there are more positive sentiments toward AVs than negative sentiments within the US. The monthly tweet count for each category had a decreasing trend simply because of a lower number of total tweets in 2020 and 2021 compared to the year 2019. The average of all tweets’ (positive, negative, neutral) sentiment scores in each month was plotted in the secondary axis of Figure 4.

|  |
| --- |
|  |
| **Fig. 4.** Distribution of public attitudes towards AVs over time, Monthly distribution of tweets, and Average monthly sentiment |

Figure 4 shows the distribution of the public attitude and acceptance toward AVs during the period from January 2019 to December 2021. The maximum value of the average AV sentiment was 0.32 in May 2020, and the minimum value of the average sentiment score was 0.012 in April 2021. However, it never went below 0. From Figure 4, the trend of the average AV sentiment in the US decreased over the years. Overall, the distribution of average sentiment scores reflected that the public in the US was optimistic about the prospects of AV technologies.

**5.3 Comparison between AV testing location and non-testing location**

Public perception of any technology significantly depends on their familiarity or exposure to that technology. AV technology is still under development and very few people had experienced or interacted with AVs on public roads. Recently, several states in the US allowed companies to test their Autonomous Driving Systems (ADS) on public roads. For example, GM/Cruise LLC is currently operating approximately 230 ADS-equipped vehicles on public roads in San Francisco, California. GM/Cruise LLC is also operating approximately 40 AVs in Milford, Michigan. However, these are limited to private roads and test team use only. Another major AV testing location is Phoenix, Arizona where 12 AVs are in operation on public roads. There could be significantly different perceptions about AVs among the states that allow public road testing of AVs and the states that do not have AV testing programs for public roads. Figure 3 shows the average AV sentiment scores in major states (with at least 100 tweets in positive, neutral, and negative categories) in the US. The result showed that the AV testing states had consistently higher values compared to other states. The average sentiment score in California, Michigan, Arizona, and New Jersey was 0.23 whereas for other states the average value was 0.18. This means that people are more optimistic about AV technology when they experience it or had some kind of exposure to the technology. This result showed that other states in the US need to introduce testing of ADS technology in public roads. This will not only help people to become familiar with the technology but also facilitate the process of developing ADS that is reliable and efficient.

**5.4 Effect of COVID-19 on AV perception**

The COVID-19 pandemic, which is also known as the coronavirus pandemic, had caused huge changes in public attitude, behavior, and acceptance of different technologies into the daily life. The outbreak first came to light in December 2021 and declared global pandemic by World Health Organization (WHO) on 11th March 2020. Autonomous Vehicle (AV) technology development and testing was also affected by the COVID-19 pandemic. To assess the impact of pandemic on publics’ attitude toward AVs, the average sentiment score before and after the pandemic was compared. The null and alternative hypothesis for this test is presented below. The null hypothesis states that the average AV sentiment score before the pandemic (before the pandemic is equal to the average AV sentiment score after the pandemic (. The alternative hypothesis for this test is that the two means are not equal.

The test statistic Z\* was calculated based on equation (1) where is the average sentiment score before and after COVID-19, Sa and Sb are the standard deviation, and na and nb are the sample size.

(1)

Overall perception towards AVs as well as perception toward AVs as a shared mode of transport was tested under this hypothesis. The average sentiment scores for these two AV features are presented in Table 3. As seen from the Table 3, before the pandemic the overall AV perception and perception toward AVs as a shared mode was both higher than the after-pandemic values. Before the pandemic the average value of overall AV perception was 0.206 and after the pandemic it was 0.1767. The test statistic Z\* based on the equation 2 was 2.06 which is larger than 1.96, the critical value for a two-tailed test at 5% significance level. This indicates that the null hypothesis is rejected, and the before pandemic AV perception was significantly different than the after-pandemic situation. The result concluded that after the pandemic the overall AV perception in the US got decreased and more people is concerned about this technology and its deployment.

One of the main benefits of AVs is that they can revolutionize the shared mobility area such as for-hire vehicles (Uber and taxis). In addition to looking at overall AV sentiments, AVs as a shared mode were also analyzed in this study. For this, tweets related to keywords ‘driverless taxi’ and ‘shared autonomous vehicle’ were analyzed to calculate the difference between sentiments before and after the COVID-19 outbreak. The tweets were fed into the developed BERT model and the sentiments were predicted. The result showed that people had an overall positive sentiment toward AVs as a shared mode. However, after the COVID-19 outbreak, the positive sentiment decreased. Before the COVID-19 outbreak the average sentiment, the value was 0.297 and after the COVID-19 outbreak, the value was 0.20 (Table 3). The test statistic value was 5.43 which is much larger than 1.96 which means the difference is significant. This result indicates that before the pandemic people

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3.** Effect of COVID-19 on AV sentiments | | | |
| **AV Feature** | **Average AV Sentiment** | | Z\* |
| **Before COVID-19** | **After COVID-19** |  |
| Overall AV perceptions | 0.206 | 0.1767 | 2.06 |
| AV as a shared mode of transport (e.g., UBER, Lyft, etc.) | 0.297 | 0.20 | 5.43 |

were more optimistic about the AVs and their potential as a shared mode of transport. However, after the pandemic people had more negative feelings about using AVs as shared mode. The result is intuitive as after the pandemic people became more concerned about sharing the spaces with others.

**6. Model Interpretation**

In addition to increasing the performance of the sentiment classification models, it is important to develop a better understanding of how the model works. In this study, an explainable AI-based approach was adopted. The developed sentiment classification model was explained with the SHAP (Shapley Additive exPlanations) values. SHAP explains the output of the model by assigning each feature importance related to a particular prediction [20]. To show how SHAP explains individual predictions, one positive and negative sentiment tweet was randomly selected. The tweets are shown below:

***Positive Tweet:*** “This is sooo true. And I know how amazing and wonderful self-driving is. The media has to drive a Tesla to know what it actually feels like.”

***Negative Tweet:*** “Now self-driving cars are racist. #CantCatchABreak. Study finds a potential risk with self-driving cars: failure to detect pedestrians.”

The associated local interpretation graph is shown in Figure 5a and 5b. The figures show the different features of these two tweets contributing to pushing the predicted truth probability from the base value to individual value. The features with positive values are shown in red color (i.e., words in red increases the base value) and the negative values are shown in blue color (i.e., words in blue reduces the base value). If the local value is lower than the base value, the tweet is a negative tweet and if the local value is higher than the base value then the tweet is a positive tweet. For example, in Figure 5a, the base value is -0.317 and the local value is 0.471 which means that this tweet is a positive tweet. The original tweet shows that the author expressed positive attitude toward AVs. The word amazing, wonderful, and like had highest impact on the model output. On the other hand, Figure 5b shows that the local value is -0.904 which is less than the base value. So, the tweet is a negative tweet. The original tweet shows that words such as risk, failure bears the negative which is also captured by the SHAP values.

|  |  |
| --- | --- |
|  |  |
| **Fig. 5.** How each feature pushes the model output from the average value to the output value for  individual samples. (a) Positive sentiment. (b) Negative sentiment. | |

**7. Contribution and policy implications**

The contribution of this research is threefold. First, this study performed a spatio-temporal analysis of AV sentiments in the US based on several aspects of the AV technology such as AV testing activities, impacts of COVID-19, and AVs as a shared mode of transport. The result showed that AV testing activities made people more optimistic about the technology. Also, the COVID-19 pandemic lowed the public sentiment towards AVs as a shared mode of transport (e.g., Uber, Lyft). Second, this study utilized a state-of-the-art NLP model called Bidirectional Encoder Representations from Transformers (BERT) to predict the public sentiment which can handle longer text data and predict the public sentiments more effectively. In addition to increasing the performance of sentiment classification, it is important to develop a better understanding of how the model works. As such, SHAP values were used in this study to interpret how the sentiment classification models worked. Finally, this study had major policy implications. Negative attitudes towards AVs were mainly related to the privacy and security issues of the autonomous driving system. Another negative aspect of AVs that was found in the tweets was the price of the AV driving system. To make AVs more popular AV manufacturers need to take these concerns into their technology development process. AVs should be made affordable so that general people can use them. Also, the privacy and security system of AVs should deal with continuous threats from hackers.

**8. Conclusion**

This study explored social media users’ attitudes towards autonomous vehicles by using large-scale spatio-temporal Twitter data from the US. Twitter data was collected using Twitter API with five keywords: ‘autonomous vehicle’, ‘self-driving car/vehicle’, ‘driverless car’, ‘autonomous driving’, and ‘driverless taxi’. Using these keywords, around sixteen thousand tweets were collected between January 2019 to December 2021 which originated within the US. Among the collected tweets 700 tweets were randomly selected which were then classified into three categories: Positive, Negative, and Neutral. The tweets were labeled based on subject knowledge and to remove subjectivity labeling was performed by three different persons. A state-of-the-art natural language processing model called Bidirectional Encoder Representations from Transformers (BERT) was used to classify the tweets into the three sentiment classes. To compare the predictive performance of the BERT model another machine learning tool called Support Vector Machine (SVM) was also developed. The performance measures of the developed models showed that the BERT model successfully predicted the correct sentiment with a validation accuracy of 90% whereas the SVM model had a validation accuracy of 81%. The effect of the COVID-19 pandemic was also explored in this study which showed that the overall AV perception decreased after the pandemic. In addition to that, this study also focused on public attitudes toward AVs as a shared mode of transport.

**References**

1. Ding et al. "How are sentiments on autonomous vehicles influenced? An analysis using Twitter feeds." Transportation research part C: emerging technologies 131 (2021): 103356.
2. Wojtowicz J at el. Use of social media by transportation agencies for traffic management. Transportation Research Record. 2016;2551(1):82-9.
3. Hu et al. Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. Journal of Medical Internet Research. 2021 Sep 10;23(9):e30854.
4. Pyrialakou VD et al. Perceptions of safety on a shared road: Driving, cycling, or walking near an autonomous vehicle. Journal of safety research. 2020 Feb 1; 72:249-58.
5. Rahman et al. "Sharing the road with autonomous vehicles: A qualitative analysis of the perceptions of pedestrians and bicyclists." Transportation research part F: traffic psychology and behaviour 78 (2021): 433-445.
6. Penmetsa et al. "Perceptions and expectations of autonomous vehicles–A snapshot of vulnerable road user opinion." Technological Forecasting and Social Change 143 (2019): 9-13.
7. Haboucha et al. "User preferences regarding autonomous vehicles." Transportation Research Part C: Emerging Technologies 78 (2017): 37-49.
8. Penmetsa et al. "Effects of the recent autonomous vehicle crashes on public perception of the technology." IATSS Research (2021).
9. Wang et al. "Investigating the Impacting Factors on The Public's Attitudes Towards Autonomous Vehicles Using Sentiment Analysis from Social Media Data." arXiv preprint arXiv:2108.03206 (2021).
10. Leung et al." Polytechnic University Department of Computer Science/Finance and Risk Engineering 2007 (2007): 123-156.
11. Buntoro et al. "Sentiments Analysis for Governor of East Java 2018 in Twitter." Sinkron: jurnal dan penelitianteknikinformatika 3, no. 2 (2019): 49-55.
12. Liu et al. "Co-training and visualizing sentiment evolvement for tweet events." In Proceedings of the 22nd international conference on World Wide Web, pp. 105-106. 2013.
13. D'Andrea et al. "Monitoring the public opinion about the vaccination topic from tweets analysis." Expert Systems with Applications 116 (2019): 209-226.
14. Xiong et al. "Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings." Neurocomputing 275 (2018): 2459-2466.
15. Hunt et al. "Fighting fake news during disasters." OR/MS Today 47, no. 1 (2020): 34-39.
16. Vaswani et al. "Attention is all you need." In Advances in neural information processing systems, pp. 5998-6008. 2017.
17. Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
18. Sousa et al. "BERT for stock market sentiment analysis." In 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1597-1601. IEEE, 2019.
19. Parsa AB et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accident Analysis & Prevention. 2020 Mar 1;136:105405.
20. Lundberg SM et al. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30.
21. Ribeiro MT et al. " Why should i trust you?" Explaining the predictions of any classifier. InProceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 1135-1144).
22. Štrumbelj E et al. Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems. 2014 Dec;41(3):647-65.