Observing the Effect of a Crash on Twitter Sentiment: Early Results from Time Series Data

Michael McGurrin
McGurrin Consulting
mike@mcgurrin.com

Background

The two topics of automated and electric vehicles generate many opinions. One approach used to automate the assessment of positive or negative opinions on these topics is *sentiment analysis*. Sentiment analysis can be conducted at multiple levels of detail, but a common use is to identify whether a piece of writing expresses a positive, negative, or neutral opinion on the subject matter. Computer algorithms are used to automate what would otherwise be an extremely labor-intensive process.

Tweets are often used as a data source because there are a very large number of tweets available for analysis, many express strong opinions, and there is an Application Programming Interface (API) provided by Twitter to make it easy to access and filter tweets.

I launched the AV and EV Sentiment Index project to collect and analyze tweets on each of these two topics on a daily basis, track how sentiment changes over time, and examine how events in the world affect the sentiments expressed on Twitter. The AV Sentiment Index and EV Sentiment Index websites display the results.

The initial versions sampled tweets that had been posted within the past seven days. I've been collecting data for about eight weeks, and the data already shows the effect of news events on the sentiment regarding automated vehicles. This is not the first time that crashes have been shown to effect Twitter sentiment regarding automated vehicles (see, for example, Jefferson and Mcdonald, 2019). However, previous effort have custom selected twitter search terms to match each incident, whereas this current work keeps a constant baseline of search terms.

Initial Observations

In November 2022, there was an eight-car pileup on the San Francisco Bay Bridge. Around December 21st and 22nd, multiple media sources reported that the cause of the pileup was a Tesla in "full self-driving" mode (see, for example, https://www.businessinsider.com/teslas-fsd-car-crash-california-phantom-braking-2022-12). Immediately, the number of negative tweets on automated vehicles went up significantly, as shown in Figure 1. The number of negative tweets in the seven-day samples remained high for about a week afterwards.

The ratio of positive to negative tweets on automated vehicles also shows this drop. Figure 2 plots this ratio over time. The blue band shows + / - one standard deviation from the average.

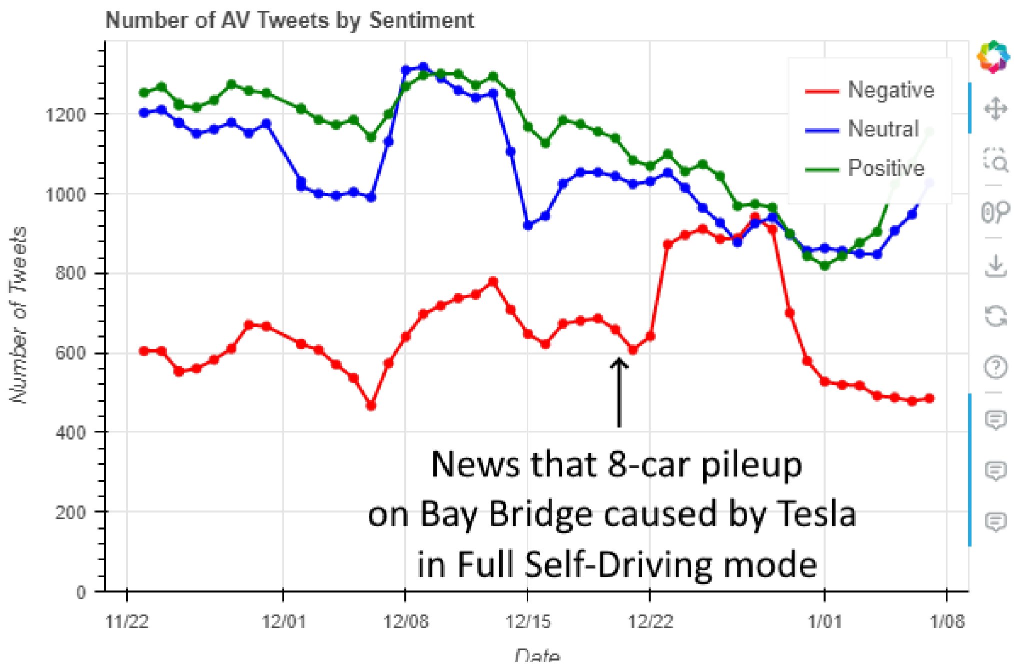

Figure 1. Number of AV tweets by sentiment,
showing the spike in negative tweets after a crash involving a Tesla appeared in the news.

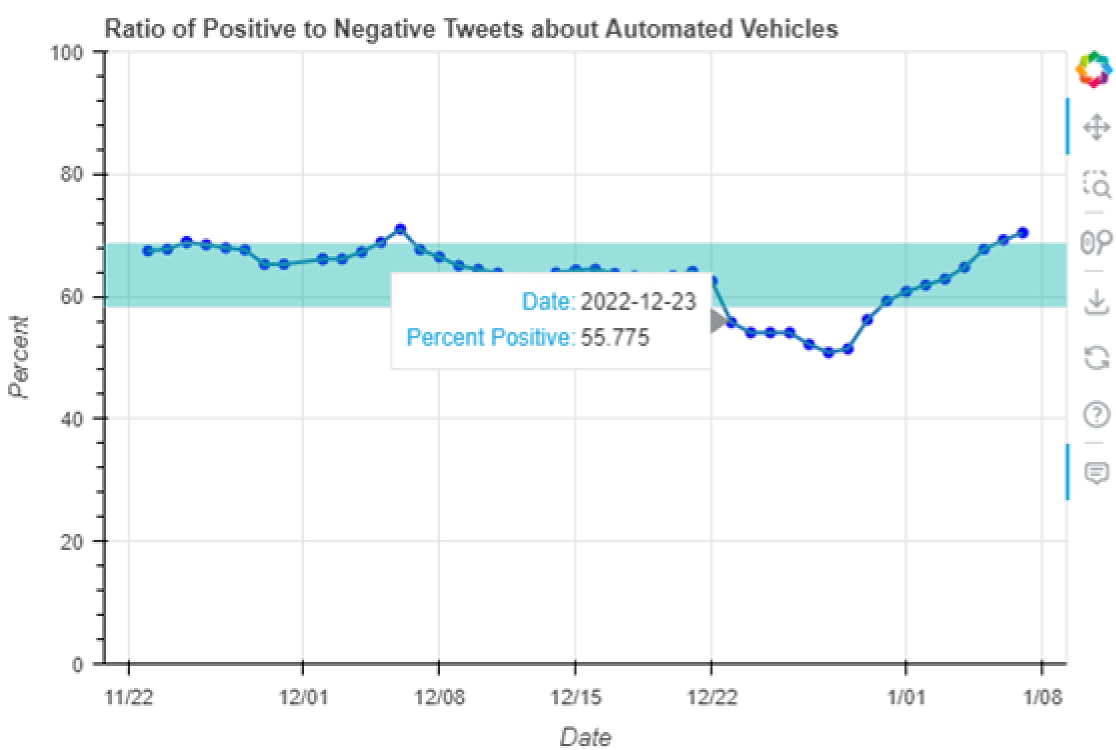

Figure 2. Ratio of positive to negative tweets. The blue band shows +/- one standard deviation.
A significant drop in the ratio can be seen after the news reports.

In order to further identify what may be the cause of sentiment changes, the analysis also tracks the top 100 one or two-word combinations that appear in the set of tweets. A comparison is made to identify the terms that appear in the top 100 for the current day but were absent from the top 100 the previous week. When looking at the list, the list of newly hot terms include “Bay Bridge,” “Tesla full,” “eight car,” “eight vehicle,” “vehicle crash,” and “tells police.” All indicating that indeed, the news of the pileup being caused by self-driving software is the likely cause for the increasingly negative sentiment.

Shortly after Christmas, the day with the highest negativity, California announced that they were banning Tesla from using the phrase “full self-driving” in their advertising (see, for example, https://www.pcmag.com/news/california-bans-tesla-from-calling-software-full-self-driving). And, when looking at words frequently appearing on tweets on December 28th, but not a week before, the terms “advertising” and “bans Tesla” appear on the list.

Interestingly, another crash involving a Tesla occurred the first week of January. In that crash, a Tesla ran over a cliff. However, it was quickly reported that the crash did not involve the self-driving features, and this crash does not appear to have affected the sentiment on Twitter regarding automated vehicles.

During the 2nd week of January, there is a rise in both the number of positive AND negative tweets, as shown in Figure 3.

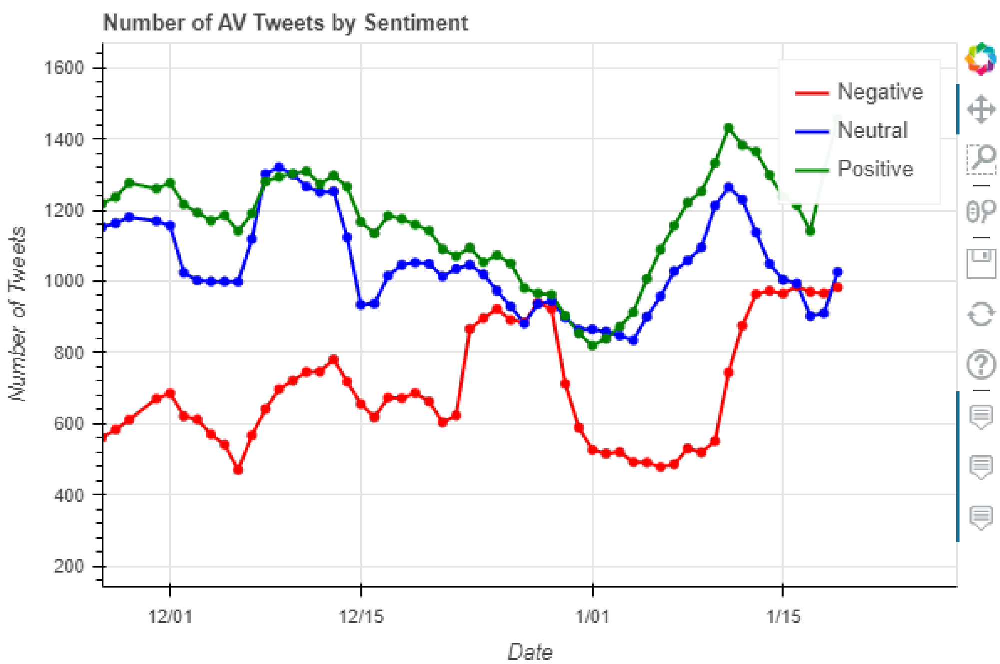

Figure 3. Number of Tweets by Sentiment, showing a rise in BOTH positive and negative tweets in the 2nd week of January.

When one examines the “hot words” for January 11th (i.e., those appearing frequently in the sample from the 4th through 11th, but not the week before) one sees words associated with the surveillance footage of the above-referenced crash, which had just come out that week (e.g., “Surveillance Footage,” Tesla crash,” “Bay Bridge”) but also words associated with the Consumer Electronics Show (CES), including Tesla’s announcement of a self-driving stroller (e.g., “CE,” “CES2023,” “baby stroller,” and “driving stroller”). The number of negative tweets remains high through January 19th, as the news concerning the allegedly staged full self-driving demonstration video. The new “hot words” for the 19th include “Tesla staged,” “Tesla video,” “engineer testifies,” “fake”, “faked,” and “video promoting.” At the same time, some of the “hot words” from the previous week concerning CES news, such as

"CES2023," "Las Vegas," and "baby stroller" are now missing. This can be seen in the word cloud for the top 100 words in automated vehicle tweets for the week preceding the 19th, as shown in Figure 4.

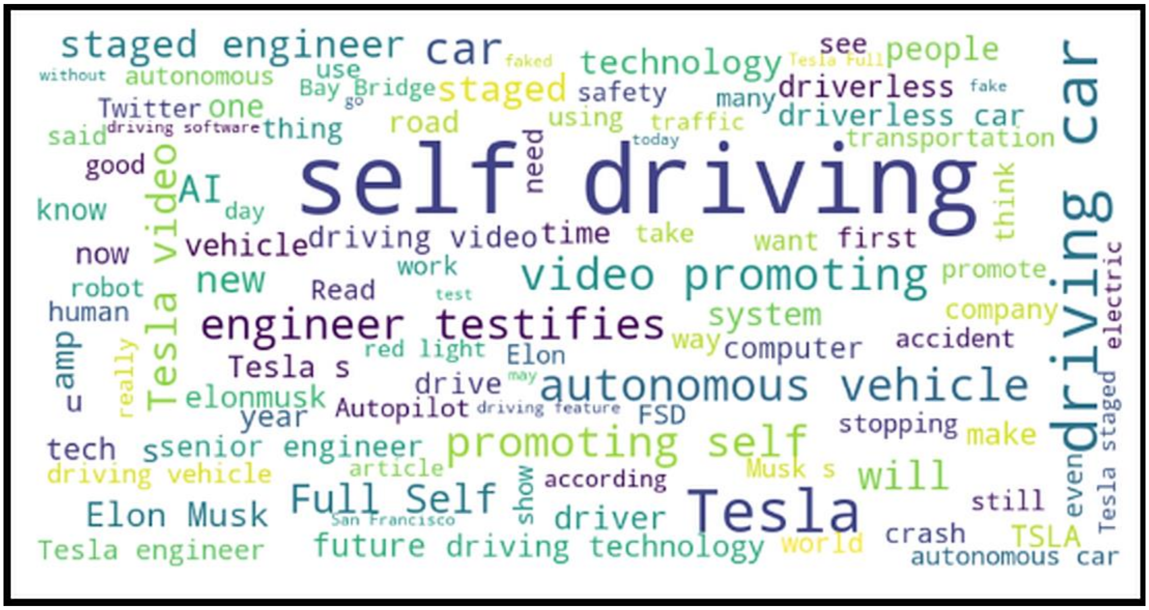

Figure 4. Word cloud of the 100 most frequently appearing one and two word phrases on tweets about automated vehicles made on January 17, 2023.

These results are highly preliminary, but they indicate that the influence of real-world events and news can be observed in the sentiment of tweets on the subject of automated vehicles.

The daily sentiment indices and hot words for both automated and electric vehicles can be seen on two web pages: mcgurrin.info/sentiment/AVsentiment.shtml and mcgurrin.info/sentiment/EVsentiment.shtml.

Next Steps

Data collection is ongoing, and a longer collection period will allow the examination of more correlations between real world events, news reports, and sentiment expressed on Twitter. However, the system is switching over to using a sample of the previous day's tweets, rather than a sample from the past week. There appear to be several hundred tweets per day on each subject (AV's and EV's), and this will allow changes to be identified more quickly.

In addition, the current system stores the top 100 words, but does not store the full content of the sampled tweets. A future revision will store tweets for some period of time, possibly a week to ten days.

Finally, Twitter has announced that the free basic API tier this system is based on will end in mid-February. If that happens and a new approach cannot be developed, the project may have to come to a premature end.

Appendix: Additional Background on the Approach

1. **What is Sentiment Analysis and how is it calculated?**

   Wikipedia defines sentiment analysis as *the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study* *affective states and subjective information.* In this case, the Twitter API is used to extract tweets that, based on the presence of key words, appear to be discussing either Automated Vehicles or Electric Vehicles. Then, for each of these tweets, the contents of the tweet is analyzed to determine the polarity (positive, neutral, or negative) of the sentiment expressed on the subject matter. Because this is highly labor intensive if done by hand, computer algorithms are used to identify the polarity of each tweet. Tweets about each of the two subjects are analyzed separately.

There are multiple approaches for using computers to assess sentiment, including various machine learning algorithms and neural networks. The AV and EV sentiment indices are calculated using VADER (Valence Aware Dictionary and sEntiment Reasoner), an open source software package. VADER is a lexicon and rule-based analysis tool that was specifically developed to perform well when analyzing social media posts such as tweets. Some minor additions were made to the lexicon to tailor it to the current subject matter.

This rule-based approach is rather simple when compared with some of the other approaches, however comparison studies have shown that it often performs as well or better than other approaches when analyzing social media and it has the advantage of running very quickly.

2. **Why analyze tweets?**

Twitter has hundreds of millions of users. Even though only a small percentage tweet about either automated or electric vehicles, the result is still thousands of new tweets per day on the these two topics. This analysis uses the Twitter API to collect a sample of up to 5,000 new tweets on each of the two subjects, once per day. It provides a large sampling of opinions, and these are readily accessed through their open API.

Tweets have disadvantages as well, however. Many bots simply repeat news posts from other sources, resulting in identical or (worse) nearly identical tweets that are simply repeats. Custom code in the AV and EV sentiment index analysis attempts to filter these out.

3. **How accurate is sentiment analysis?**

Machine based sentiment analysis can be difficult. Tweets can include negated words ("not good" is negative, whereas "good" is positive), slang words, emojis, and sarcasm. VADER is designed to handle many of these cases, but it is far from perfect. In fact, research has shown that human raters only agree with one another less than 80% of the time, even when dealing with just two categories (https://skystance.com/sentiment-analysis-everything-you-need-to-know/).

With three sentiment values and a roughly equal number of tweets expressing each sentiment, random chance would get about 33% right. The algorithm used in this work was tested using a similar set of tweets concerning automated vehicles (but which was originally generated using different key word selection criteria), this algorithm agreed with a human annotator 53% of the time. The goals of the indices, however, is not to get a precise value of the sentiment polarity, but rather to track changes over time, and how these changes correlate with external events.