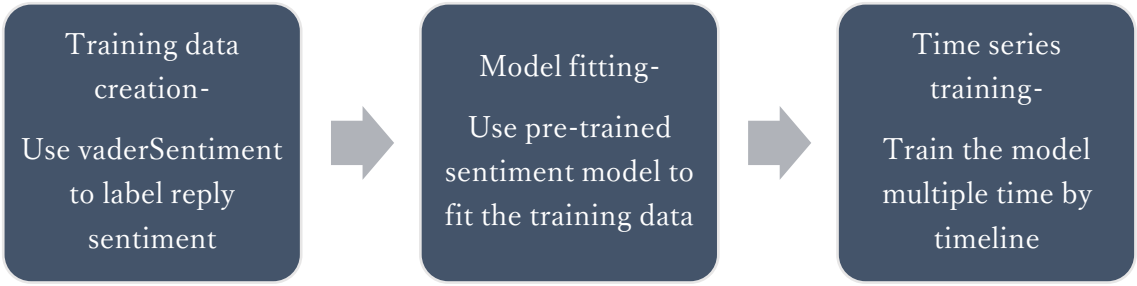
How to predict if Elon musk is saying something positive?

Elon musk is one of the “Twitter masters” whose tweets were highly watched. But do his tweets matters all the time? Can we build a natural language sentiment model to tell if his tweet is positive? Negative? Or it doesn’t really matter?
Here I’m demonstrating a time-series trained sentiment model that can tell the sentiment of the tweet while reflecting the context and history of someone’s tweets.



Case study – Elon’s tweets about his upcoming FSD

To start with, what is “FSD”???
Someone like myself, who is not a specialist in the automobile industry nor an expert on Elon Musk, would be confused when we see something like “FSD”.
That is why I wanted to develop a model that can just tell me if the tweet (or news, article, anything) is positive or negative without asking me to understand the whole context or any background knowledge.

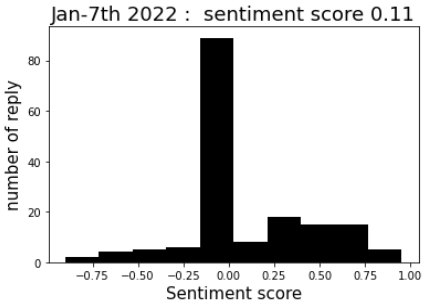
FSD: Tesla’s Full Self Driving system enables the user to free their hands and let the car drive itself. As it is still in beta version, Tesla currently puts users in a rather strict safety condition by asking users to pay attention to the road condition. If the car detects you are not paying enough attention to the road, it will give you a ‘strike,’ and if you reach a certain number of ‘strikes,’ you will be banned from using the FSD.

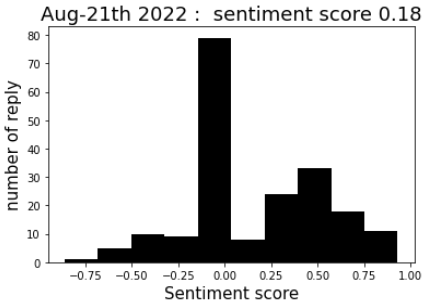
<u>When Elon Musk tweeted about the FSD price increase</u>

Jan 8, 2022

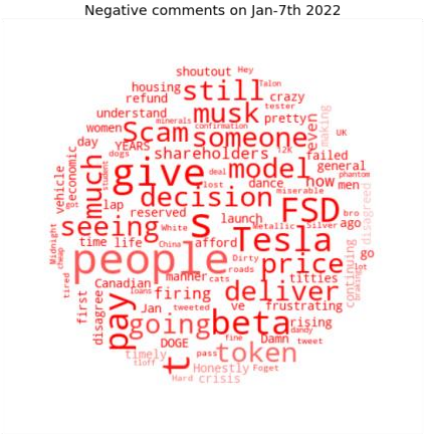
Aug 21, 2022

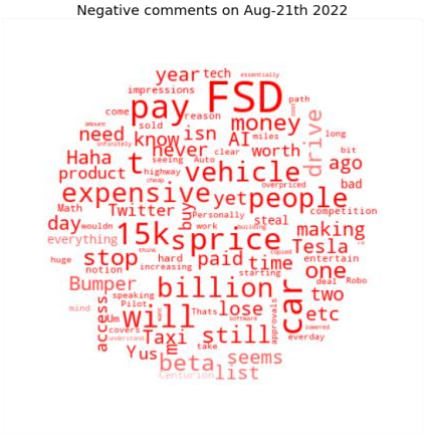
Most people replied with neutral sentiment but sentiment improved on Aug 21



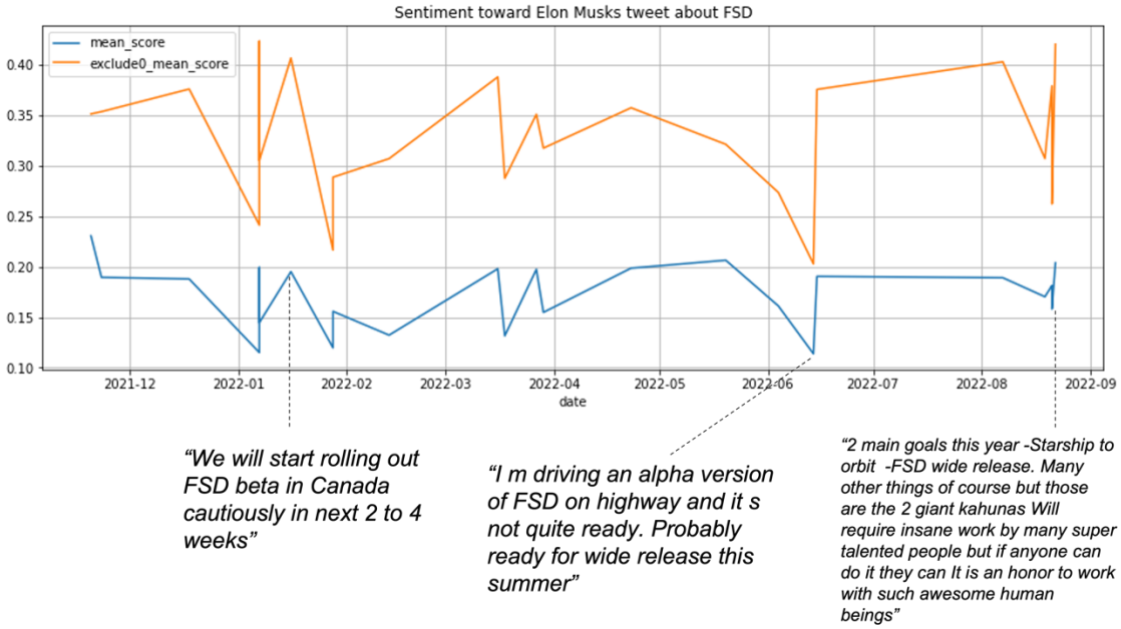


Word cloud map





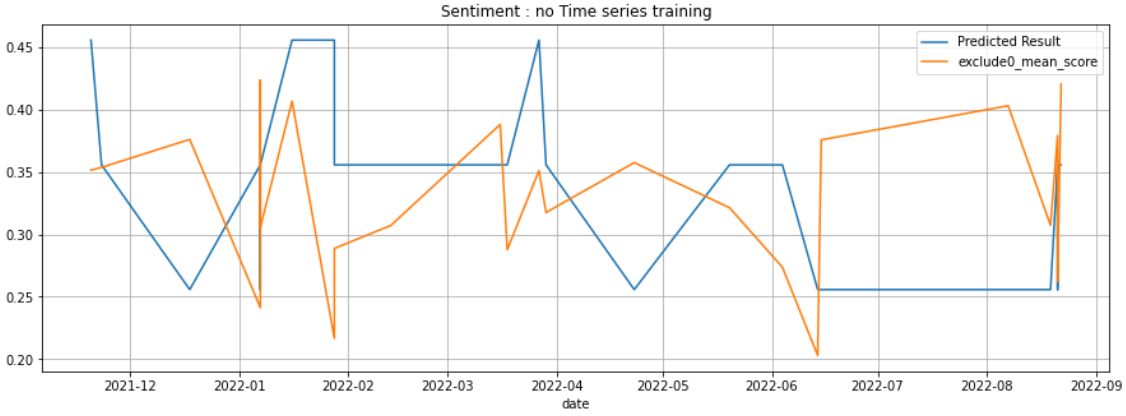
Replies' sentiment matches the tweet's content.



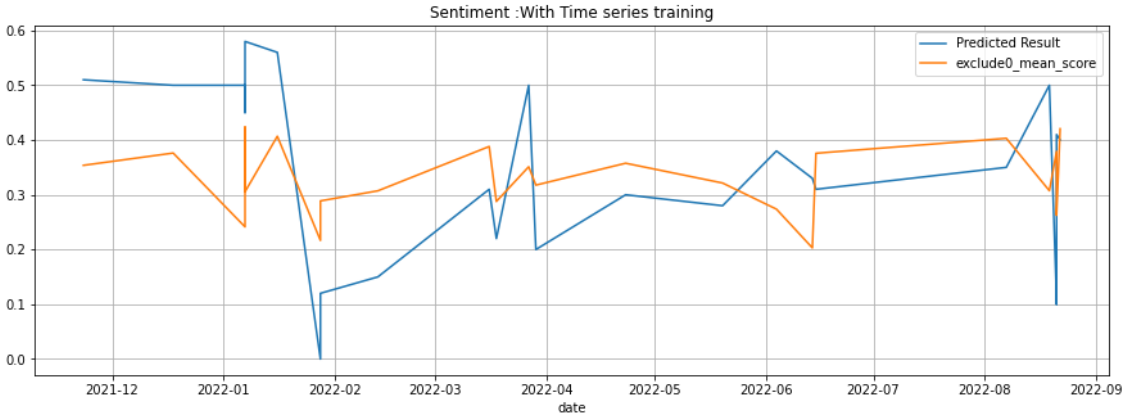
We can observe that people generally react positively to positive content, such as new service launches, and negatively to worrisome content, such as ''not quite ready". Thus, we can use the average reaction of thousand's replies as training labels for tweets' sentiment. Because the pre-trained sentiment model is a categorical model, which is different from the Vader model. We will use the "0 excluded mean score" to have a larger variance when we multiply the score by ten and round it into an integer.

Model result

Model without time series training



A model with time series training



Accuracy improved with the time series training model, as we can compare the two charts above. Due to the data access limitation, our training data size is relatively small. We expect our validation accuracy to improve more as we have a better dataset.