* **Research problem (10pts)**: Describe the task you want to achieve. What is the outcome of interest? What are you trying to predict? Why is it important? What are the potential benefits of having a predictive model for this outcome?

On October 27, 2022, Elon Musk, famed owner of Tesla and SpaceX, acquired twitter for a total price of $44 billion (Korn, 2022). Arguably, most if not all social media in some case is dipped in the paint of chaos. Twitter is no exception and often the center of many opinions on social events. Long before the idea of acquiring Twitter became public knowledge Elon was a known social media troll. One look no further than 2021 when he was quoted as saying “I keep forgetting you’re still alive” to Senator Bernie Sanders after he suggested the wealthy pay their “fair share” (Maruf, 2021). Since his acquisition, the tables have been turned with Elon in the sights. With this somewhat significant change in social perception, his archive of tweets is volatile in nature.

I’m interested to see if I can use a model trained on different tweet data to predict the sentiment of Elon’s tweets. Then looking at how the likes and tweets fluctuate with positive and negative sentiment. A model like such is important as it can be a solution or prodromal to a solution that helps others understand context. Given the natural variability in the human brain, there are plenty of disorders making speech and text language difficult to interpret. While I am interested in Elon’s metrics I’m mostly interested in looking at developing a model that aids in providing positive and negative context.

* **Description of the data (15pts)**: Describe core features of the data, any additional features you produced from existing features and how, basic descriptive statistics about these features, and any missing data analysis you conduct. The description should be sufficiently clear that the instructor understands all the variables included in your modeling.
* **Description of the models (15pts)**: List at least three different modeling approaches you apply to this dataset. Describe each model fit, why the given model was selected, which hyperparameters to be optimized and how, assumptions of the model, and a high-level (think broad audience) description of what the model is doing and why it is appropriate (even as an initial starting point). Also, discuss how you plan to evaluate model performance.
* **Model fit (20pts)**: Provide the results of your model evaluation. Compare and contrasts results from different fits, including a discussion of model performance. Discuss your final model selection and the evidence that led you to this selection. If it is a classification problem, how did you choose a cut-off point for binary predictions? Did you consider different cut-off points?
* **Data visualization (5pts)**: Include at least two plots (or more) to help communicate your findings. The plots may be of initial data explorations, fits of individual models, and plots displaying the performance of competing models.
* **Discussion/Conclusion (25pts)**: Discuss and summarize what you learned. Which variables were the most important in predicting your outcome? Was this expected or surprising? Were different models close in performance, or were there significant gaps in performance from different modeling approaches? Are there practical/applied findings that could help the field of your interest based on your work? If yes, what are they?

References

Korn, J. (2022, May 17). *Elon Musk’s bumpy road to owning Twitter: A timeline | CNN Business*. CNN. https://www.cnn.com/2022/05/17/tech/twitter-elon-musk-timeline/index.html

Maruf, R. (2021, November 14). *“I keep forgetting you’re still alive:” Elon Musk trolls Bernie Sanders on Twitter | CNN Business*. CNN. https://www.cnn.com/2021/11/14/business/elon-musk-bernie-sanders-tweet/index.html