* **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?

As someone with a deep interest in investing, I like to pay attention to the different social factors influencing volatility. Currently, the company twitter is experiencing a significant amount of social turmoil due to its recent takeover. Now CEO Elon Musk has been frequently tweeting irrationally causing some uncertainty in the stability of the company. By understanding the sentiment of his tweets, we can get a good idea for how people feel about him. However, I cannot just ask people to look at all his tweets and tell me whether they are positive or negative in sentiment. That would take far too long and would be far too subjective.

* **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?
* **Reproducibility (10pts)**: Provide a link to the GitHub repo at the beginning of your report as a note.