Deliverables

The scope of the project is broad, please feel free to aggressively adjust it given your time constraints. Please don’t spend a lot of time chasing the last bit of accuracy. What we would like to see:

● All the code you wrote to solve the problem, including the model and feature generation. ● A short document answering the following questions:

○ What features did you consider?

○ What model did you use and why?

○ What was your evaluation metric for this?

○ What features would you like to add to the model in the future if you had more time?

○ What other things would you want to try before deploying this model in production?

**Features Used**

All features except Follow because it had a very low standard deviation, and it did not have much correlation with the class value of the model

**Model Used**

I chose Multilayer Perceptron to do the regression because it performs much better than a simple linear regressor on a limited number of features.The models also scales much better than linear regression over a large amounts of data.

**Evaluation Metric**

I used the correlation coefficient as the evaluation metric

**Other things to be considered**

* Clean the data
* Fill in the missing values
* Check for missing data outliers
* Test the model out with various other algorithms, and probably implement ensemble methods to increase performance.
* Tune with the algorithm with various parameters to optimize performance.

**Implementing the model to production**

* Creating a web service that uses the model to serve predictions
* Streaming the data using streaming service such as Kafka or Kubeflow
* Regularly retraining the model with new data