Proposed modeling approach

Since the value we want to predict is the daily return of certain stock, one of a classical method would be using time series approach.

Here is an outline of the steps we want to take:

1. Naïve model
2. Rolling Average
3. Exponential Smoothing
4. ARIMA/SRIMA
5. SRIMAX with sympathy play as an indicator
6. Long Short Term Memory (LSTM)
   1. LSTM with previous stock prices
   2. LSTM with previous stock prices and sympathy play indicator

The model 1 typically will not offer good estimation or prediction. But this will serve as a fundamental benchmark model. If a model is performing worse than this, then the model choice is very biased.

Model 2 and 3 will increase the model complexity a little bit, and the model 4 will offer the most classical time series approach for stock price forecasting. With proper choices of the parameters, model 4 should offer a better estimation and prediction compared to 2 and 3. (at least not worse) If this is not the case, then we might directly start with model 6. (We might need to consult with our mentor if this happened.)

Model 5 is where we use classical linear model to assess if the sympathy play indicator is a relevant variable. Because model 5 is still linear model, we can check the statistics like p-value to see if it is significant. We can also check its predictability.

1. If it indicates the that the sympathy paly indicator is a significant factor and the predictability is significantly improved, we are done and have a very good result. Sympathy play indicator is an useful variable to forecast stock prices.
2. If the sympathy play has statistical significance but no improvement on predictability. We move on to model 6. We would conclude that the sympathy play indicator is a relevant factor but there might be some more complicated relationships underlying.
3. If the sympathy play indicator has no statistical significance and no improvement on predictability. We move on to model 6.

Model 6 is a nonlinear model and is hard to give rigor inference. Model 6 should outperform the previous 5 models in predictability. If 6(b) improve 6(a) in predictability, then we can conclude that the sympathy play is a potential valid investment strategy, despite we do not fully understand its relationship with daily return. If 6(a) and 6(b) offer the same predictability, or 6(b) is even worse, then we conclude that for daily return, our sympathy indicator is not useful for predication. This can cause by three reasons:

1. Our sympathy play indicator is not properly defined.
2. Daily data is too coarse. We need very fine data like 5-minute data or shorter.
3. Sympathy play is not really a valid strategy for our selected sectors.