I decided to approach this hackathon with more focus on feature engineering than on model selection and data processing. After reading the problem, I decided to use gradient boosting with binary logistics.

I always submit a preliminary model, generally with all the variables, to set a benchmark score.

There were 4 moving averages in the data set and I expected them to be correlated. So, I plotted correlation matrix and as expected 10 days and 20 days moving average were highly correlated with other moving averages. I removed these two variables and trained my model on rest of the data. This model was giving a 0.68 (approx.) score on public leaderboard.

I checked for null values and there were 4000+ rows which had missing values. I left it as it because it was very small percentage of the train data set. (Wanted to come back to it, didn’t get time)

After this I started creating features. Features which improved my score were (1,0,-1 values) :- comparison of 3 days moving average with other moving averages, comparison of 5 days moving average with other moving averages and sum of these comparison value. I created this to use price movement direction based on moving averages. After creating this, my model was giving a score of 0.677(approx.) on public leaderboard.

I think that hardest part in any mini-hackathon is to create features. It takes some thinking and not every feature you create will add values. But, it is important to keep on doing it even if first few features are not able to improve your model.