Are there any changes to your problem description or the approach that you are taking?

Getting the data in the correct format was more difficult than we expected, and took us more time than we expected. Moreover our classifiers (see below) did not behave as we expected them to behave (think of only predicting one label for the complete dataset). Therefore we will focus on SVM and see if we can optimize our features in such way that the classifiers can actually use them. We hope this will lead to better predictions (for now SVM was our best classifier with a accuracy on the test set of 56%).

Also, normalizing the data is an issue because of the non-linear properties of the data. Normalizing of min-max approaches skew the data and do not favor the outcomes of the classifiers. We will try to find a way to pre-process the data such that it does not affect the correlation, but only scales the features.

Which classification/clustering/regression algorithms are you going to use?

We tried quite a lot of classifiers, most of the ones from sklearn, such as: Logistic regression, Neural networks, decision trees and random forest and SVM. Most of the accuracies on the test set were around 40%. For example Neural networks scored 60 percent but after investigation of the confusion matrix and the classification report, we found that this classifier only predicted increase, which was apparently the label of 60 percent of the test set.

However, SVM scored 56%, with a better distribution of predictions, but a seemingly still biased prediction towards increase.

How will you evaluate/compare them?

For now we use the accuracy on the test set and a cross-validation score. To study the predictions, we use the classification report, which includes the recall, precision and F-score, and the confusion matrix to find possible errors as mentioned above.

What is your progress till now? Describe shortly, no need for figures or code.

We understand our dataset, and made a selection of features. Also we created some new features ourselves (moving averages, ratio between moving averages and the change in price per day). However, the bad performance of the classifiers is still a big issue and makes us rethink the selection of features. We possibly want to add more features, since our current data has only 3 features, but these features are complex features, constructed from data that is left out in our feature selection right now.
For now we also used the standard classifiers, but we are planning to optimize the ones that perform best with their default settings.