Springboard Data Science Career Track
Jesse Mailhot
Mentored by Tony Paek

Capstone Project 2
Milestone Report 2
Predicting Microsoft Stock Price Movement With LSTM

Updates

1. Data has been prepared for the models
2. Models have been compiled

Preprocessing

LSTM

The first step in preprocessing is to transform the data into a supervised learning problem. To do this, each row of data will contain the previous 60 days and the future 60 days from each data point. Next, the data will be split into randomized testing and training sets to avoid any bias from time series. This is a very important step; without it, the model would be biased towards the historical value of the stock, and would not be adverse to future changes in the stock price differing from the historical trend. After the data has been split into testing and training sets, each set is then reshaped into the LSTM format: actual data, number of periods (steps), and number of features.

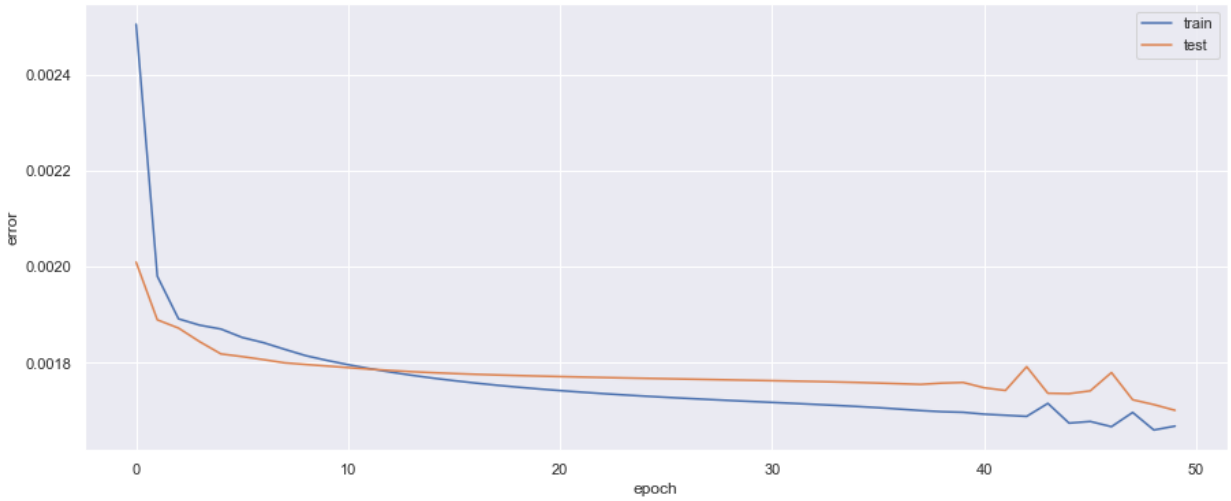
Prophet

For the prophet model, the data preparation is much simpler. Prophet takes as input a Pandas dataframe with 2 columns: ‘ds’ (datestamp) and ‘y’ (historical data). To get comparable results for the 2 models, the prophet model will train on 1 year of data, then make a prediction for the following year. Each iteration will call a new instance of prophet in order to avoid historical bias.

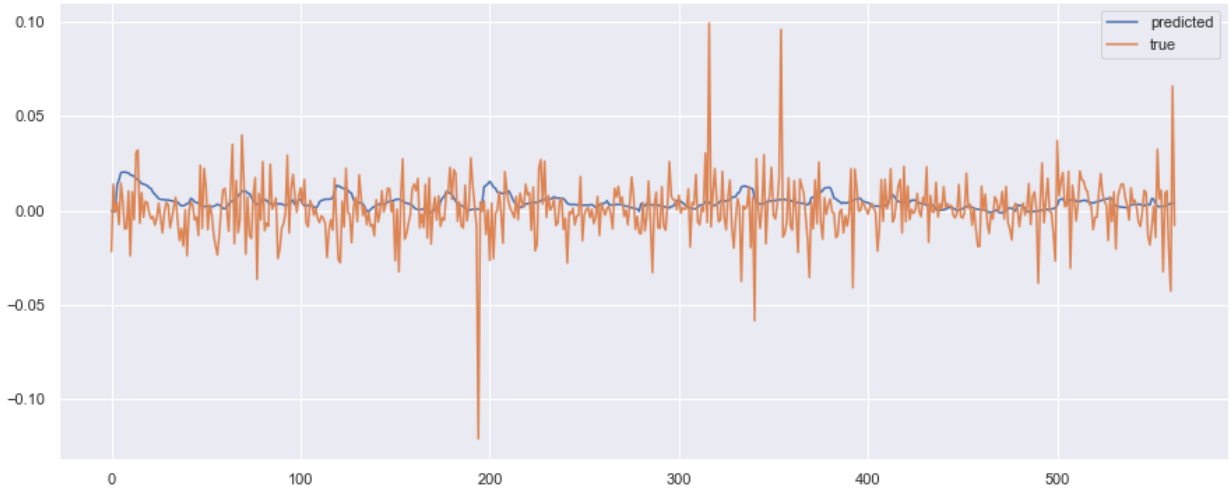
Model Analysis

LSTM

The LSTM model for this project will have four layers: three 50 node LSTM layers and one 4 node dense layer for output. The will use mean squared error as the loss function with the ADAM optimizer. The model will be trained over 50 epochs with a batch size of 72. Below is a plot of the training and validation loss.



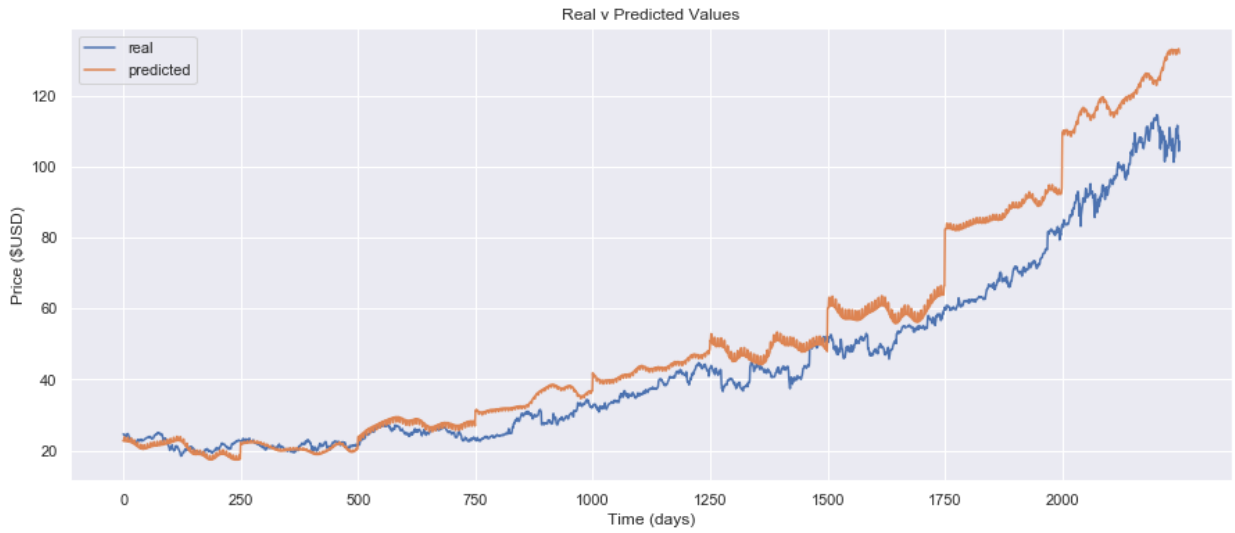
After compiling and training the model, it is run on the test set. The results for daily returns are visualized below.



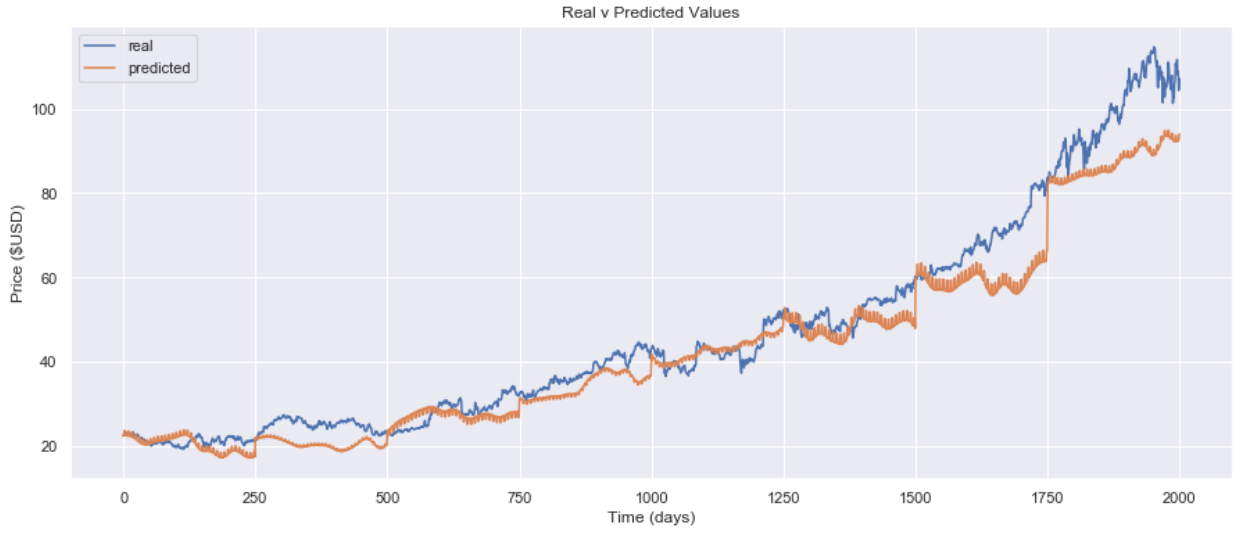
The LSTM model correctly predicted the direction of the daily stock price movement 53% of the time; whereas it correctly predicted the direction of the quarterly stock price movement 74% of the time.

Prophet

The prophet model will be the standard out-of-the-box model, using the yearly seasonality. Running the model in a for loop for each year produced the following results.



From the plot it is interesting to note that as time progresses, the model starts making large jumps from the previous endpoint, especially at 1750 and 2000. Running each model separately (not in a for loop) produces significantly different results, as shown below.



There is a clear difference in the way the model reacts when run in a loop compared to being run individually. In a production setting, this model would likely only be run on one year at a time, so it is not necessarily a problem.

Model Comparison

The prophet model correctly predicted the direction of the stock price movement for each of the 7 years it predicted on. This should be taken with a grain of salt; however, due to the biased nature of training a regression model on historical stock data. Given that the underlying goal of the stock market is to create value and increase prices, simply predicting that the stock price would increase every year would yield a correct prediction 6 out of 7 times.

The LSTM model is designed to be averse to this kind of bias, and therefore I assert that it is the better of these two models for the purpose of predicting stock price movement. The LSTM model performed better on longer term data, 74% correct quarterly direction prediction compared to only 53% correct daily direction prediction.