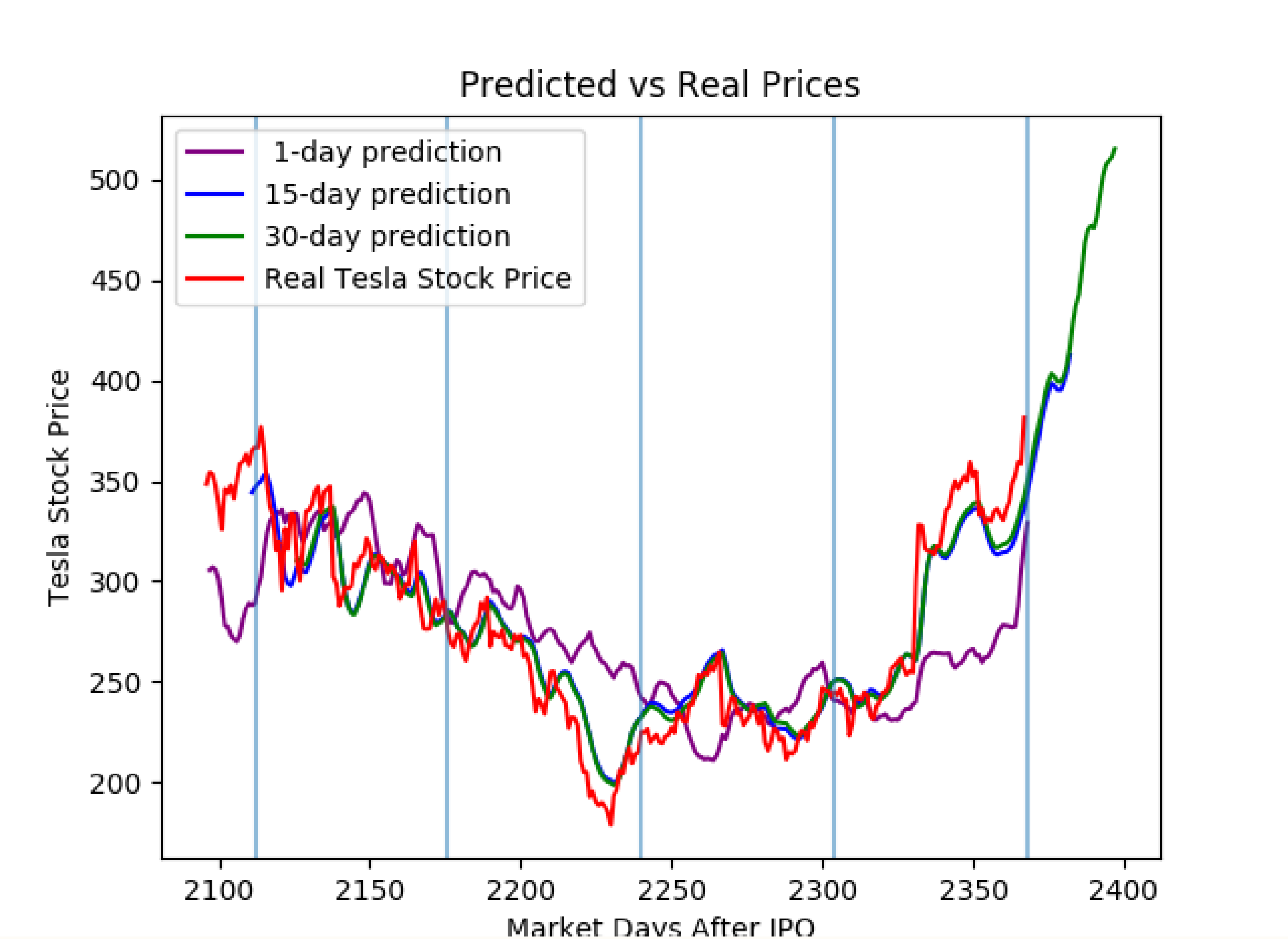
Stock Price Prediction

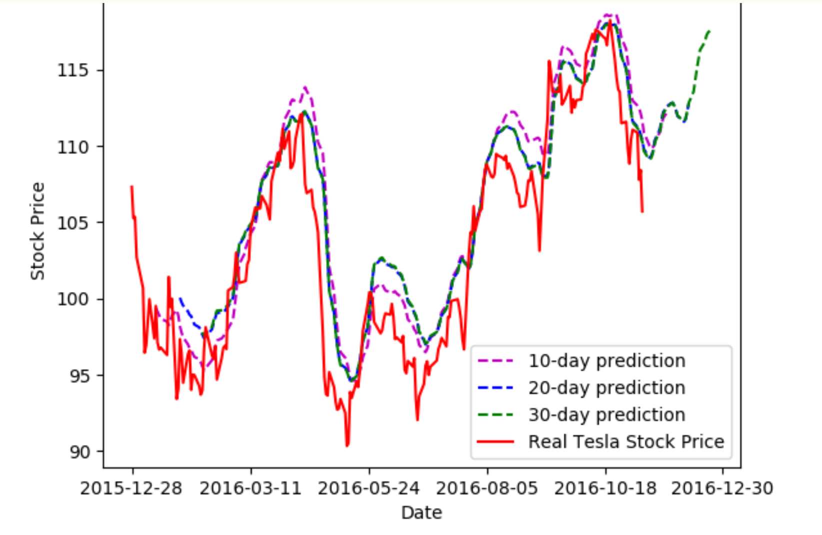
Using Historical Data

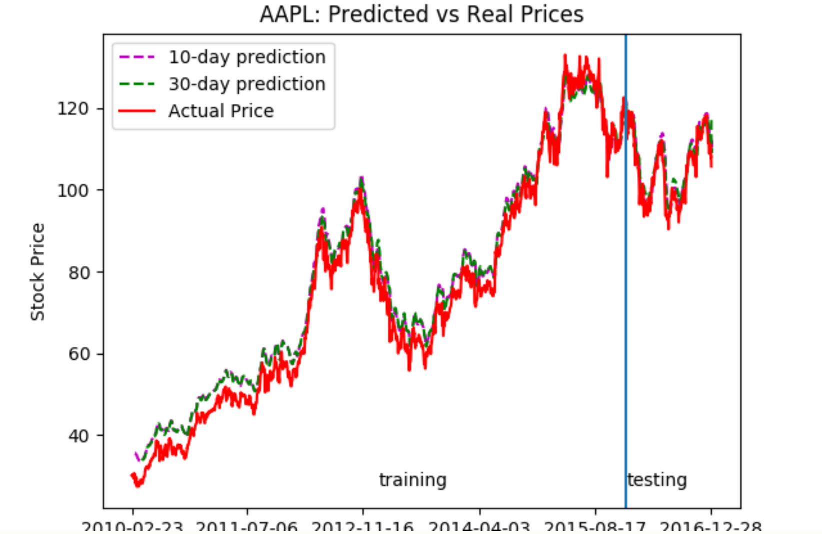
By Christopher Pearson



Earlier Models

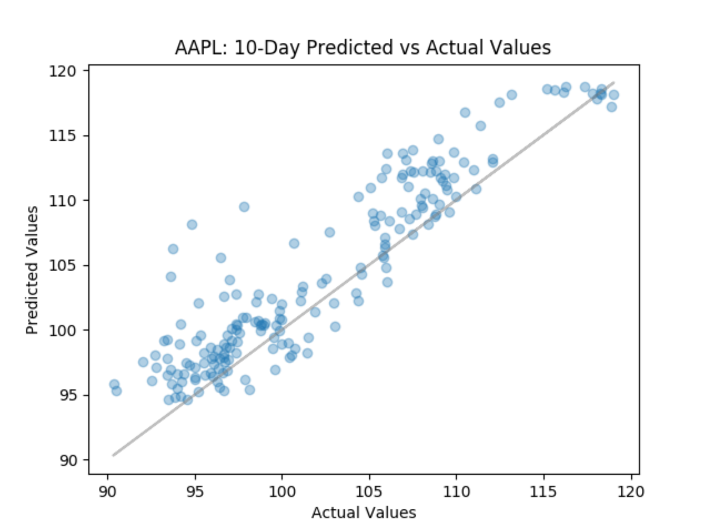
- Earlier we tried predicting Tesla's stock price using machine learning.
- We had some success for predicting the closing price 15 and 30-days in the future
- This time, we are going to see how this model predicts with the closing price for other stocks

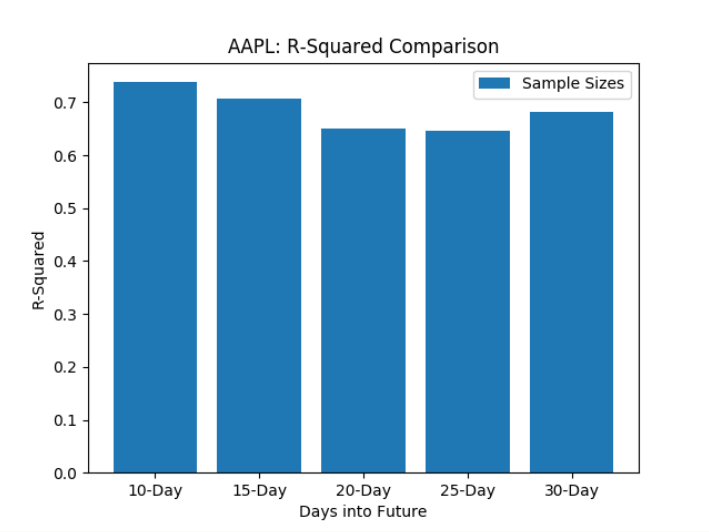




First Stock: Apple

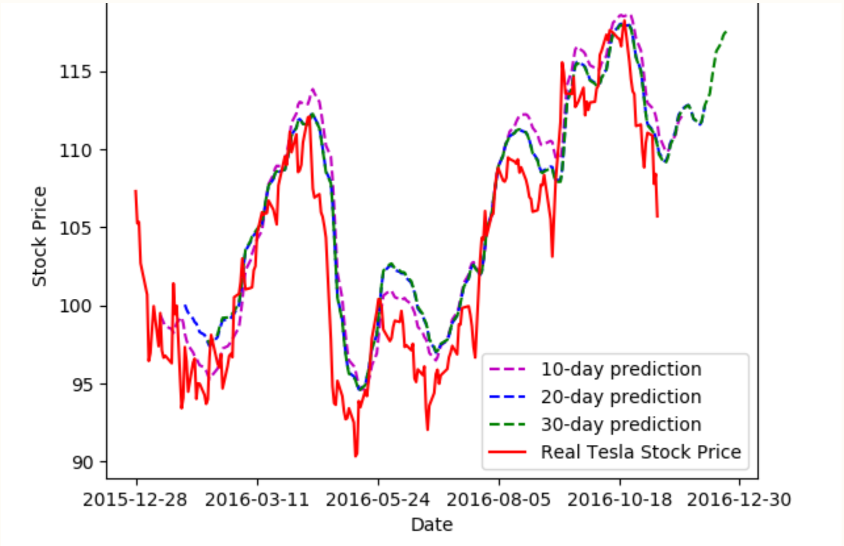
- Predictions match training data fairly well
- Testing data seems to do well, but some looseness





Accuracy

- R-Squared between .60 and .70
- Apple stock seems more volatile during test period
- Residuals indicate y-hat a little too high in test data

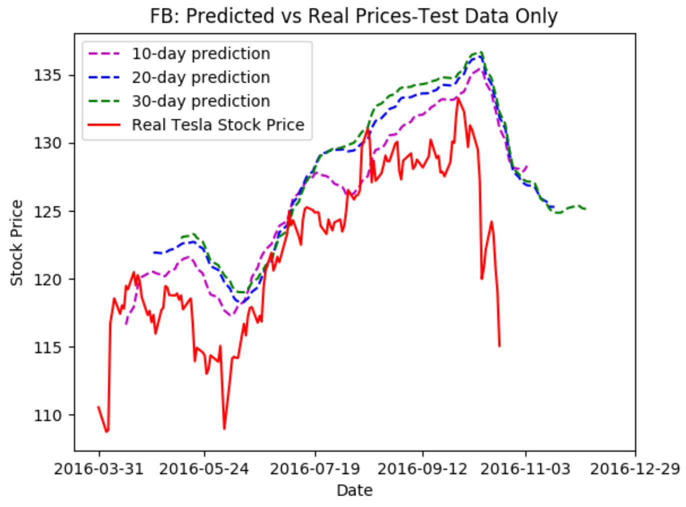


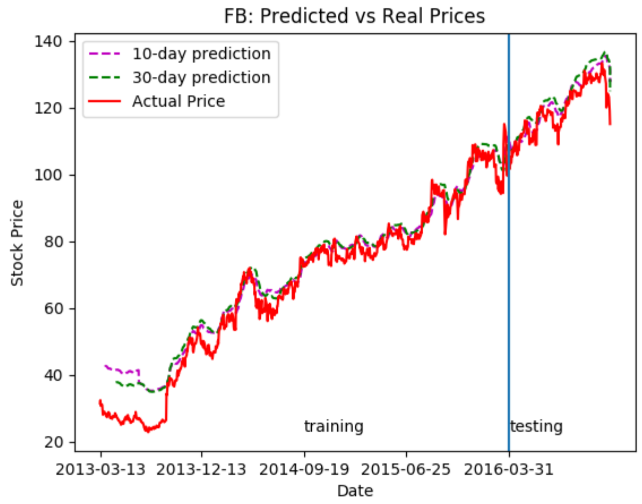
Does it pan out?

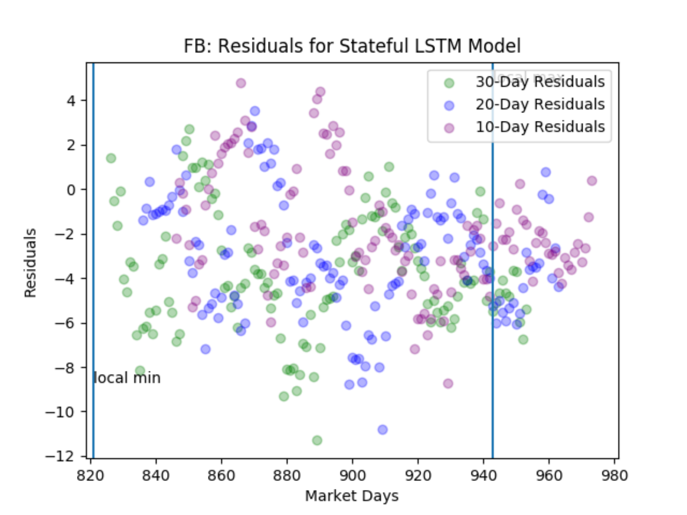
- Model predicts a December recovery after the October Slide
- There was a December Recovery

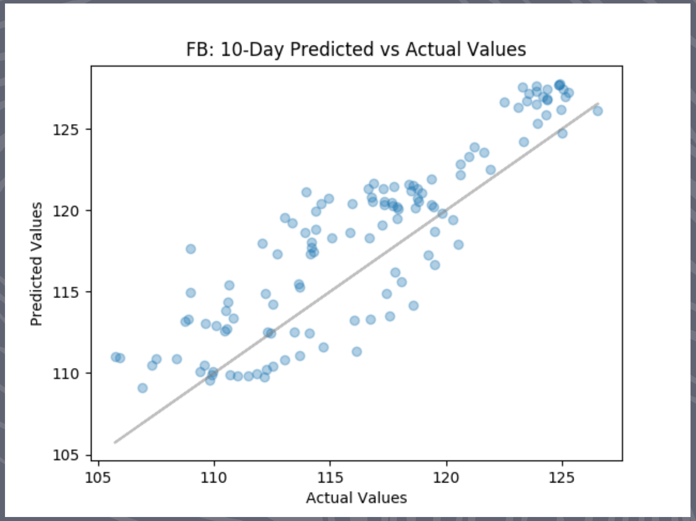
Second Stock: Facebook

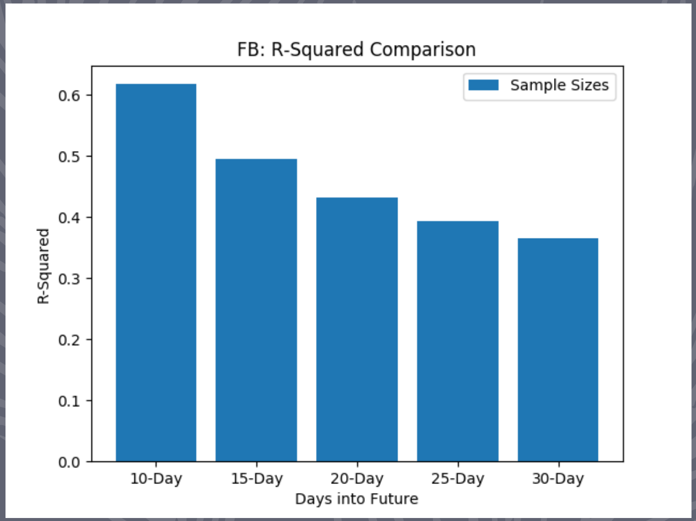
- Looks good for training data
- Model struggles with test data
- What is going wrong?





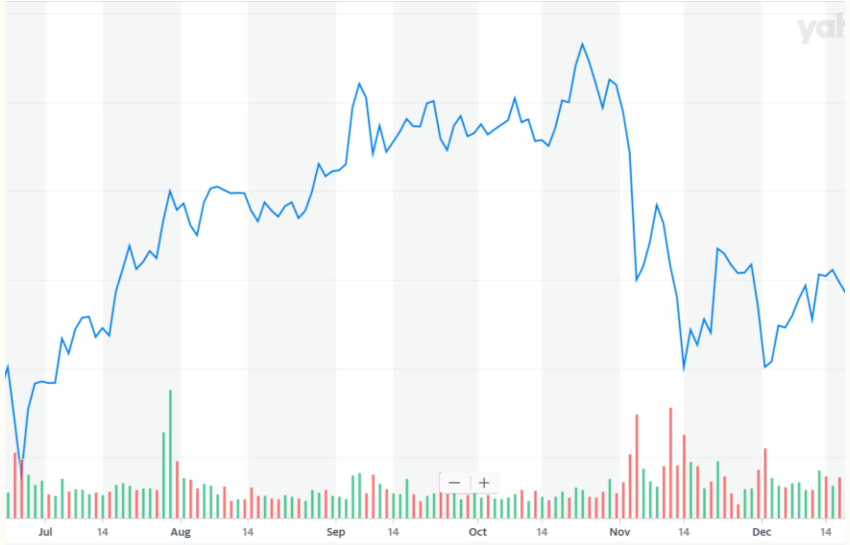


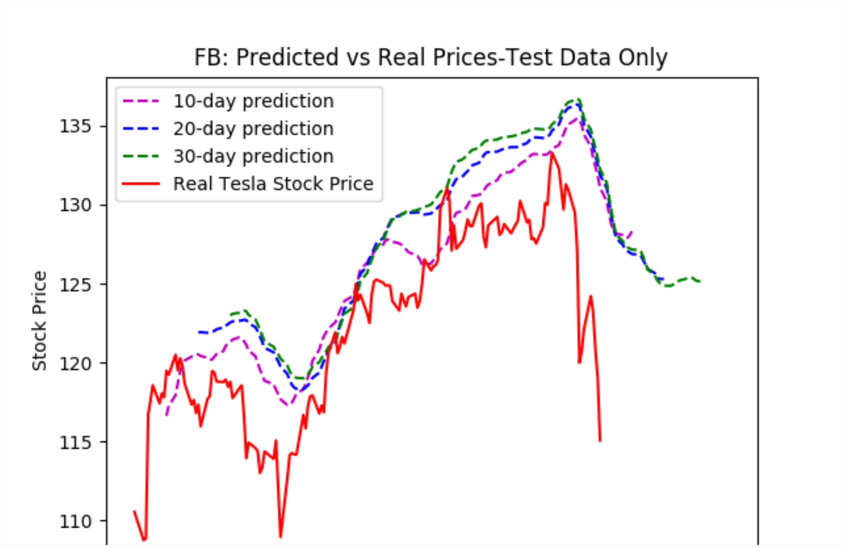




What do Residuals tell us?

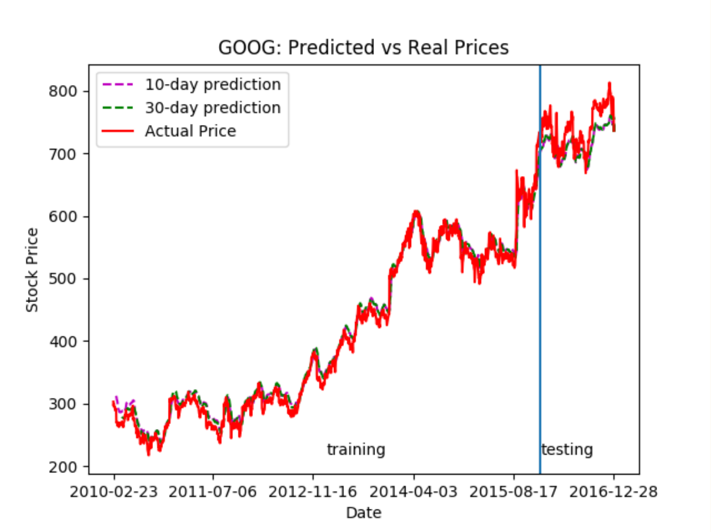
- No clear pattern for predicted vs actual values
- Residuals vs Time shows no clear pattern
- R-Squared best for 10-Day Predictions
- R-Squared ranges between 0.35 and 0.60
- Worst for 30-Day Predictions

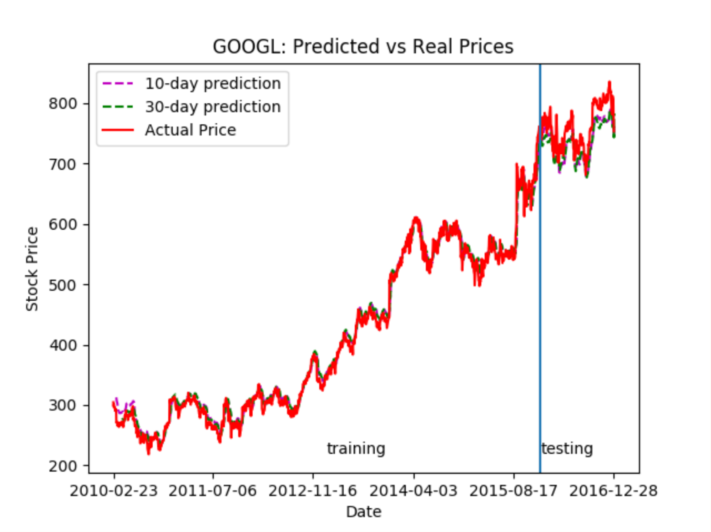




How's its predictions?

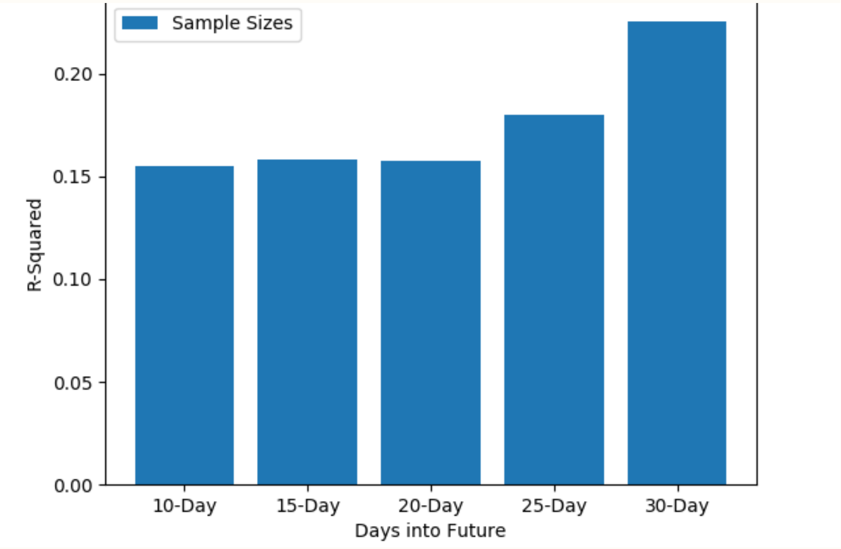
- Model does not think much of October slide from 130 to 115, sticks to 125
- Stock does recover in January
- But…failure to predict major dip not a mark of success

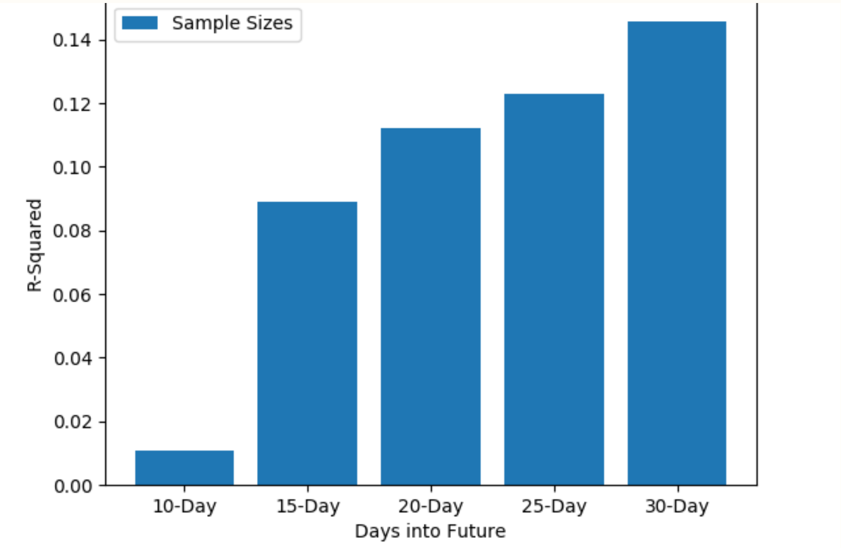




3rd and 4th: Google A and Google C

- Google A (GOOGL) shares come with votes
- Google C (GOOG) do not
- Same Company, similar price
- We’re using their machine learning code (Tensorflow)

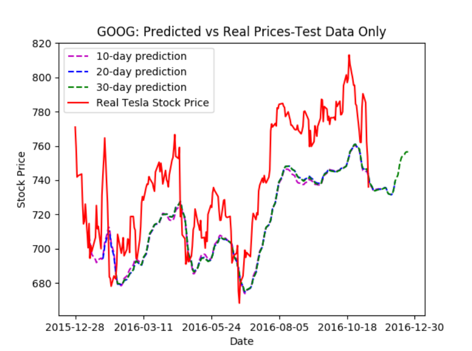


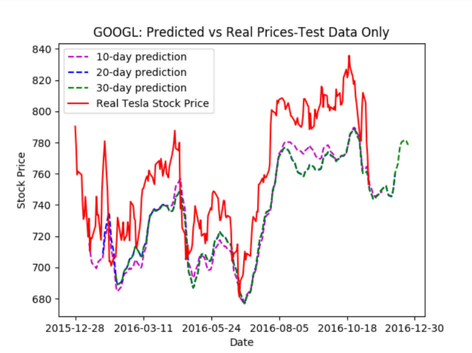


R^2 Look Different

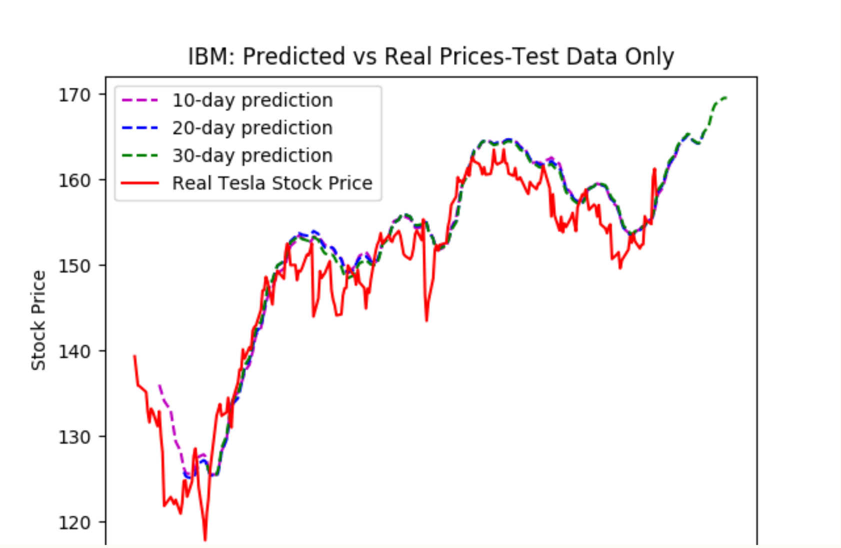
– 30-Day Predictions strongest
– Look different, but it’s the narcissism of small differences
– Both are near-zero

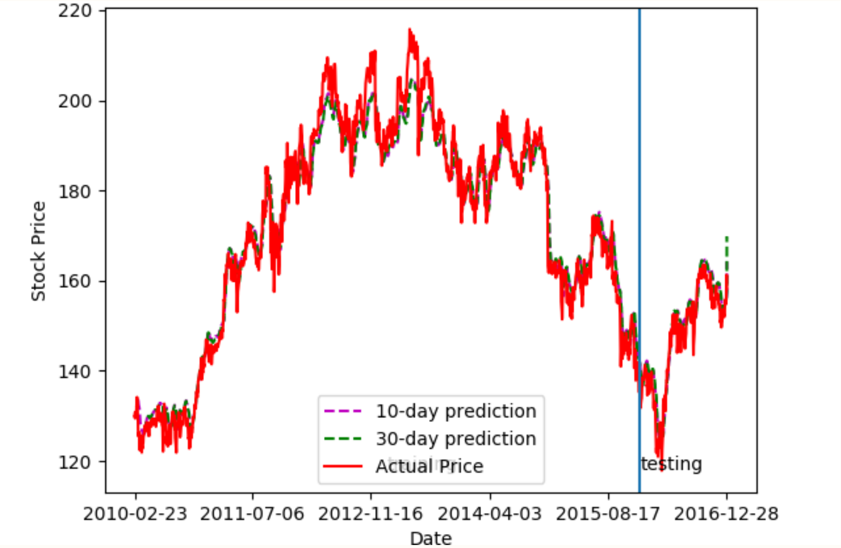
So what happened?





- Google stock spiked and stayed volatile
- Model stayed a step behind
- Y-Hat consistently below actual price

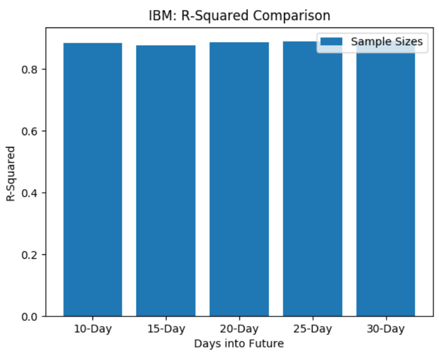


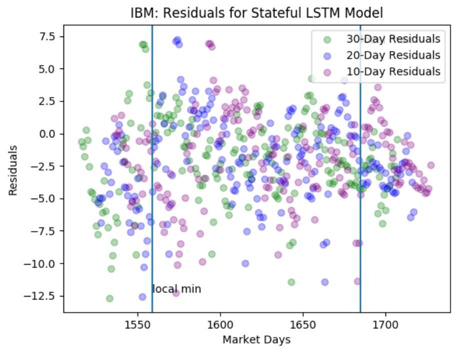


Fifth: IBM

Perhaps an older technology company is more predictable

Predictions seem to track well with actual data for both training and testing sets

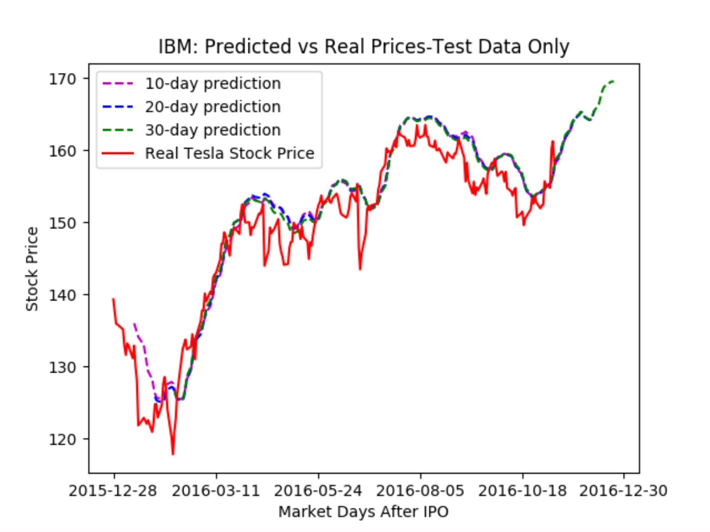




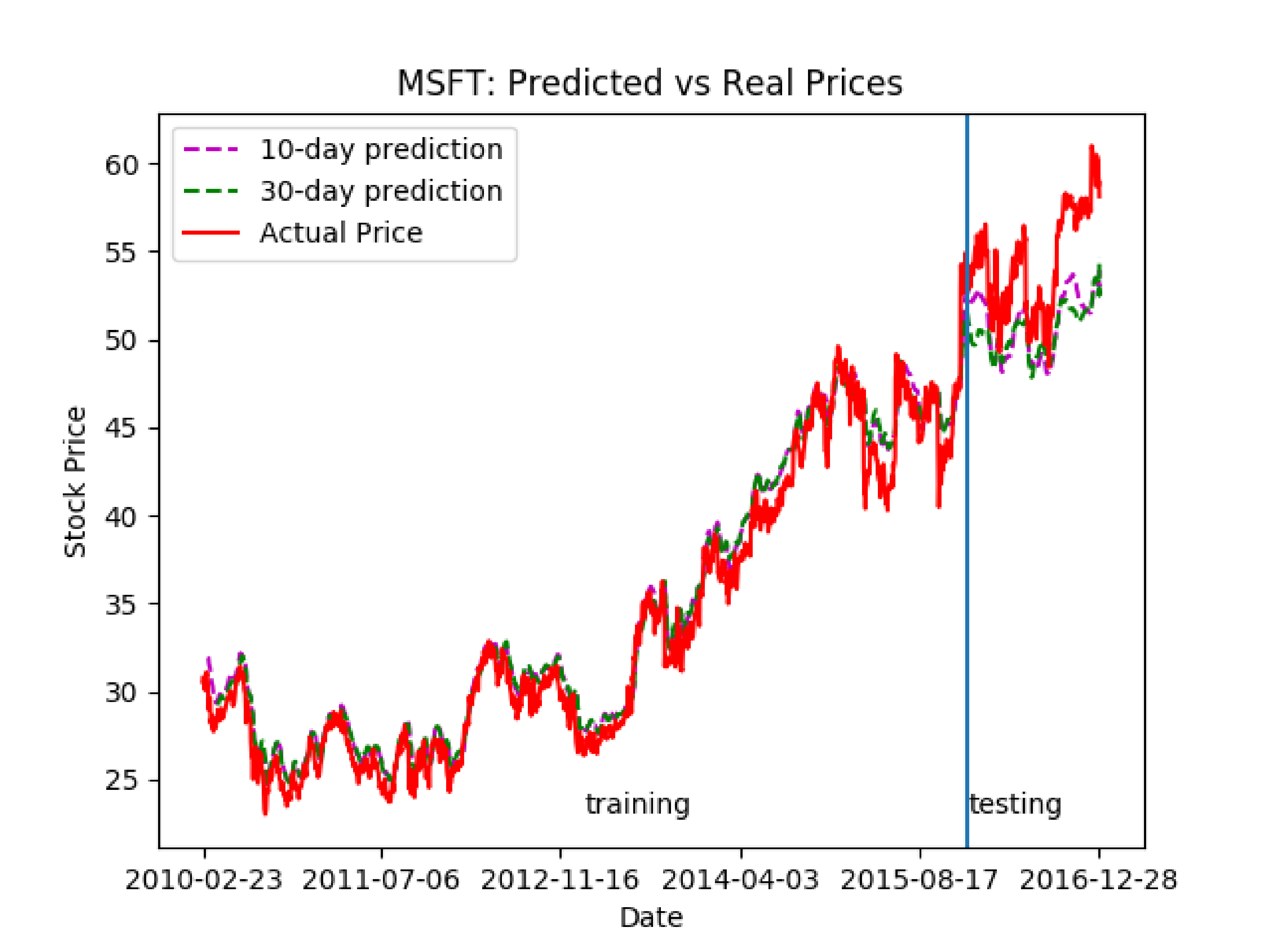
Residuals and R-Squared

- Residuals consistent around 0.85
- No clear pattern to residuals (though this is not regression, so patterns harder to see)

How did predictions go?

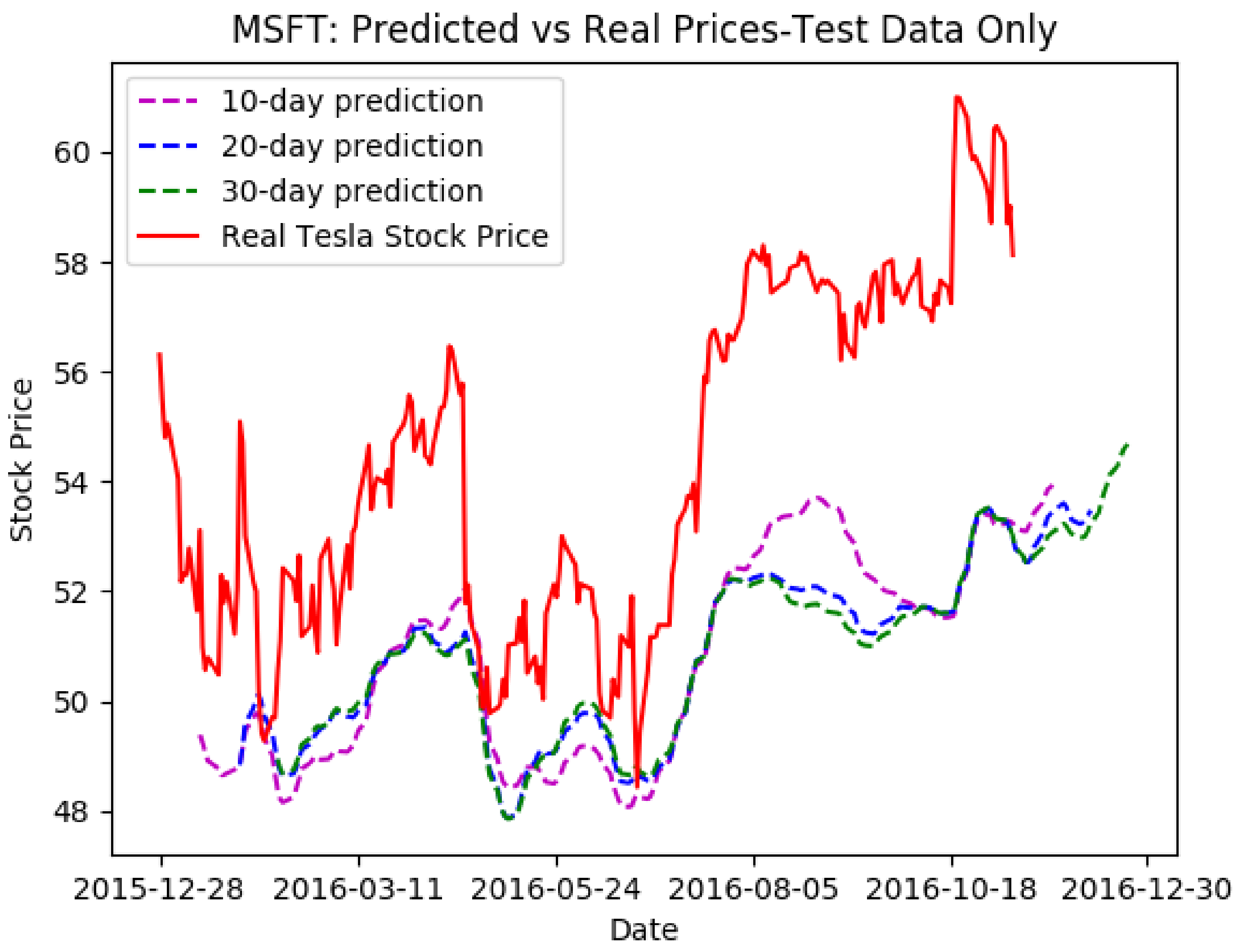


- Bullish on IBM—Rise to 170
- IBM stock rises as predicted
- Initially not as high as predictions
- But by February, stock peaks at 169.53
- Model seems to be doing well



Sixth: Microsoft

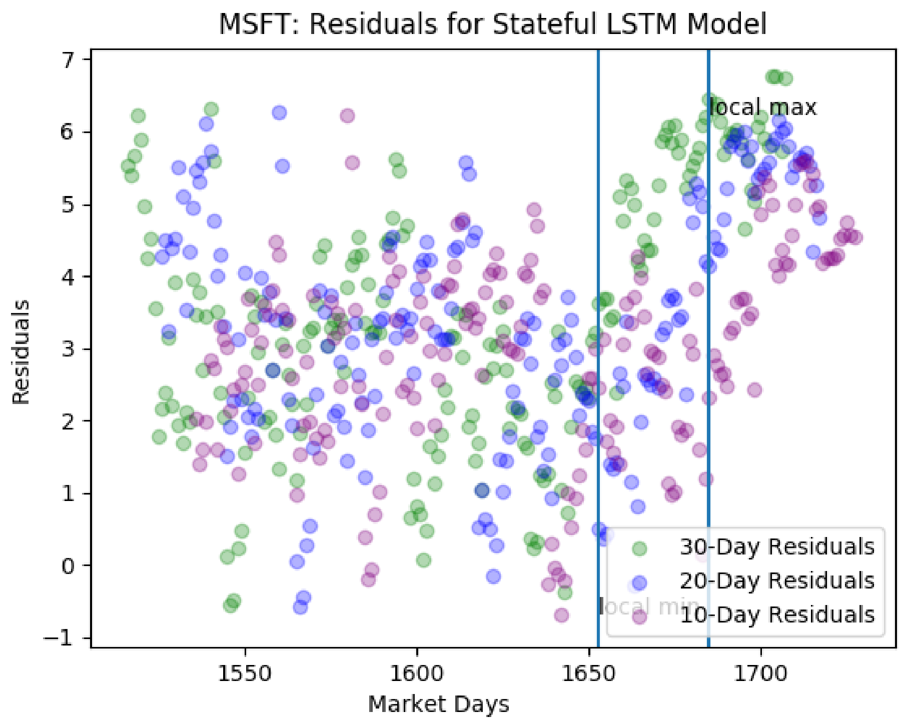
- Looks good for training data
- But seems to be a problem when we shift to testing data



Zoomed In

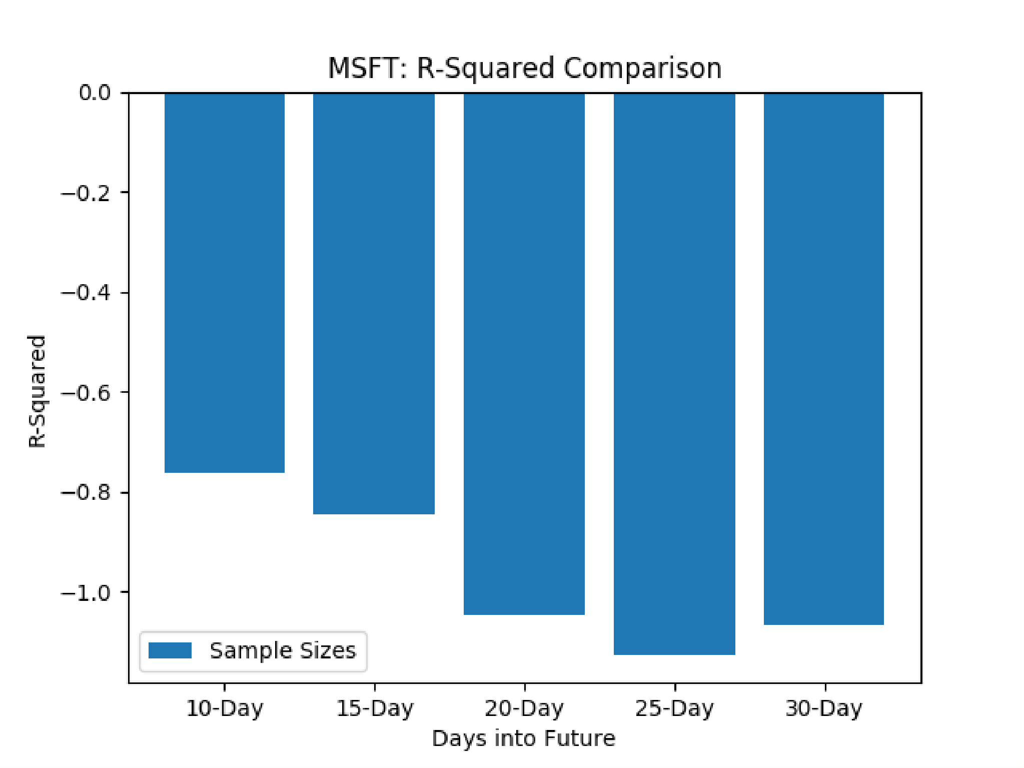
- Looks worse close up
- Seems to be having trouble with that price spike that occurred at the end of the training data
- The shape of the data may be OK if it were adjusted six points up

Residual Plot



- Residuals range -1 to 7
- Shifting the graph up could improve the graph somewhat
- Residuals are worst toward the end.

At least the model knew it was failing



Summary

- Some stocks more predictable with machine learning than others
- We can tell if model is working well or not
  - Only trust well-performing models
- If we calculate on many stocks, should be able to find a few good investments
- Room for improvement
  - Only using daily closing price data
  - Plenty of room for adding more data
- Do sudden price changes at junction between training and testing data lead to model failure?