# Introduction to Data Analysis

## Predicting Stock Prices

## Final Project

5/10/2019

### Jarred Cutlip

Nick Esposito

Jon Mierzwa

The stock market is a crazy unpredictable part of our everyday life. It has one of the

three largest effects on the economy. Everyday millions of dollars are traded in the stock market

accounting for either a gain or a loss in profit. With such a large part of our economy being

trusted in a way that seems to be unpredictable we wonder what safety measures the country or

companies should take to ensure maximum profit. A lot of companies make algorithms to try

and predict the stock market. These predictions are not 100 percent accurate however, these

algorithms are getting better everyday. Collecting all this data is useful, for the more you have

the more accurate your result can be. It is important to remember that even with all this data,

sometimes the market can still stray off path. Stocks are very unpredictable but using algorithms

we believe computer scientists can predict the future of the stock.

What we did in the first part of our project was make a stock predictor for aggressive

small cap stocks. How it works is a couple of simple strategies. Before anything, we needed a

function in our code to be able to open up google chrome, navigate to the correct yahoo page,

and then gather the data from the amount of stocks that the user inputs. Next, we pulled 2 years

of historical data from the chosen stocks and put them into separate CSV files. The 2 years of

data consisted of the variables date, open, high, low, close, and volume. These are very basic

variables in the stock market, but we thought it was a good idea to start simple and then judging

by our predictions, adjust the variables from there.

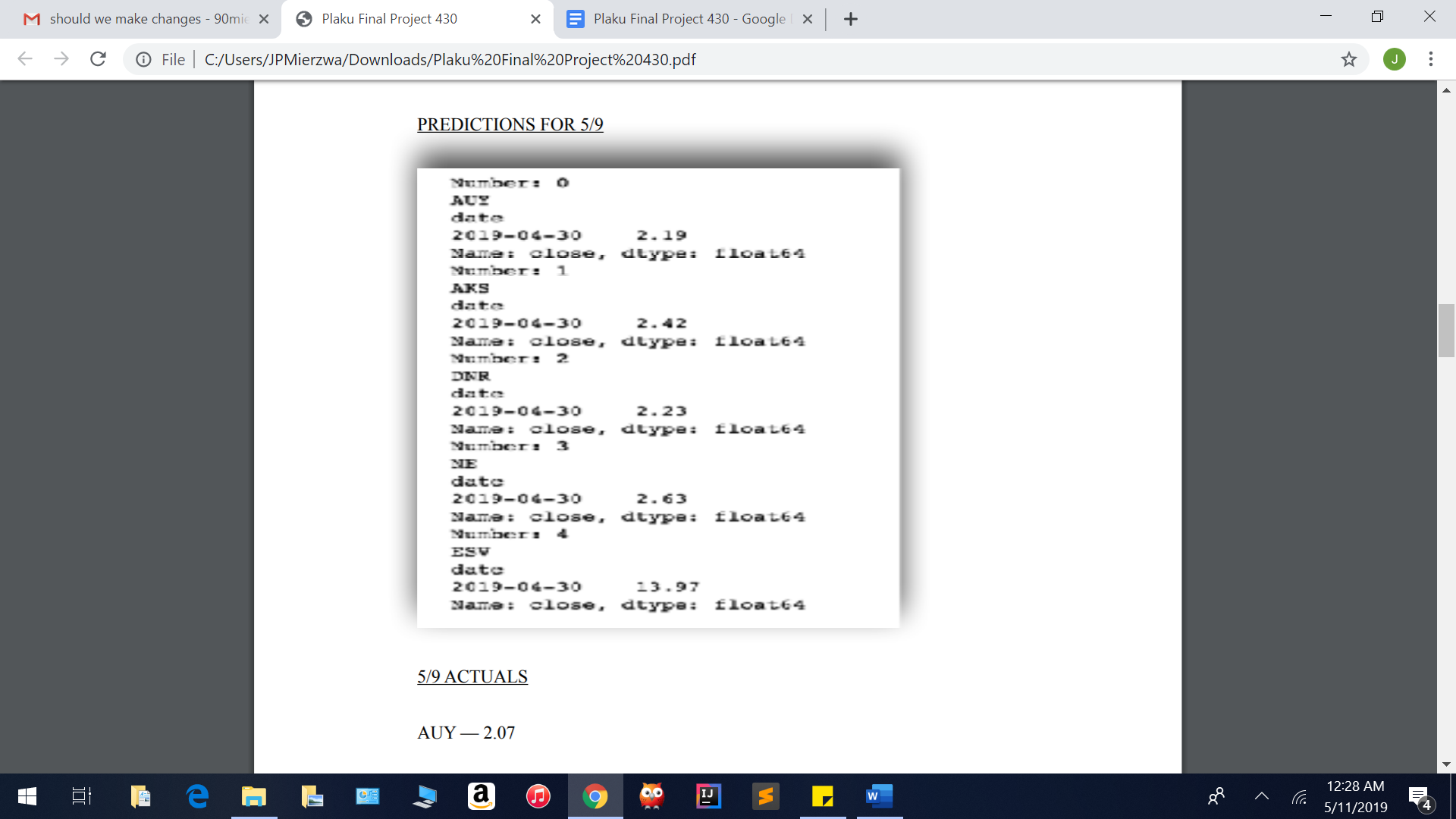
For example, if you run the program on Sunday night and choose a 5 day prediction, the

results would be for when the market closes on Friday. Here are our predictions for 5 aggressive

small cap stocks. We ran the program on 5/2and our results were gathered on 5/9. Lets see how

we did.

#### PREDICTIONS FOR 5/9



5/9 ACTUALS

AUY — 2.07

AKS — 2.45

DNR — 1.90

NE — 2.62

ESV — 12.14

As you can see, our predictions did ok for a very basic model. We plan to further develop our model as we learn more about the stock market, and the possible variables that affect stocks price. We also plan to use a more accurate predicting algorithm in the future. For the second part of our project we wanted to take a look at predicting prices for big name stocks, in our case we chose Apple.

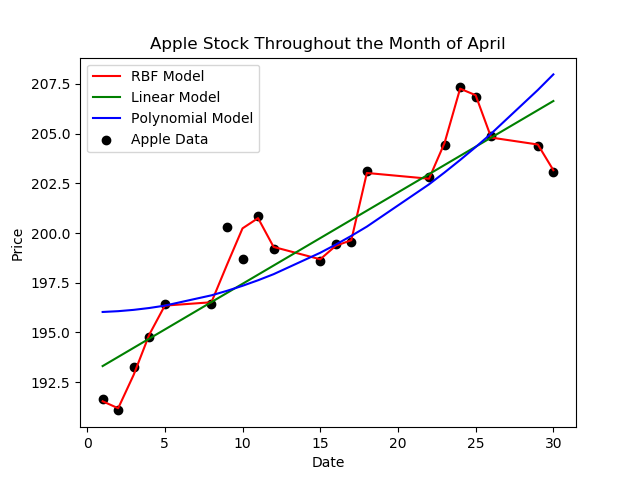
Founded by Steve Jobs back in 1976, Apple has been at the forefront of technical innovation throughout the past couple of decades. Let’s take a deeper look into the financial trends of arguably the most powerful tech company in the world – Apple. The figure below shows a brief snapshot of the historical data regarding Apple’s stock. These data points were taken from April 1st to April 30th of this year.

A screenshot of a cell phone

Description automatically generated

This trend in the data clearly depicts how Apple’s stock did better as the month of April went on. From starting out at $191.64 on the first day of April, it was able to climb all the way up to $203.06 on April 30th. However, this was not the highest number the stock was able to reach in April of 2019. Apple accomplished this feat on April 24th when it rose to $207.36. With many factors affecting the rise and fall of a company’s stock, trying to predict when to invest or take out money can be quite difficult.

With this month of historical data, we applied three different forms of data modeling, linear regression, polynomial regression, and Radial Basis Function (RBF). Alongside simple vector regression (SVR), we were able to observe which type of predictive model performed the best in its goal of closely mimicking the real data. The figure below represents all three of the data models we created in comparison to the real data from Apple’s finances. It is evident in the representation of the data that the RBF model was superior to the linear and polynomial regressions.



The RBF appears to almost perfectly mimic the real Apple stock data, except for a couple data points around the April 9th and 10th. Linear regression has an unfair advantage in this situation, since no stock on the market will ever trend in just a linear direction. Even with this shortcoming, the linear model achieved the second highest accuracy rate of the three. It split the model evenly by overestimating 8 data points and only underestimating 9 of them. The polynomial regression model appeared to do the worst of the three, generating the least accurate parabola to represent the data. The polynomial regression places its first data point at $196.08, which results in an error of $4.44, the largest of the three models. The last data point of the month created by the polynomial regression curve has the highest error, as well.