Christine Saw

ECE 4271 – A

MATLAB Project #2

Linear Prediction with Real Data

March 7, 2021

Part I

(a)

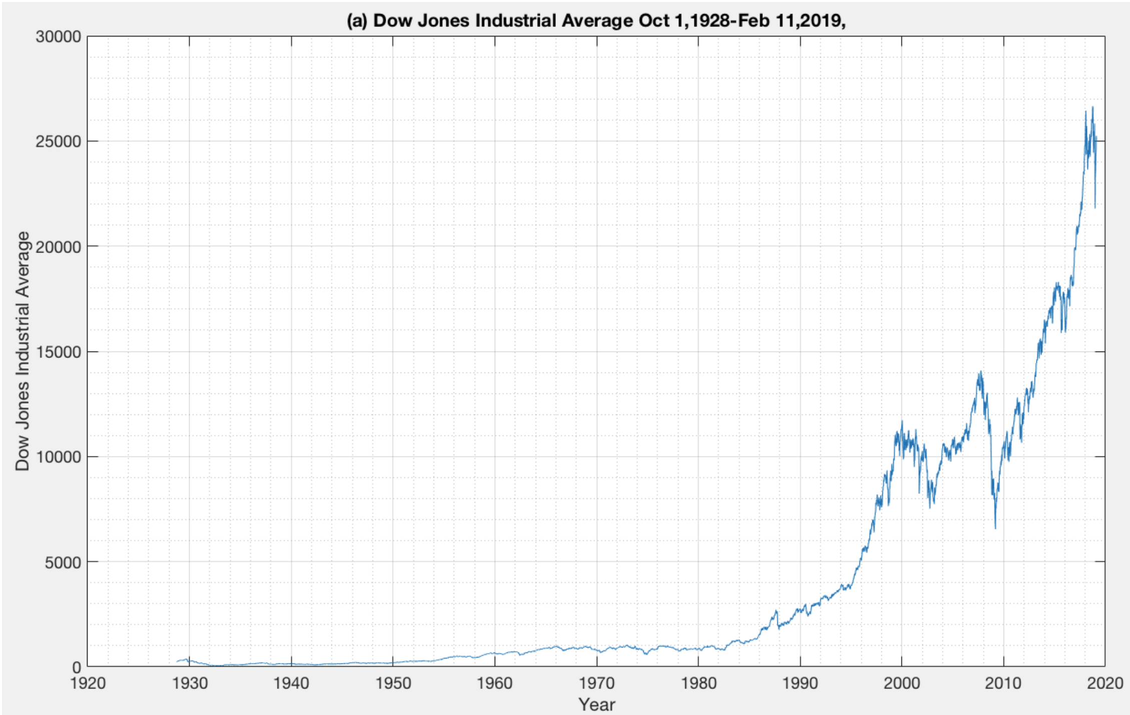

Figure 1. Plot of Dow Jones Industrial Average on a linear scale.

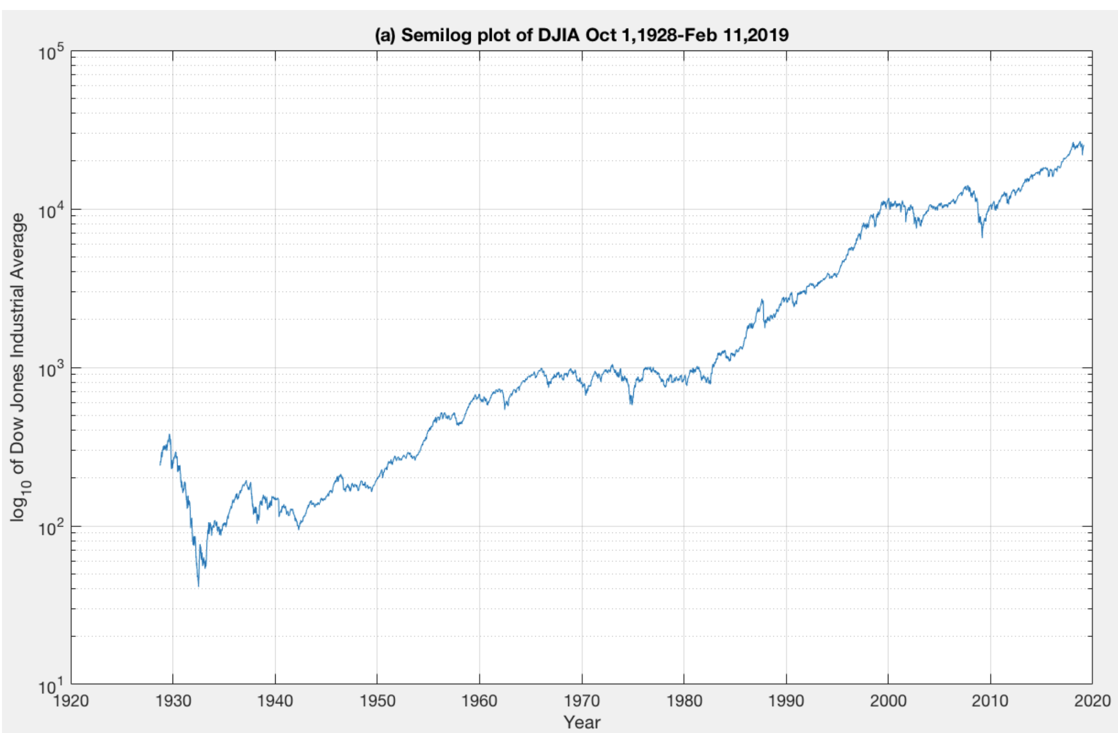

Figure 2. Plot of Dow Jones Industrial Average on a semi-logarithmic scale.

Assuming we started with $1000 and invested all of our money in the DJIA, we would have **$104,196.93** at the end of the investment interval.

An annual percentage rate (APR) of **5.128%** would be needed to achieve the same level of performance if we put all of the money in a savings account for the same duration. It will also take **8056 weeks** to reach the same level of performance with a fixed APR of 3%.

(b) Given p = 3, N = 520, the linear predictor coefficients obtained from solving the matrix is:

$a_1 = 0.0268$

$a_2 = 0.0938$

$a_3 = -1.1183$

(c)

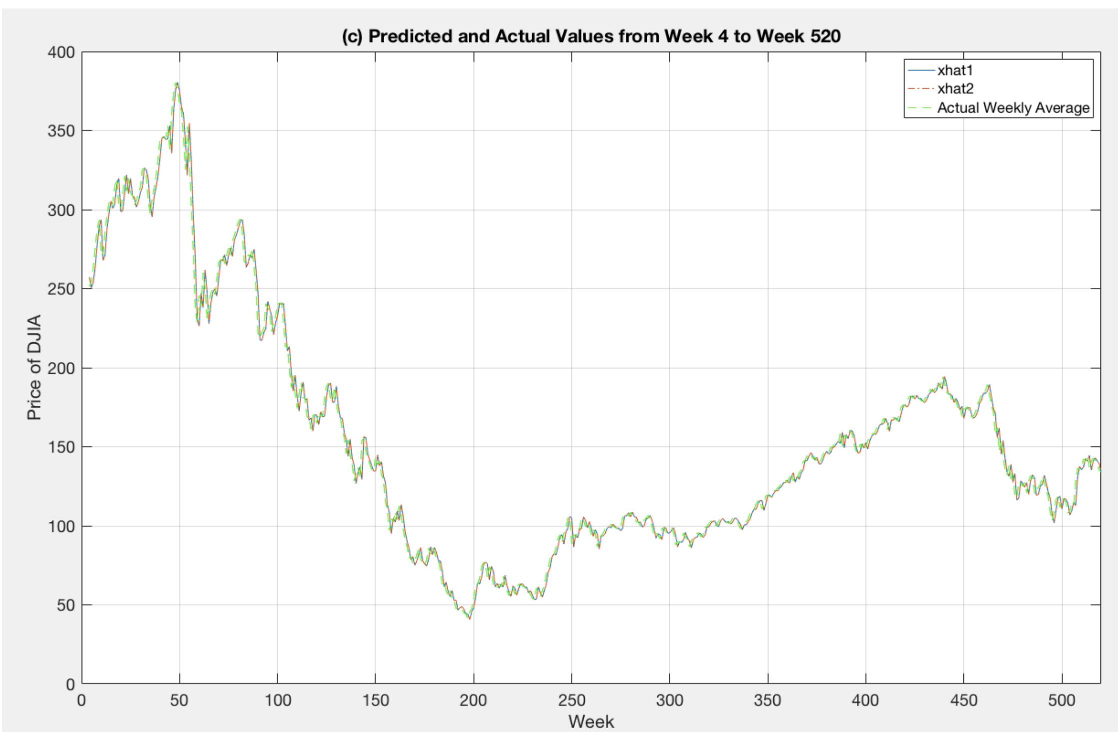

Figure 3. Plot of predicted and actual weekly values of DJIA data from week 4 to week 520.

The total squared error for xhat1 = 23638.1

The total squared error for xhat2 = 23638.1

Therefore, xhat1 and xhat2 share the same values.

(d)

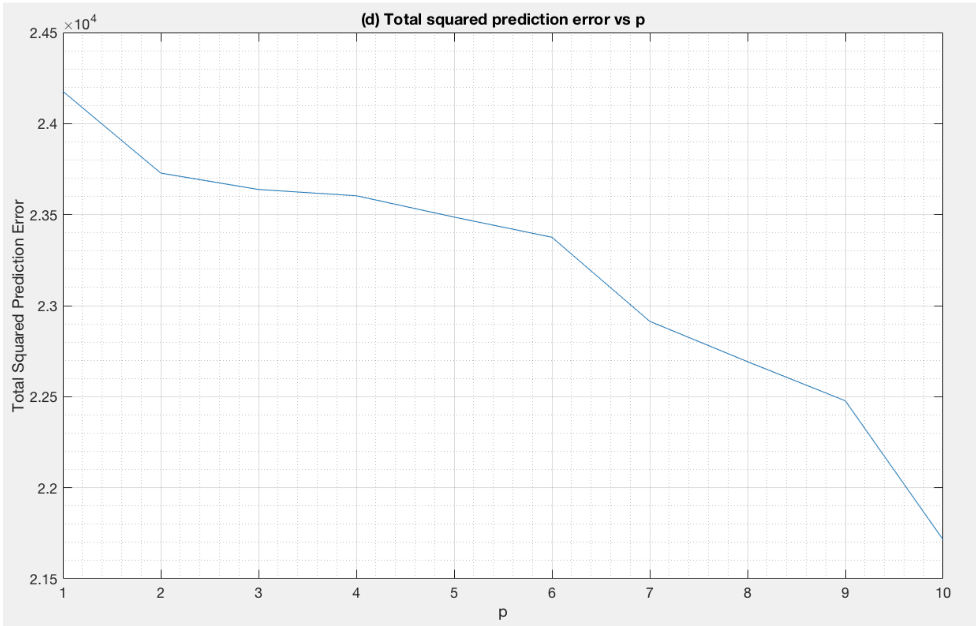

Figure 4. Plot of total squared prediction error as a function of p for p = 1 to 10.

We observe a linear decreasing relationship between p and total squared prediction errors. There is no observable "knee". Given that the range of p is limited from 1 to 10, p value is chosen such that it has the overall lowest total squared prediction error. Therefore, p = 10.

(e) Given p = 10, we want to make 520 trading decisions starting from the 11th week i.e. week 11 to week 530. Using several investment strategies, the results are as follows:

- The upper bound of returns, the best possible method, that is if we were always right and invested in the better of the bank or the DJIA would be $4,700,565.98, giving us a profit of **$4,699,565.98.**
- The lower bound of returns, 'all in the bank account' method, that is if we left all of our money in the bank would be $1,349.74, giving us a profit of **$349.74.**
- The second lower bound of returns if we use the "buy-and-hold" strategy in the DJIA would be $544.44, giving us a profit of **$ -455.56.** We lose money here.
- If we applied our linear predictor and invested in the better of either the bank or the DJIA, the return would be $1,422.65, giving us a profit of **$422.65**.

The bank would have to provide an APR of **3.526%** to achieve the same gain as the predictor.

(f) Using the same p = 10 and computed linear prediction coefficients from the first decade of data, we try to predict the DJIA values for the most recent 520 weeks. By repeating the investment strategies as before, we get the results as follows:

- The best possible bound would return $104,798.92. A profit of **$103,798.92.**
- The 'all in the bank account' bound would return $1,349.74. A profit of **$349.74**.
- The 'buy-and-hold' strategy would return $3,521.28. A profit of **$2,521.28**.
- If we used our linear predictor, it would return $2,265.53. A profit of **$1,265.53**.

The bank would have to provide an APR of **8.185%** to achieve the same gain as the predictor.

<u>Part II</u>

In this section, I used the autocorrelation method to answer the related questions.

(a) Using July 30, 2015 to Dec. 31, 2015 to predict Jan. 1, 2016 to June 30, 2016.

[a]*(a)* Predicting data from Jan. 1, 2016 to June 30, 2016 using July 30, 2015 to Dec. 31, 2015 as training data.

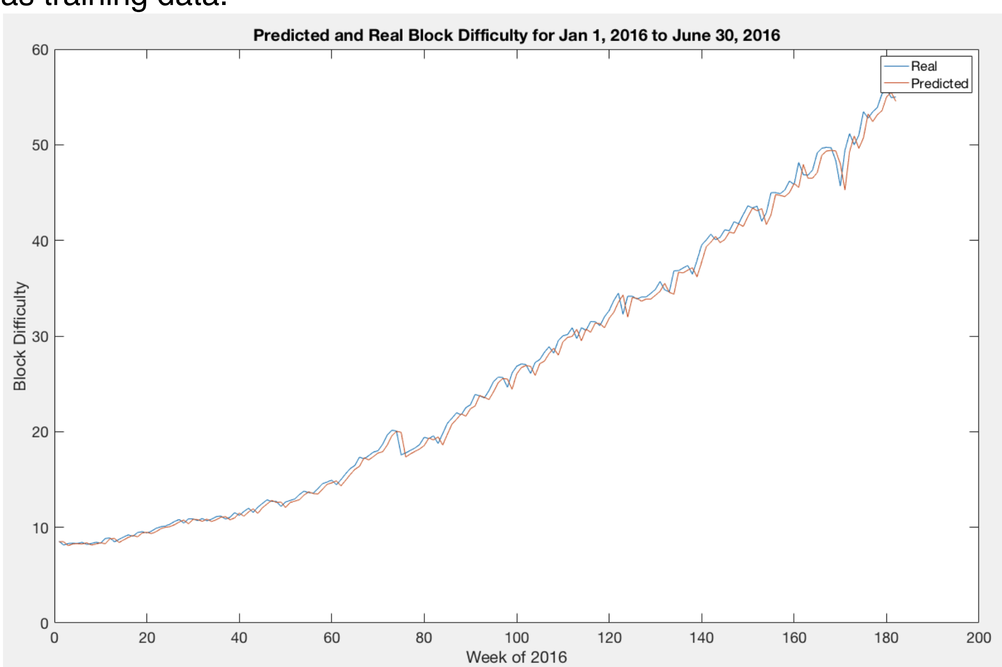

Figure 5. Plot of predicted difficulty and real difficulty with P = 2.

[a] *(b)* Computing least squares errors for P = [2:4:50]. The error is computed between the real and predicted values from Jan. 1, 2016 to June 30, 2016. E is defined as:

$$E = \sum_{n=1}^{182} (x[n] - \hat{x}[n])^2$$

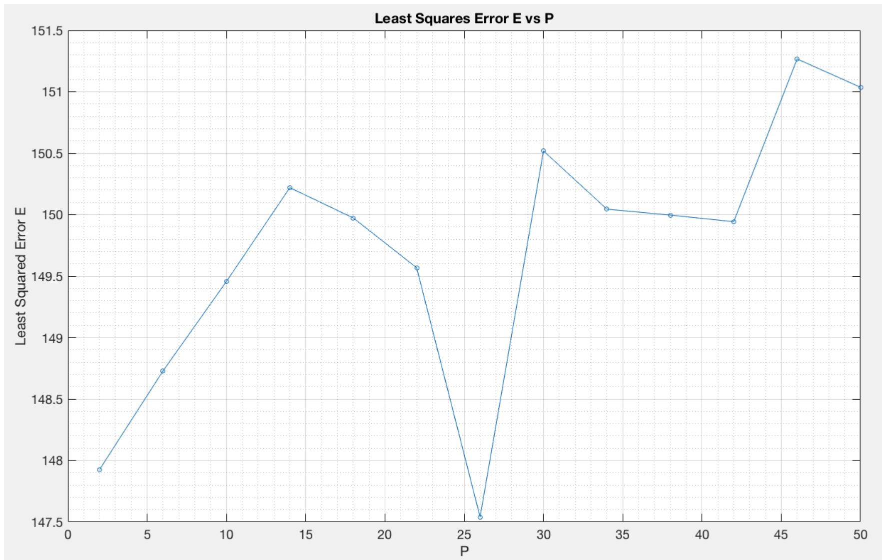

Figure 6. Plot of least squares error E vs P for P = [2:4:50].

[a] *(c)* Computing average predicted errors for P = [2:4:50]. The error is computed between the real and predicted values from Jan. 1, 2016 to June 30, 2016.

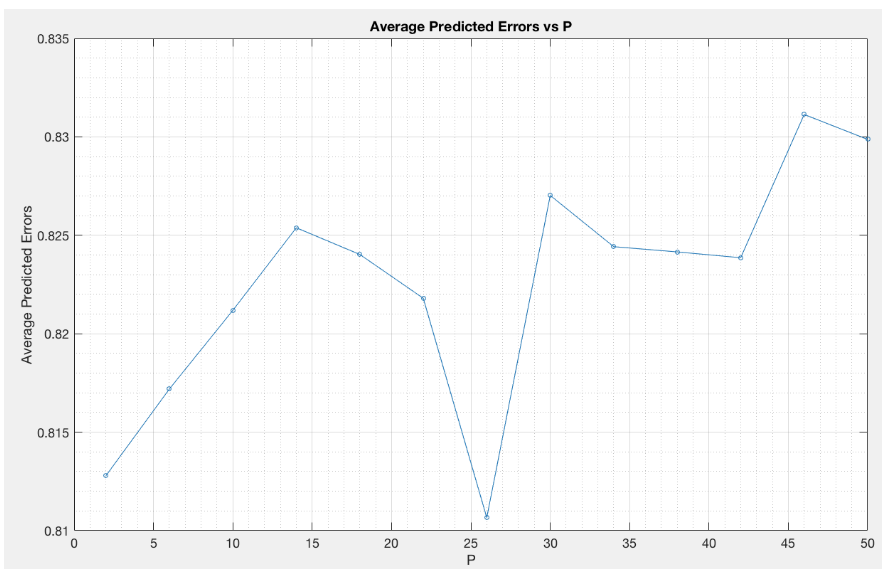

Figure 7. Plot of average predicted errors vs P for P = [2:4:50].

Both the plots for (b) and (c) look similar with different y-axis scaling. This is because the average predicted errors is just the least squares error divided by the total number of predicted days, which was calculated to be 182 days from Jan. 1, 2016 to June 30, 2016. Therefore, the y-axis values are scaled by 1/182 for part (a)-c.

(b) Using Jan. 1, 2016 to Dec. 31, 2016 data to predict values, P = 2.

b(a) Predicting difficulty from Jan. 1, 2017 to Dec. 31, 2017.

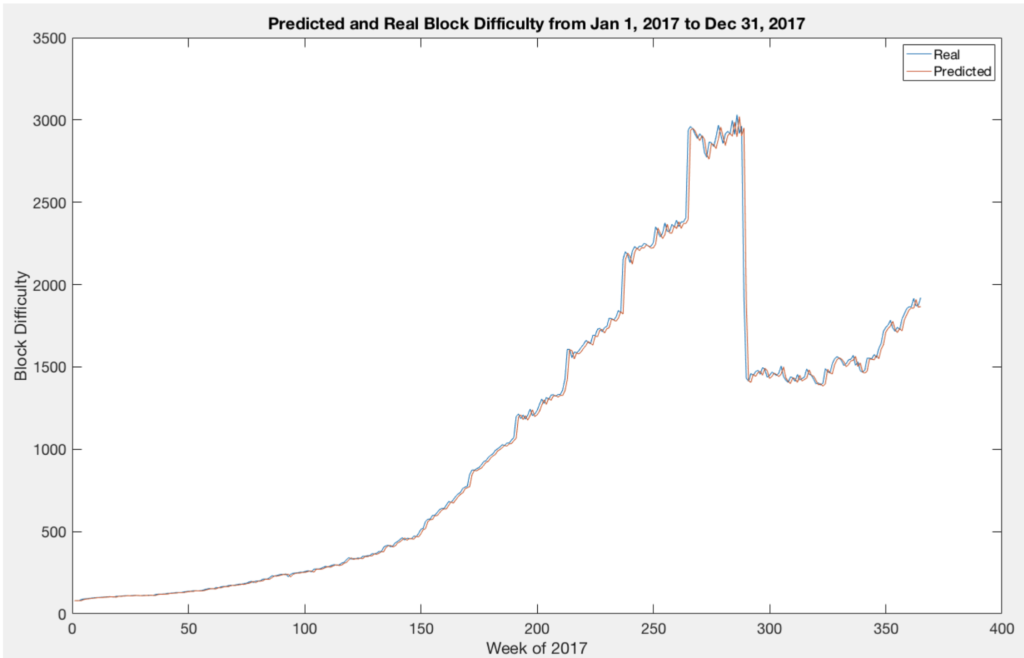

Figure 8. Plot of real and predicted block difficulty values from Jan. 1, 2017 to Dec. 31, 2017.

b(b) Predicting difficulty from Jan. 1, 2018 to Dec. 31, 2018.

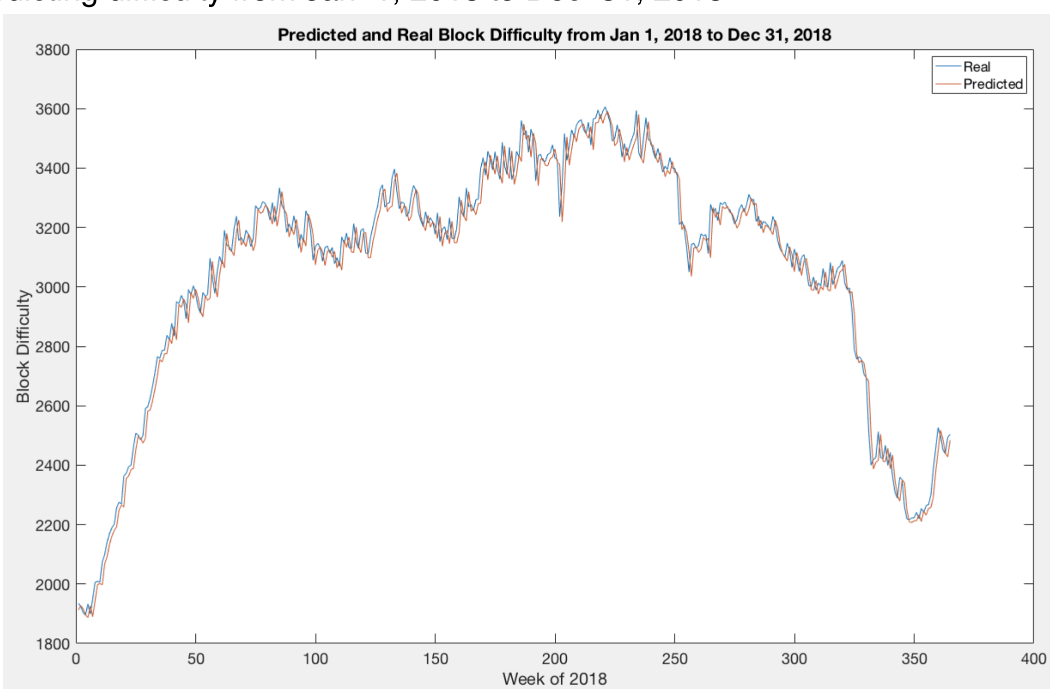

Figure 9. Plot of real and predicted block difficulty values from Jan. 1, 2018 to Dec. 31, 2018.

[b]*(c)* The average predicted errors for 2017 is 5531.24. The average predicted errors for 2018 is 3049.19. The error for 2017 is larger than 2018. This could be due to a sharp decline in block difficulty around week 288 to 290 in 2017 (around 16 to 18 October 2017). Also, looking up news related to ETH around those dates, it seems that Ethereum's developers released a Byzantium hard fork to make adjustments to the difficulty formula, causing the difficulty to decline dramatically. Since the linear predictor only uses past values and does not incorporate external factors to predict the block difficulty, it is not able to predict such a dramatic decline. Therefore, it produces a higher error compared to 2018 where there wasn't much of a sudden decline.

(c) Predicting difficulty data from Jan. 1, 2018 to June 30, 2018. P = 2.
[c]*(a)* Using previous year's data (365 days)

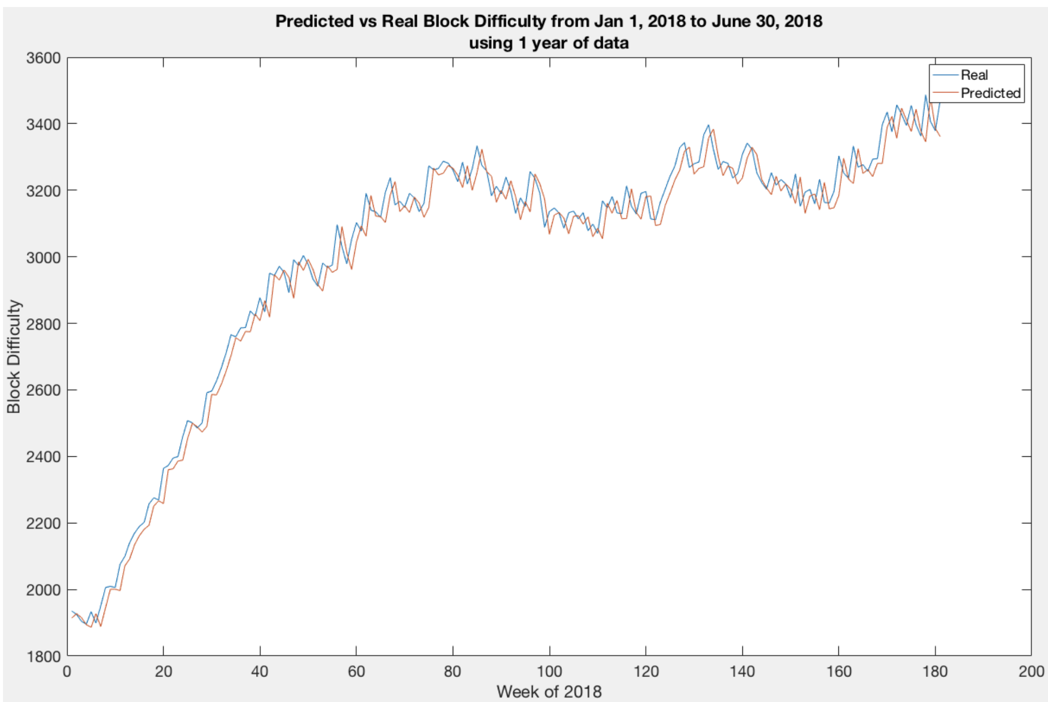

Figure 10. Plot of real and predicted block difficulty values from Jan. 1, 2018 to June 30, 2018.

[c](b) Using previous 6 months' data (180 days)

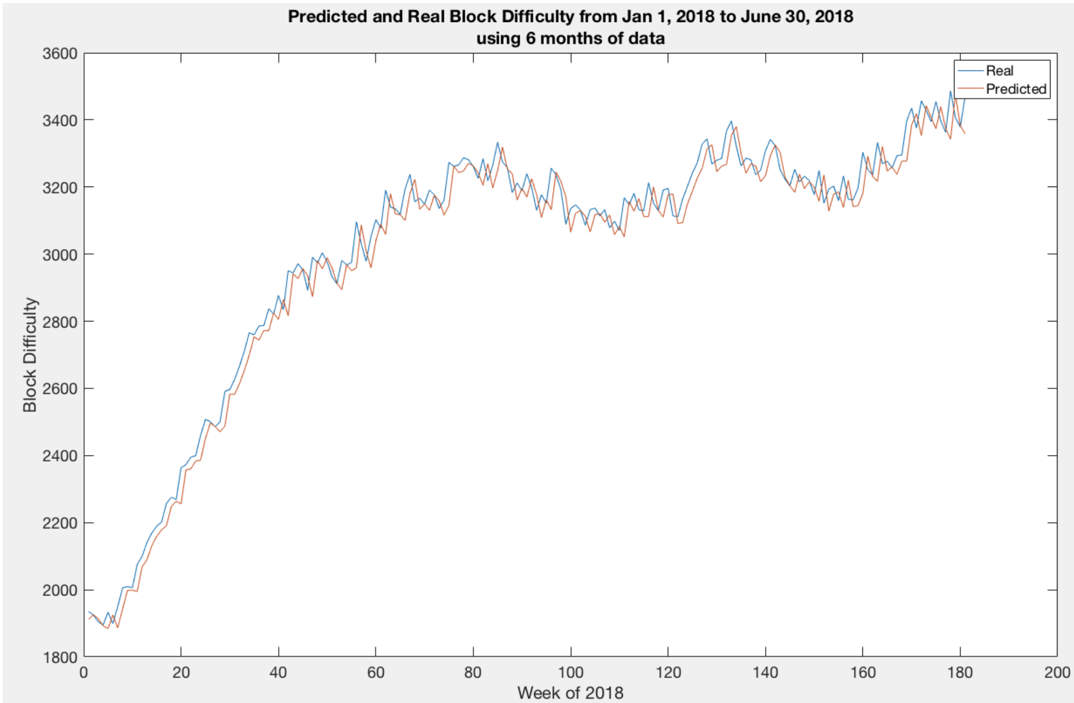

Figure 11. Plot of real and predicted block difficulty values from Jan. 1, 2018 to June 30, 2018.

[c](c) Using previous month's data (30 days)

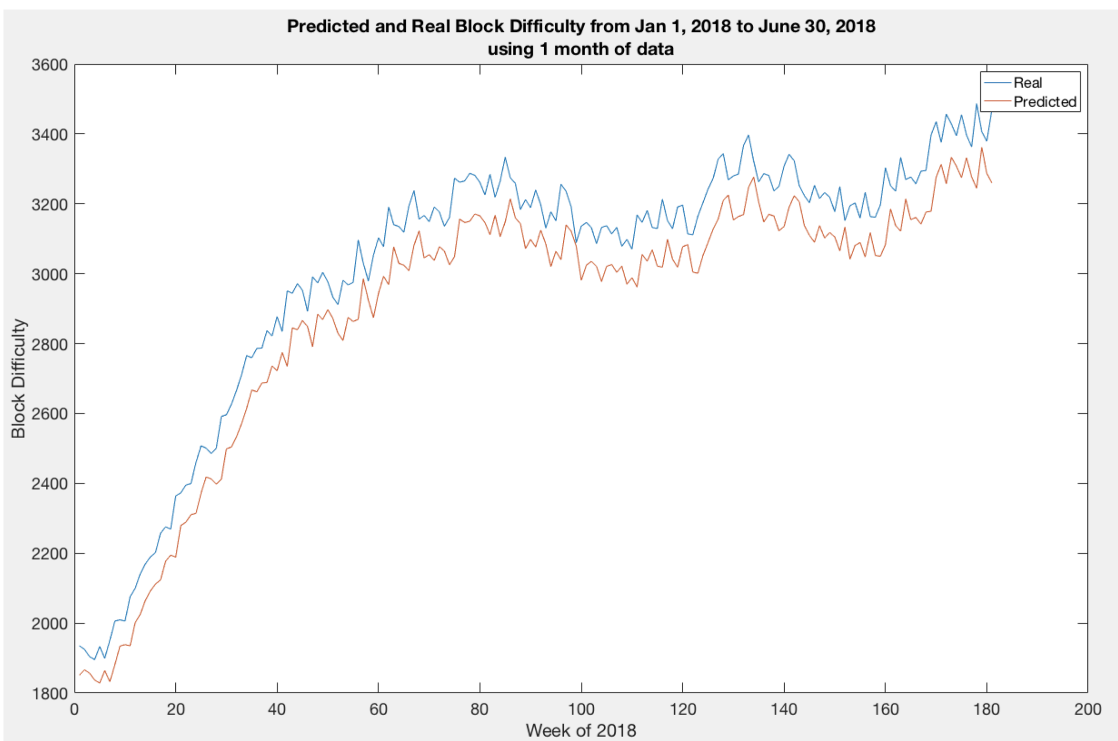

Figure 12. Plot of real and predicted block difficulty values from Jan. 1, 2018 to June 30, 2018.

The average predicted error for the three cases are as follows:

- One year (365 days) data : 3012.30
- 6 months (180 days) data : 3135.98
- 1 month (30 days) data : 15578.56

Since the predictor's accuracy should increase if we use more data, the computed values make sense. The shorter the duration of data used, the lower the accuracy of the predictor, the higher the average predicted errors. The predictor would perform better if we increased the training data.