The data source has been obtained from alpaca markets for historical data of JPM stock for past 7 years.

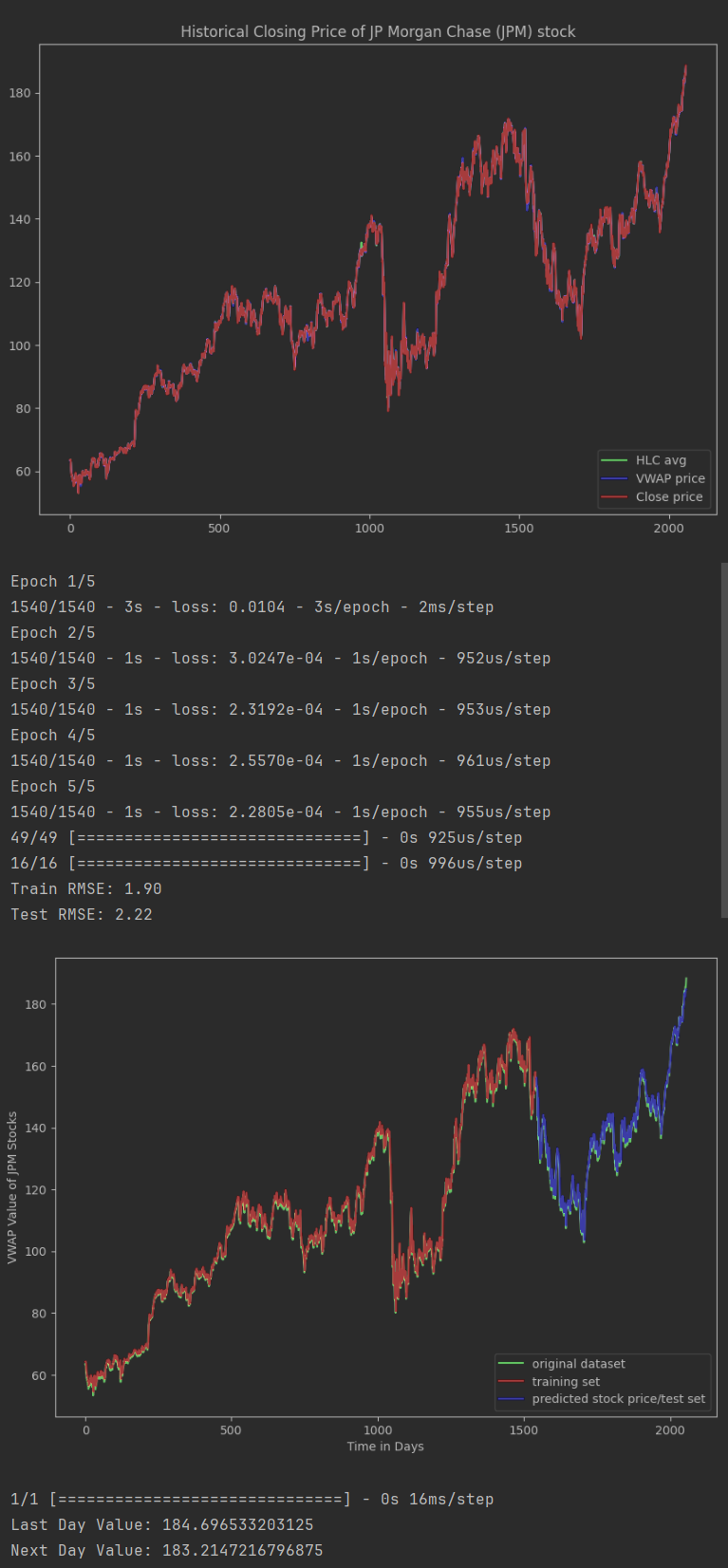
I have defined the LSTM with 32 neurons in the first hidden layer, 16 neurons in the second hidden layer and 1 neuron in the output layer for predicting pollution. The input shape will be 1 time step with 2 features.

I have used the Mean Squared Error (MSE) loss function and the efficient Adam version of stochastic gradient descent.

After the model is fit, we can forecast for the entire test dataset.

I have combined the forecast with the test dataset and invert the scaling. I have also inverted scaling on the test dataset with the expected pollution numbers.

With forecasts and actual values in their original scale, we can then calculate an error score for the model. In this case, we calculate the Root Mean Squared Error (RMSE) that gives error in the same units as the variable itself.



Since difference among VWAP average price, HLC average and closing value is not significant, so only VWAP average is used to build the model and prediction. The training and testing RMSE are: 1.90 and 2.22 respectively which is pretty good to predict future values of stock. Stock price of last day of dataset was 188.55 and using this model and price of next two days are predicted as 184.69 and 183.21 - which were 189.53 and 187.87 on 6th and 7th March 2024 according to Yahoo Finance. However, future values for any time-period can be predicted using this model.

Vincent, you are correct the problem is missing values!

The bizarre thing is the loss function should throw run-time error or propagate a warning cannot calculate a NaN value.