Stock Price Prediction

By Lilit Hovsepyan

Predicting:

We will work with historical data about the TATA stock prices of a publicly listed company. We will implement the Machine Learning algorithms to predict the future stock price, starting with simple algorithms, and then move on to advanced techniques.
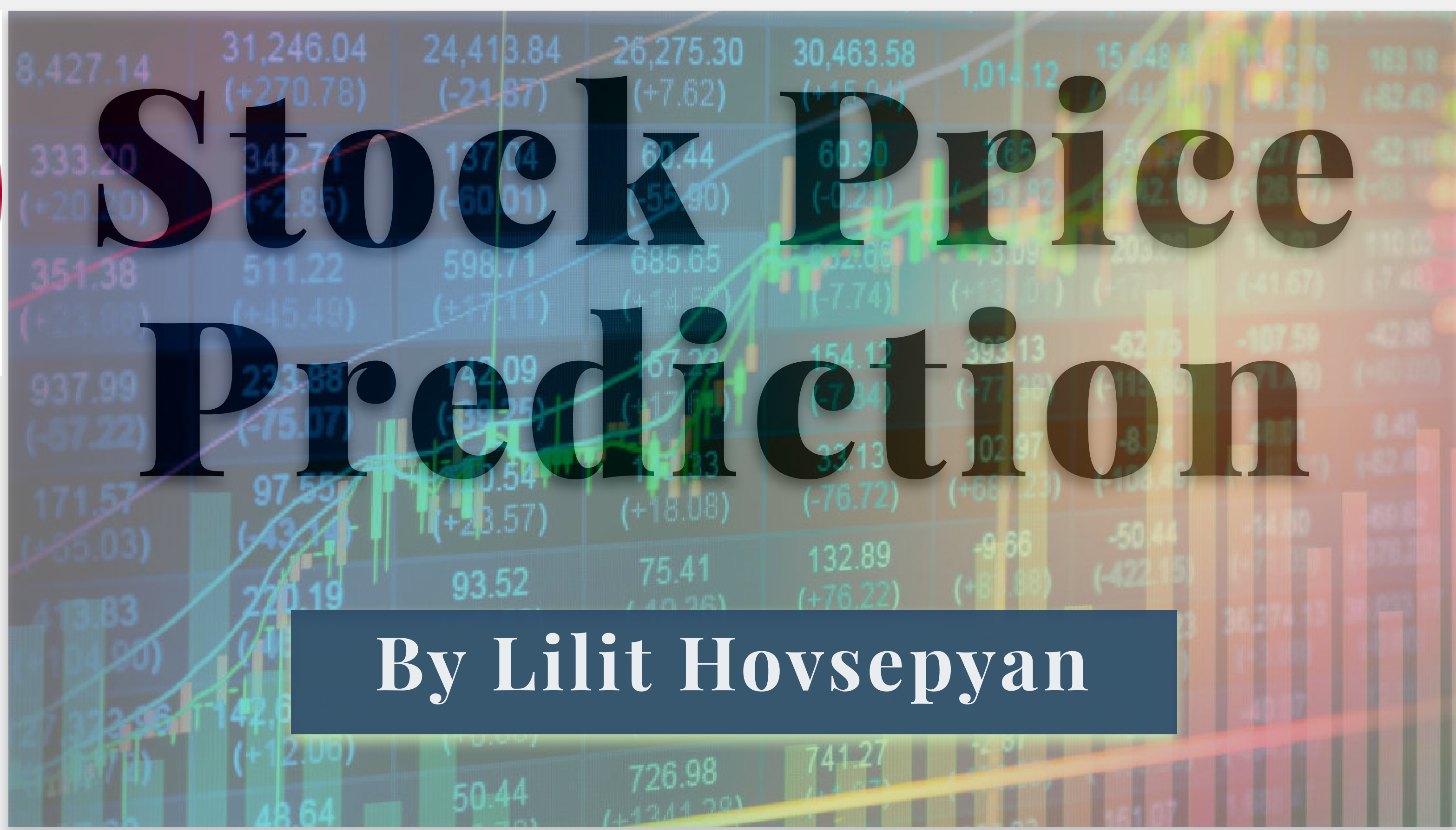
Data:

It's been used 'Tata Global Beverages' dataset from Quandl. We have Time Series data which shape is (2035,8) where columns are Date, Open, High, Low, Last, Close Prices, Total Trade Quantity and Turnover.

Features:

Our features are Close Price and Date. Afterwards we will split the Date feature into Year, Month, Week, Day, DayofWeek, DayofYear, Is_Month_End, Is_Month_start, Is_Quarter_end, Is_Quarter_start, Is_Year_end, Is_Year_start and will add one more feature which is Monday_Friday

Results:

Already from the Figure 2. it's clearly visible that LSTM predicted very well as the test set (orange line) and the prediction (green line) are almost the same. This model could detect all the ups and downs fitted in a way that looking at the lines it's very difficult to separate which one is the line in observed dataset and which one is in predicted dataset.

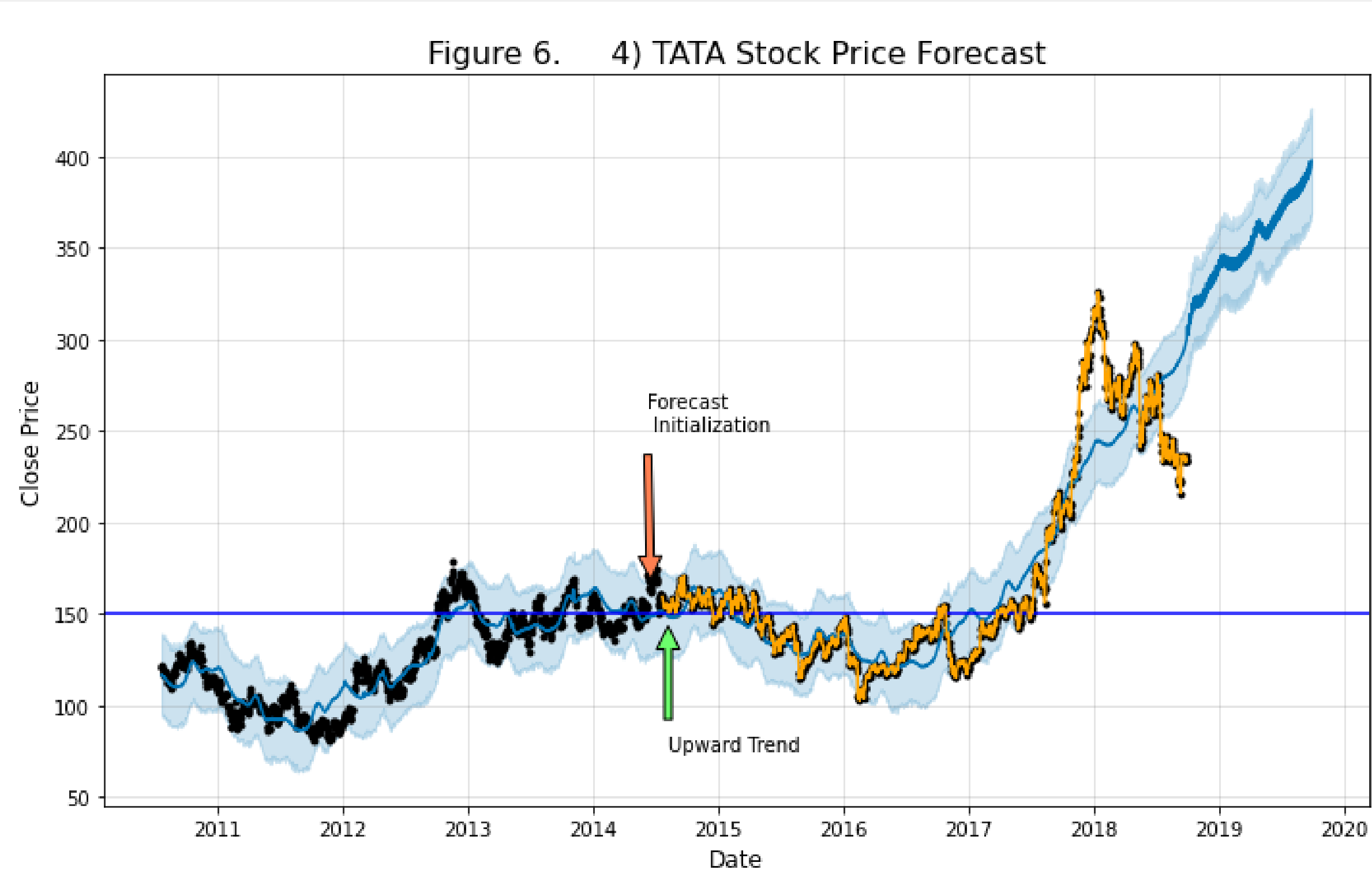| Table	LSTM	Prophet	LR	k-NN	ARIMA
RMSE	9.7	50.0	50.8	63.5	95.9

Figure 1. Prophet Prediction

Figure 2. LSTM Prediction

Models:

Here we used used Linear regression, k-NN algorithm, ARIMA and afterwards shift into the widely used algorithms such as Facebook Prophet (open source library) and LSTM by showing that these fit the model very well compared with the above-mentioned algorithms.

Future Work:

If I had another 6 months to work on this and additional resorces maybe I would rather go dipper into LSTM model, RNN and would try to add other features besides past price data, i.e. political decisions, S&P 500 changes over time etc.

Conclusion:

Concluding from our results, we see that using only past price data we could predict the future prices for a long time period. Lots of companies for their strategic goals (for 3-5 years) need long-term predictions. Here we've got the RMSE = 9.7 result for the long time period using LSTM method. And it's quite obvious that this model performs better than others.