Crude Oil Price Prediction Using IBM WATSON

Introduction

The accuracy of forecasting and prediction is more important and can assist organizations to take up-to-date decisions for better planning and management. So, we proposed the crude oil prediction model performance, different evaluation measures are used including Root Mean Square Error (RMSE) test. All the forecasting accuracy measures confirmed that our proposed model performs well in crude oil prices forecasting as compared to other hybrid models.

Crude oil is the world's largest fuel, and its prices have a significant impact on the global environment, economics, and oil exploration activities Although many methods are designed to predict oil prices, it remains one of the many challenging predictive problems due to severe oil price instability. To test the predictive ability of our broadcast learning model, we compare it with the other three popular oil price-prediction models. Test results show that our broadcast learning model achieves the highest accuracy in both definitions of square prediction error and measurement accuracy according to different weather time horizons. crude oil is a natural liquid that is found in the earth's crust under the surface of the earth. Crude oil prices are determined by many factors and have a significant impact on the environment and the economy. Although crude oil prices were strong in early 2014, they have fallen sharply since mid-2014. The earth's environment is affected by the fall in oil prices. In addition, there is little incentive to improve renewable and clean energy sources. On the other hand, lower oil prices could continue has led to a decline in oil and gas exploration activities around the world. Falling oil prices also play an important role in the global economy. Falling oil prices will lead to moderate global economic growth employment, although oil sector owners are losing revenue. The World Bank shows that for every 30% decline in oil prices, global GDP will increase by 0.5%. There is no doubt that forecasts for crude oil prices are of great industrial value, governments, and individuals. Therefore, predicting crude oil prices has been based on research by both academics and the industry. The key benefit of our way of learning to broadcast is that the predictive model can capture a variable pattern of oil prices as the model is continuously updated whenever new oil pricing data are available, with the smallest head always. Specific accuracy measures over a wide range of climatic conditions.

Problem Statement

Crude oil is the largest source of energy for the world. The prices of fuel price are in continuous fluctuation which makes it difficult to predict the price of fuel. Being such an important commodity it is important to know the price of the fuel. Hence our project aimed to make a model which can predict the fuel price for the next day based on previous data

Solution

We will use LSTM (Long Short Term Memory). and mean square error to estimate the crude oil prices for the given data set.We will predict the test and train data and will plot the graph accordingly. Predicting the price of crude oil for future 10 days. Plotting the graph for the prediction.

Literature Survey:

S.no	Reference Paper	Authors	Citation	Link
1	**Price Prediction of Share Market using Artificial Neural Network (ANN)**	Zabir Haider Khan Department of CSE, SUST, Sylhet, Bangladesh Tasnim Sharmin Alin Department of CSE, SUST, Sylhet, Bangladesh Md. Akter Hussain Department of CSE, SUST, Sylhet, Bangladesh	International Journal of Computer Applications (0975 – 8887) Volume 22– No.2, May 2011	As researchers and inventors strive to out-perform the market, the use of neural networks to forecast stock market prices will be a continuing area of research. The ultimate goal is to increase the yield from the investment. It has been proven already through research that the evaluation of the return on investment in share markets through any of the traditional techniques is tedious, expensive and a time consuming process. In conclusion we can say that if we train our system with more input data set it generate more error free prediction price

2	**House price prediction: Hedonic price model vs.artificial neural network**	Limsombunchao, V.	Limsombunchai, V. (2004). House price prediction: Hedonic price model vs. artificial neural network. New Zealand Agricultural and Resource Economics Society Conference, 25-26 June 2004. Blenheim, New Zealand: New Zealand Agricultural and Resource Economics Society.	The objective of this paper is to empirically compare the predictive power of the hedonic model with an artificial neural network model on house price prediction. A sample of 200 houses in Christchurch, New Zealand is randomly selected from the Harcourt website. Factors including house size, house age, house type, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location are considered. Empirical results support the potential of artificial neural network on house price prediction, although previous studies have commented on its black box nature and achieved different conclusions
3.	**Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction**	HarunaChiromaaSameemAbdulkareemaTututHerawanb	Applied Energy Volume 142, 15 March 2015, Pages 266-273	This paper proposes an alternative approach based on a genetic algorithm and neural network (GA–NN) for the prediction of the West Texas Intermediate (WTI) crude oil price. Comparative simulation results suggested that the proposed GA–NN approach is better than the baseline algorithms in terms of prediction accuracy and computational efficiency. Mann–Whitney test results indicated that the WTI crude oil price predicted by the proposed GA–NN and the observed price are statistically equal. Further comparison of the proposed GA–NN with previous studies indicated performance improvement over existing results.

Experimental Investigation

First, we take the dataset and look for any null values. If there are any null values as in this case the jupyter notebook will inform us about it.

```
In [7]: dataset.isnull().any()

Out[7]: Date             False
        Closing Value     True
        dtype: bool

In [8]: dataset.isnull().sum()

Out[8]: Date             0
        Closing Value    7
        dtype: int64
```

After finding out that null values are present we will find a way to replace null values with real values and then again run the same code as above to find out null values but this time we will not get any null values

```
In [10]: dataset.isnull().any()

Out[10]: Date             False
         Closing Value    False
         dtype: bool

In [10]: dataset.isnull().sum()

Out[10]: Date             0
         Closing Value    0
         dtype: int64
```

We will select the column that will be taken as input. for example in this case 'Closing Value' and plotting the graph according to the values of the input column in the dataset

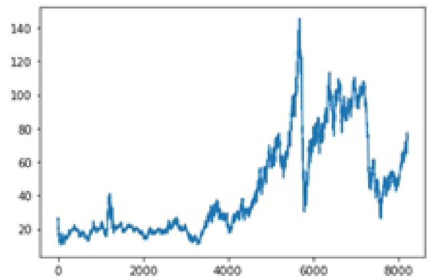
```
In [16]: oil = dataset.reset_index()['Closing Value']
```

```
In [17]: oil
```

```
Out[17]: 0       25.56
         1       26.00
         2       26.53
         3       25.85
         4       25.87
                 ...
         8211    73.89
         8212    74.19
         8213    73.05
         8214    73.78
         8215    73.93
         Name: Closing Value, Length: 8216, dtype: float64
```

```
In [18]: plt.plot(oil)
```

```
Out[18]: [<matplotlib.lines.Line2D at 0x2852c845790>]
```



On splitting the dataset into train and test we will get 5240 values in train and 2870 values in the test. For creating a Long Short-Term Memory model we will add dense layers

```
In [32]: from tensorflow.keras.models import Sequential
         from tensorflow.keras.layers import Dense
         from tensorflow.keras.layers import LSTM
```

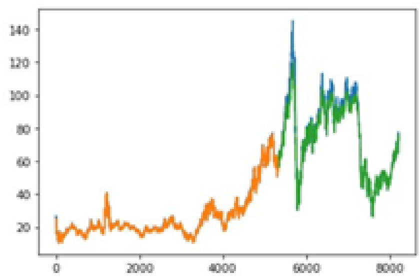
```
In [33]: model = Sequential()
         model.add(LSTM(50,return_sequences = True,input_shape = (10,1)))
         model.add(LSTM(50,return_sequences = True))
         model.add(LSTM(50))
         model.add(Dense(100))
         model.add(Dense(70))
         model.add(Dense(40))
         model.add(Dense(10))
         model.add(Dense(1))
         model.compile(loss = 'mean_squared_error',optimizer='adam')
```

For the model, evaluation means square error will be used.

```
In [38]: import math
         from sklearn.metrics import mean_squared_error
         math.sqrt(mean_squared_error(Y_train,train_predict))

Out[38]: 29.190087079864316

In [39]: math.sqrt(mean_squared_error(Y_test,test_predict))

Out[39]: 76.08525324099969
```

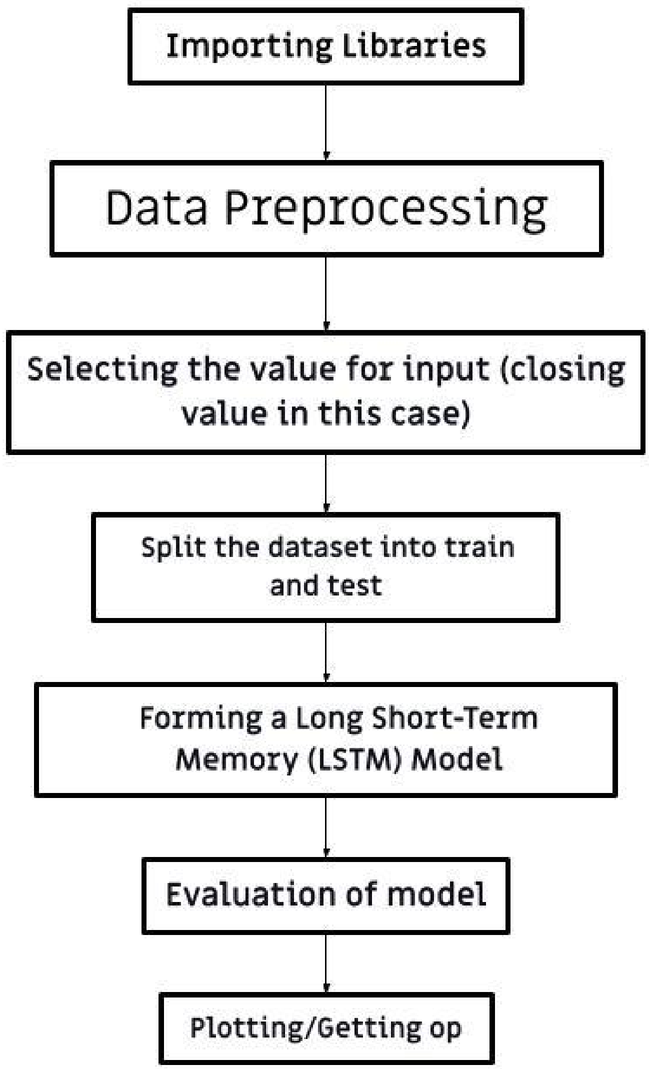
Train, test, and baseline predictions need to be shifted

```
In [40]: look_back = 10
         trainPredictPlot = np.empty_like(oil)
         trainPredictPlot[:, :] = np.nan
         trainPredictPlot[look_back:len(train_predict)+look_back, :] = train_predict
         testPredictPlot = np.empty_like(oil)
         testPredictPlot[:, :] = np.nan
         testPredictPlot[len(train_predict)+(look_back*2)+1:len(oil)-1, :] = test_predict
         plt.plot(scaler.inverse_transform(oil))
         plt.plot(trainPredictPlot)
         plt.plot(testPredictPlot)
         plt.show()
```



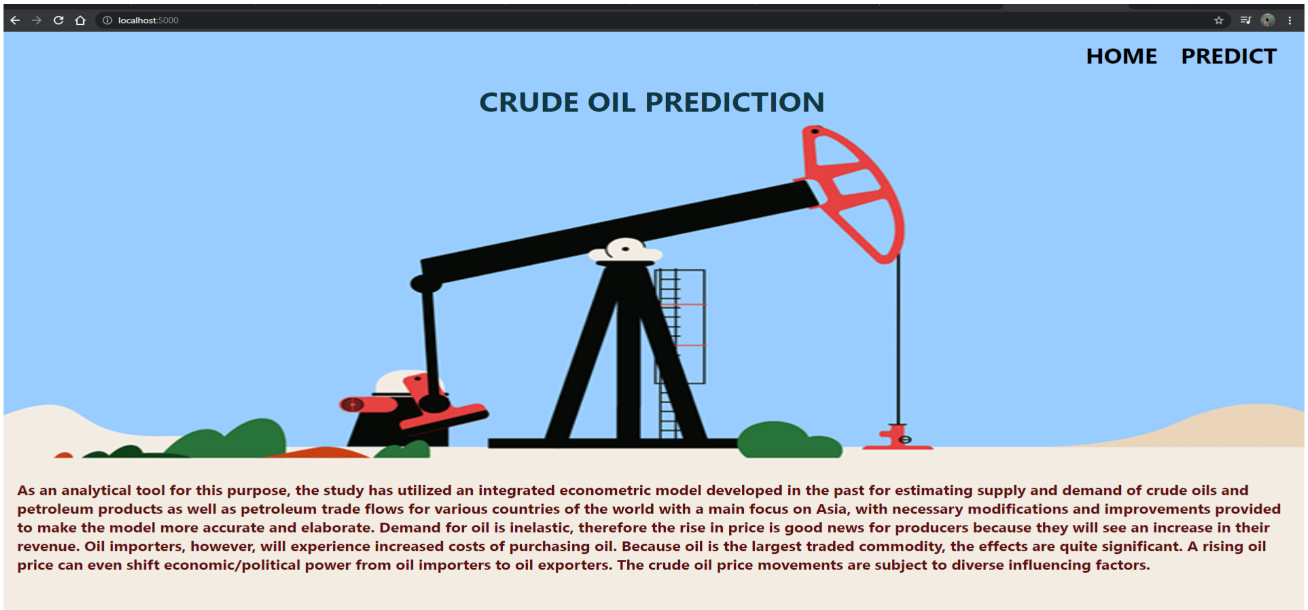
Hardware and Software Specification:
Flask==1.1.2
google-auth==1.21.3
google-auth-oauthlib==0.4.1
importlib-metadata==2.0.0
Keras-Preprocessing==1.1.2
numpy==1.18.5
oauthlib==3.1.0
tensorboard==2.3.0
tensorboard-plugin-wit==1.7.0
tensorflow==2.3.0
tensorflow-estimator==2.3.0

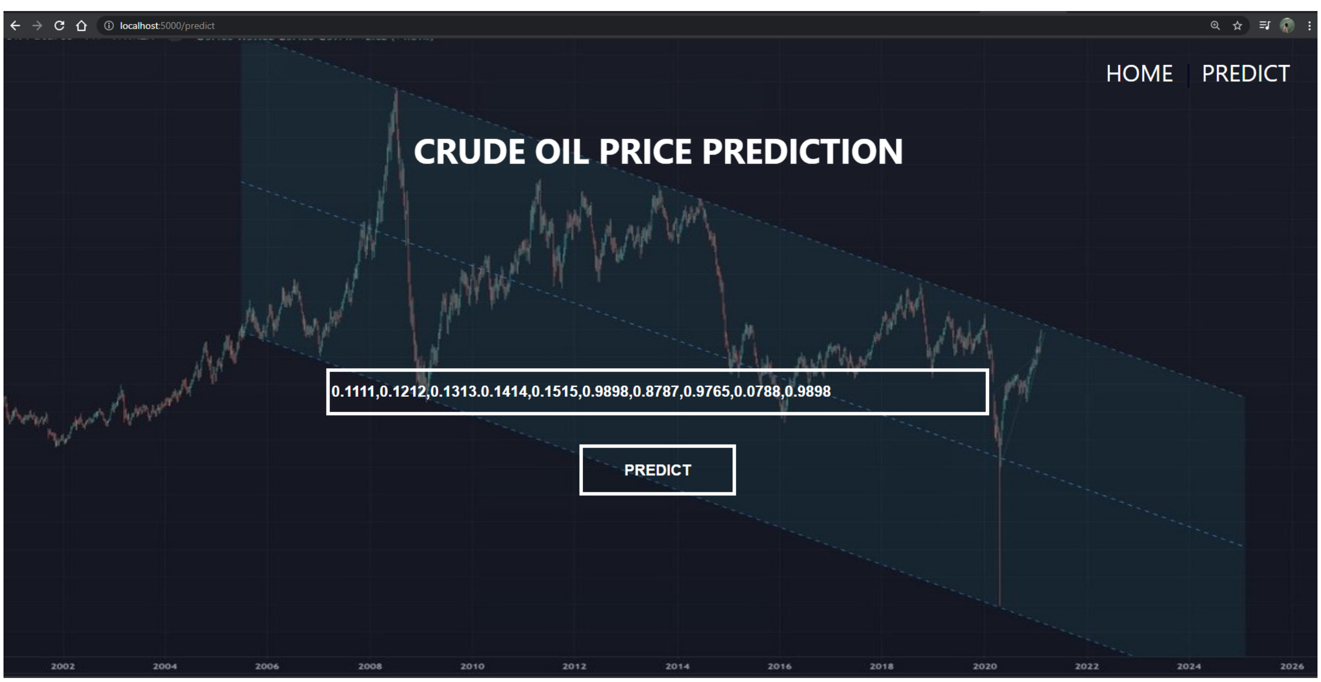
FlowChart

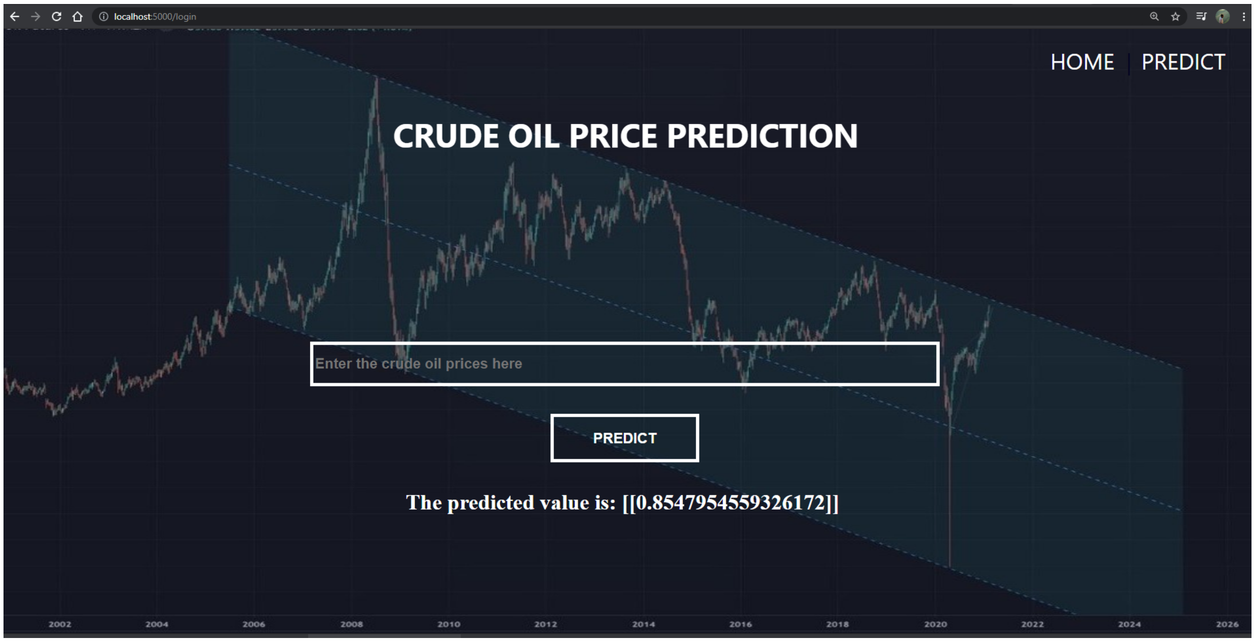

Importing Libraries
Data Preprocessing
Selecting the value for input (closing value in this case)
Split the dataset into train and test
Forming a Long Short-Term Memory (LSTM) Model
Evaluation of model
Plotting/Getting op

Conclusion:

The Prices are predicted using a Flask integrated website as shown below. We should input 10 values and according to the weights we got in training it will predict the output.







Future scope:

We can still use this LSTM method to predict various data and complex algorithms. Also if we can have huge data of prices in future we can increase its accuracy. As researchers and investors strive to out-perform the market, the use of neural networks to forecast prices will be a continuing area of research. The ultimate goal is to increase the accuracy of prediction. In conclusion we can say that if we train our system with more input data sets it generates more error free prediction prices.

Bibliography

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.4394&rep=rep1&type=pdf

http://researcharchive.lincoln.ac.nz/handle/10182/5198

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1066.9415&rep=rep1&type=pdf

http://eprints.lmu.edu.ng/3224/

https://www.sciencedirect.com/science/article/abs/pii/S0306261914013026