**CRUDE OIL PRICE PREDICTION**

**ABSTRACT: -**

Crude oil, also known as "black gold," has played a crucial part in the evolution of the world's wealth in the financial sector. Consequently, dynamic data of anticipated pricing in the future will result in improvement of making decisions at several levels. This an attempt was made to anticipate prices using long instead of employing, short-term memory neural network neural convolutional network. We have encountered testing several model iterations with various lookback and different tuning techniques. This study's conclusions are encouraging and indicate a more accurate forecast for crude oil cost in the upcoming days.

Keywords: - Recurrent Neural Networks, LSTM, Crude Oil Prediction.

**INTRODUCTION**

Since crude oil is one of the main items sold worldwide and hence incorporates global measurements, it plays a crucial role in the world. Errors in crude oil forecasting have a composite supply-demand structure at their root. Crude oil volatility has a significant impact on economic variables such as the country's economic growth, unemployment, and exchange rate, which also depends on the export and import of crude oil. Machine learning techniques can now be used in a variety of applications. Powerful computational tools and algorithms that can learn for themselves and anticipate data with a lengthy short-term memory are provided by machine learning. Recurrent neural networks built on LSTM are used in this paper to predict the price of crude oil. The most effective and powerful models for processing time-series-based sequential data are recurrent neural networks (RNNs). In addition to prediction, LSTM variations can be applied to tasks including voice, handwriting, and polyphonic modelling. Through the use of random search and variance framework analysis, the hyperparameters of the variations were evaluated. ANNs are made up of a group of nodes that resemble the neurons in a biological brain. The goal of ANN was to carry out tasks in a manner similar to that of the human brain. Due to its tendency to overfit relationships, ANN is rarely used in model prediction. It is also employed in situations where a similar recurrence of the past is anticipated in the future.

In data mining and machine learning, backpropagation is used to increase prediction accuracy. Based on the parameters of the neural network, it calculates the loss function's slope. The use of decision trees is another strategy for predicting oil prices. It is not only a very effective technique for prediction and classification but also for understanding how various variables behave. Superintend learning algorithm's Decision Tree employs categorial input and output variables. The economic variables that have an impact on the forecasting of crude oil prices can be the input values or qualities.

**SURVEY STUDY**

For finding, exploring, and projecting crude oil prices, numerous scientists and researchers have found various and distinct models. Uncertain research on forecasting and predicting prices that depend on economic models. Additional research on clever and smart algorithms is also taken into consideration. It was noticed that irregular events were causing variation to increase in the short term. The fact that the future movement of the oil price was derived from the historical data was one of the methodology's flaws. For the purpose of predicting oil prices, numerous machine learning strategies that can be categorised by ANN have been developed. Scientists believe that online discussion about a topic conveys a point of view or outlook depending on the issue, and that topic models may thus have an impact on market values. However, machine learning techniques often differ from conventional techniques in that they use a constant set of training data to guide a machine learning model, followed by a sequence of requests to apply the model to a test set. One of the most essential data sources identifies the crucial information that should be included in unstructured big data, such as text data that can be used for oil cast forecasts. In the field of text mining, a lot of studies have recently provided crucial information for retail prediction. A neural network built on the foundation of deep learning is the convolutional neural network. Its applications have mostly been in audio, picture, and sentence modelling recognition. Midterm Weighty Incident is the primary source of the price fluctuation, which the EMD approach validates. Unexpected and unplanned occurrences that crash a market are the result of the decomposition integration process, which causes performance to degrade.

**METHODOLOGY**

1. **System design**

The primary goal of the structure used in the study is to gather data from the economic news and suggest these sets into predictive model. Data collection and pre-processing, feature and factor selection, price evaluation, and price prediction are the main stages of the developed system. News, financial, and market data are acquired and analyzed in the initial hand. Additionally, CNN categorization transforms unstructured documents into structured extracts.

1. **Data retrieval and pre-processing**

Datasets including news data, black gold price data, and market data can all be retrieved using data retrieval. Because it is simpler to collect and justifies on a single line, headlines can be used to retrieve news datasets. The stock market, future business, and export business are all factors that influence the prediction.

1. **Sentimental Analysis**

Big data is assisting in this modern period by studying sentiment analysis, which focuses on collecting data from news and offering prediction models. To acquire information about marketplaces and key factors influencing them, this type of study takes a dictionary-based method. The sentiment and prediction models are regarded as variables in trend prediction.

1. **Back Propagation**

Back-propagation is thought of as a method that can be used to train feed-forward neural networks for prognostic learning models. Due of this, it is possible to teach multi-layer networks using gradient approaches by adjusting the weights to cause the least amount of loss. The procedure gathers the inputs and outputs and modifies its internal state to be able to calculate an output that will be extremely accurate to the outcome that was anticipated.

Backward propagation of mistakes is another name for backpropagation. Artificial neural networks can be taught naturally.

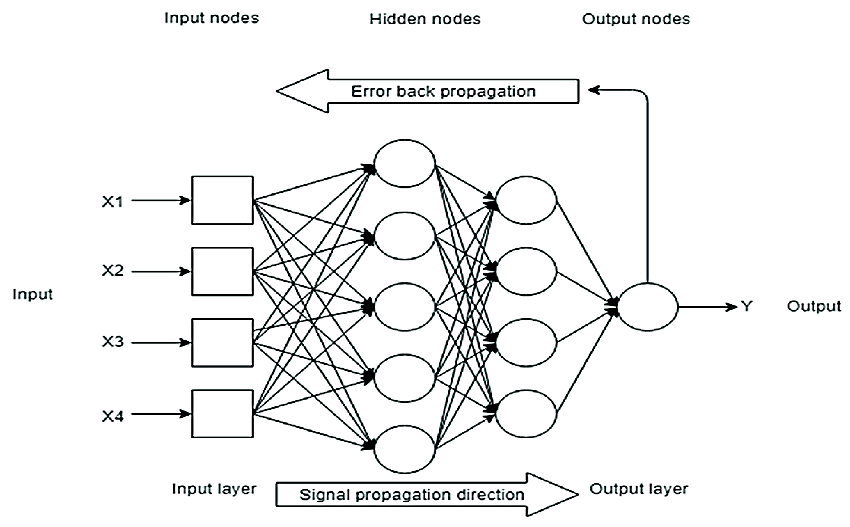


Fig 1:- Back Propagation in neural networks.

1. **CNN (convolution neural network)**

Another name for CNN is a kind of neural network that uses convolution and a pooling layer to forecast time series. The next layer receives the result from the convolution layer. On the other hand, a pooling layer can be used to reduce the dimensionality or the number of parameters. TensorFlow can be used for predicting, which is followed by data collection.

creating features, creating a machine learning model, training the ML model, and using the ML model for testing and prediction. CNN categorization can be thought as as the movement of the price, which might go up or down.

In general, the pricing activity can be described by:

Mt = {0 , pt < pt-1 }{1, pt >= pt-1}

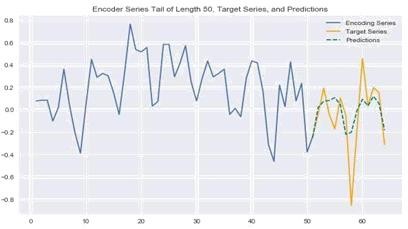


Fig 2: Time Series Forecasting with CNN

1. **RNN (Recurrent Neural Network)**

Recurrent Neural Network (RNN) are a type of Neural Network where the output from previous step is given as input to the current step. They use their internal memory for prediction. RNN have a “memory” which remembers all information about what has been calculated and evaluated. It uses the same parameters for every input as it does the same task on all the inputs or hidden layers to predict the result. This reduces the complexity of parameters, unlike other neural networks. They are implemented for many applications such as handwriting recognition and speech recognition. The Neural Networks uses sequential information which are available to them.

RNNs and feedforward networks are not the same. It draws on its internal memories to make predictions. It can perform tasks that humans are not capable of, including speech and handwriting recognition. These networks receive additional input in the form of sequential data. Our presumption that inputs are unrelated to one another is incorrect. Knowing the preceding words will let us predict the following ones without difficulty.

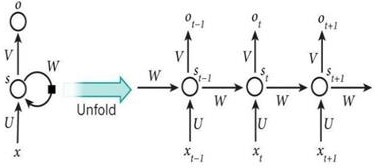
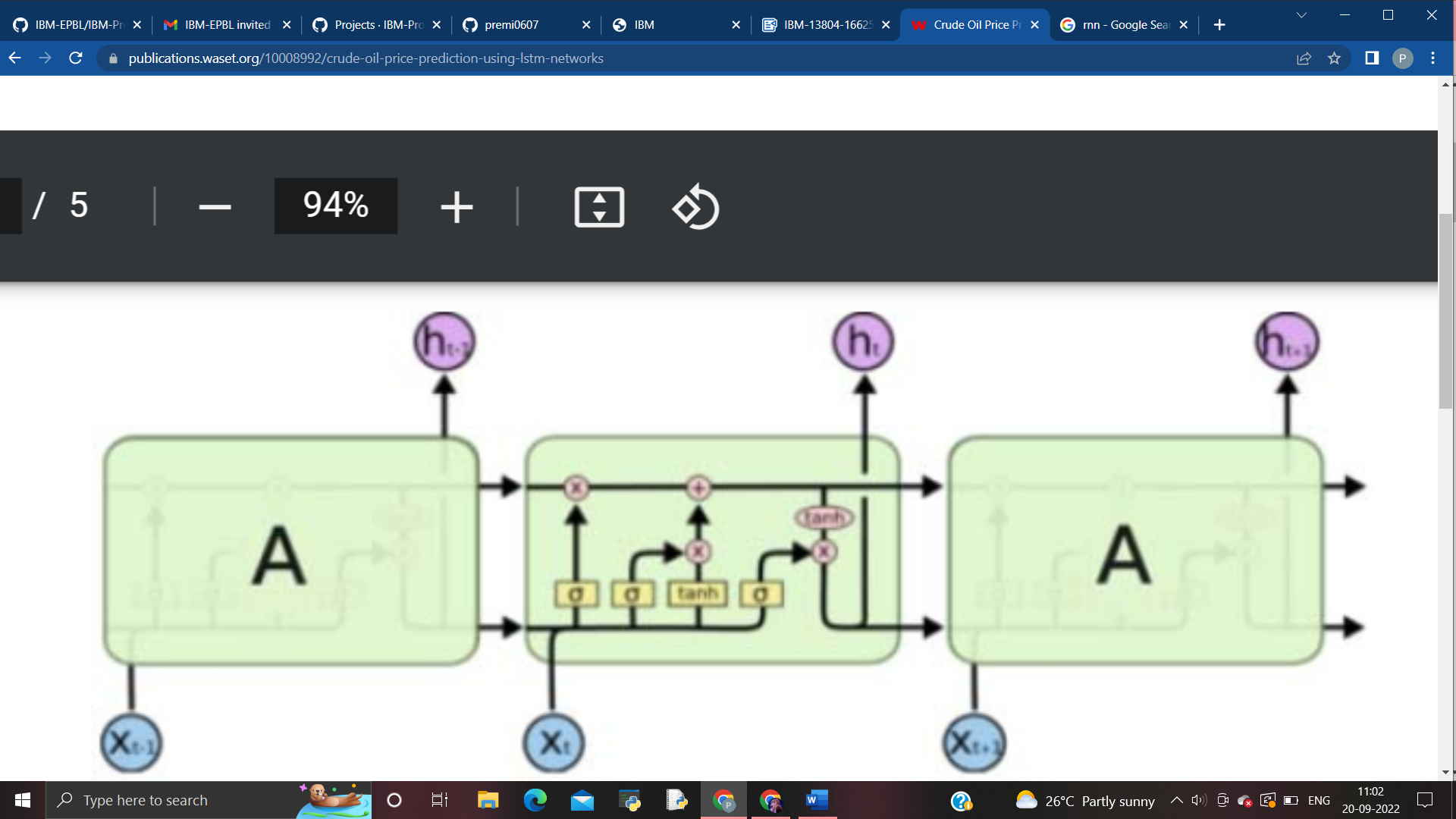


Fig 3:- RNN

Bidirectional RNN is one of RNN's extensions. The output in this case at time t also depends on subsequent inputs.

1. **The LSTM networks**

The most popular and widely used type of Recurrent Neural Network is the Long Short-Term Memory (LSTM). These recurrent neural networks learn order dependence in continuous values or sequence prediction problem. The LSTM networks overcomes two major issues which is encountered in RNN. The two issues are vanishing gradients and exploding gradients. The key to the solution of these problems was the internal structure that has been used in LSTM. The simple architecture of LSTM networks is called as vanilla LSTM which performs very well in all sequence related prediction problems.



**RESULT DISCUSSION**

Various designs were tested in order to choose the final architecture for the specified network. The test-cases are represented in the following table.



Table 1:- The image above represents a few of the outcomes of the defined network method. Finally, it completes the appropriate outcomes, which are then regarded as the ultimate configuration of the network architecture that was suggested.

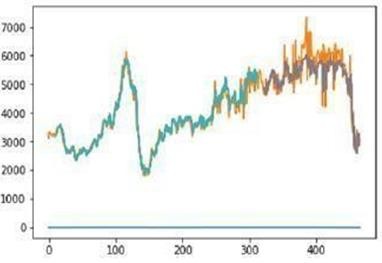


Fig 5:- Results using 100 epochs, 10 look back, and 4 LSTM layers.

224.19 RMSE for the train 550.50 RM on the test SE

Figure 5 illustrates the graph representation of the data we collected, which contains 10 lookbacks, 100 epochs, and 4 LSTM layers.

Thus, we arrived at the train score of 224.19 RMSE and the test score of 550.50 RMSE.

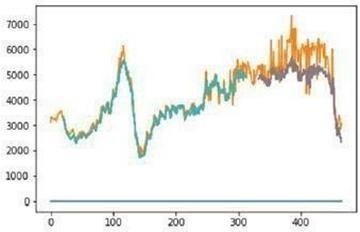


Fig 6:- Results with 20 look back,100epochs and 6LSTM layer.

Test Score:793.24RMSE Train Score:235.12RMSE

Figure 6 illustrates the data in this case, which consists of 20 lookbacks, 100 epochs, and 6 LSTM layers. As a result, we can draw the following conclusion about the values:

The train has a 235.12 RMSE rating. The test result is an RMSE of 793.24.

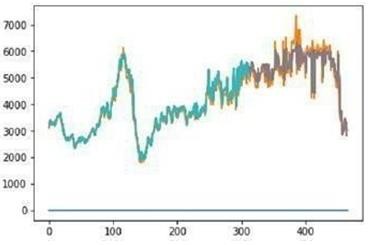


Fig 7:- Results with 10 look back,100epochs and 3LSTM layer.

Train: 269.17 RM in score SE Exam result: 566.34 RM SE

The train score is shown to be 269.17 RMSE and the test score is shown to be 566.34 RMSE in Figure 7, which describes the data that contains 10 look backs, 100 epochs, and 3 lstm layers. Results with 10 look backs, 50 epochs, and 4 LSTM layers, shown in Fig. 10.

The aforementioned outcomes make it very evident that 4LSTM, with 10 lookbacks and 50 epochs, is the best option for this specific problem statement. This shown that LSTM is a superior model for recurrent neural networks to forecast crude oil prices when compared to typical RNN feed forward models.

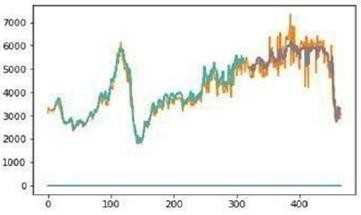


Fig 8:- Result with 10 lookback, 50 epochs and 4lstm layer.

The final test-case explains the graph using data from 50 epochs, 10 lookbacks, and 4 LSTM layers. The train score is 283.34 RMSE, and the test score is 532.13 RMSE.

**CONCLUSION**

As it uses a backpropagation model, this paper clearly shows that an LSTM network is superior to other standard neural networks for price predicting. On the other hand, a traditional neural network, such as a rrn or a CNN, predicts the next output without necessarily saving the previous data or connection, as the previous data is not required to predict the future data. The rather encouraging and more approximative LSTM focuses on retaining the prior data and prediction. The conclusions reached are generally positive. The findings demonstrate that large look ups do not always increase the precision of crude oil price predictions. Thus, it can be said that the model that uses a single LSTM model is undoubtedly the most accurate.

**REFERENCE:**

[1] <https://www.sciencedirect.com/science/article/pii/S1877050920305913>

[2] <https://www.sciencedirect.com/science/article/pii/S167498711630086X>

[3] <https://link.springer.com/chapter/10.1007/11758549_63>

[4]<https://corporatefinanceinstitute.com/resources/knowledge/economics/crude-oil-overview/>

[5] <https://www.digitalrefining.com/article/1002330/forecasting-crude-oil-prices#.YymJQHZBxPY>