ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

PREDICATION OF CRUDE OIL PRICES USING SVR WITH GRID SEARCH CROSS VALIDATION ALGORITHM

Dr.Adelina Johnsana[1], Dubba Pranay[2], J.karthik Chary[3], Talari Harathi[4], Nikhil Kumar Bellamkonda[5]

[1] Associate Professor, Dept. of AI-ML, Sri Indu College of Engineering and Technology, Hyderabad,
[2 3 4] Research Student, Dept. of AI-ML, Sri Indu College of Engineering and Technology, Hyderabad

ABSTRACT

Crude oil is the world's leading fuel, and its prices have a big impact on the global environment, economy as well as oil exploration and exploitation activities. Oil price forecasts are very useful to industries, governments and individuals. Although many methods have been developed for predicting oil prices, it remains one of the most challenging forecasting problems due to the high volatility of oil prices forecasting models that predict future events are used in numerous fields such as economics and science because they are useful tools in decision making. A perfect forecast provides insight into the implications of an action or inaction and serves as a metric to judge one's ability to influence future events, The world's environment is affected by the oil price falling. With the drop of oil prices, the fuel bills are lowered. As a result, consumers are very likely to use more oil and thus increase the carbon emission. In addition, there is less incentive to develop renewable and clean energy resources. On the other hand, sustained low oil prices could lead to a drop in global oil and gas exploration and exploitation activities. Fluctuating oil prices also play an important role in the global economy The fall in oil prices would result in a modest boost to global economic activity, although the owners of oil sectors suffer income losses. Recent research from the World B an shows that for every 30% decline of oil prices, the global GDP (Gross Domestic Product) would be increased by 0.5%. At the same time, the drop of oil prices would reduce the cost of living, and hence the inflation rate would fall. So there is a chance of prediction of the Proper and most approximate prediction in order to fix the situation if any occurs.

K**eywords:** —fore casting, machine learning,GDP,oil and gas

1 INTRODUCTION

Crude oil is a mixture of comparatively volatile liquid hydrocarbons (compounds composed mainly of hydrogen and carbon), though it also contains some nitrogen, sulfur, and oxygen. Those elements form a large variety of complex molecular structures, some of which cannot be readily identified. Regardless of variations, however, almost all crude oil ranges from 82 to 87 percent carbon by weight and 12 to 15 percent hydrogen by weight.

Crude oils are customarily characterized by the type of hydrocarbon compound that is most prevalent in them: paraffins, and aromatics. Paraffins are the most common hydrocarbons in crude oil; certain liquid paraffins are the major constituents of gasoline (petrol) and are therefore highly valued. Naphthene are an important part of all liquid refinery products, but they also form some of the heavy asphalt like residues of refinery processes. Aromatics generally constitute only a small percentage of most crudes. The most common aromatic in crude oil is benzene, a popular building block in the petrochemical industry. Because crude oil is a mixture of such widely varying constituents and proportions, its physical properties also vary widely. In appearance, for instance, it ranges from colourless to black. Possibly the most important physical property is specific gravity (i.e., the ratio of the weight of equal volumes of a crude oil and pure water at standard conditions). In laboratory measurement of specific gravity, it is customary to assign pure water a measurement of 1; substances lighter than water, such as crude oil, would receive measurements less than 1. The petroleum industry, however, uses the American Petroleum Institute (API) gravity scale, in which pure water has been arbitrarily assigned an API gravity of 10°.

2 RELEATED WORK

To further investigate the role of variable selection, we use the OLS estimates to replace the LASSO estimates for the coefficients of selected predictors. The results show that the OLS regression models based on the predictors selected by the elastic net and lasso still exhibit better out-of-sample forecasting performance than the competing models. We also find that the elastic net and lasso with a fixed number of selected predictors yield acceptable out-of-sample performance. In summary, the elastic net and lasso can parsimoniously select powerful predictors, which contributes to the improvement of oil price predictability from both statistical and economic perspectives\

Our paper is related to the studies of Li et al. (2015) and Li and Tsiakas (2017). However, our paper provides two additionally major contributions. First, we focus on the oil price predictability, while Li et al. (2015) and Li and Tsiakas (2017) investigate the predictability of exchange rates and stock returns, respectively. Moreover, we provide some additional analysis such as forecast encompassing tests and out-of-sample tests for the elastic net and lasso with a fixed number of selected predictors, which also provides insightful results. In addition, we use a relatively efficient algorithm to estimate LASSO parameters, thus generating accurate forecasts with substantially shorter computation time. 5 Second and more importantly, we provide insights into the role of variable selection, which is, however, neglected by Li et al. (2015) and Li and Tsiakas (2017). Li et al. (2015) and Li and Tsiakas (2017) recommend not only the elastic net and lasso but also the similar method of ridge regression, whereas we regard ridge regression as a competitor and thus recommend only the elastic net and lasso. This is because ridge regression uses the L2 penalty that only shrinks coefficients to zero but never sets them to zero exactly. In contrast, as the lasso uses the L1 penalty and the elastic net uses both the L1 penalty and the L2 penalty, the coefficients of unselected predictors are set to zero exactly. Furthermore, the ridge coefficients of correlated predictors are shrunk towards each other, while the elastic net and lasso tend to pick one and ignore the rest of the correlated predictors. For this motivation, we elaborately investigate the role of variable selection for the elastic net and lasso. gate the role of variable selection for the elastic net and lasso. The remainder of the paper is organized as follows. Section 2 presents the econometric methodology. Section 3 describes our data. Section 4 provides the out-of-sample empirical results and discussion. In Section 5, we present the robustness checks. Section 6 reports the results of the asset allocation analysis. Finally, Section 7 conclude

3 implementation study

Existing System:

The prediction of the crude oil rates based on the previous datasets on the data and prices as the feature _list are inputs and and target list are predicted values. The implementation was on the Linear Regression Model which is feasible to some extend for the prediction of the crude oil prices. The implementation is on predicting the crude oil prices for the days using Linear Regression Python Machine Algorithm and plotting the graph based on the prediction.

3.1 Disadvantages:

Using Linear Regression algorithm gives less approximate prediction compared to SVR Algorithm in the proposed model in the project. As well the feature list and target list fitted into the algorithm gives less predicting prices compared to the SVR, Comparatively

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

Linear regression performs poorly when there are non-linear relationships. They are not naturally flexible enough to capture more complex patterns, and adding the right interaction terms or polynomials can be tricky and time-consuming.

4. Proposed System & alogirtham

We have implemented SVR algorithm (Support Vector Regression) of Machine learning using Python. The predictions are most approximate with SVR Algorithms as they Linear or Gaussian. The algorithm automatically uses the kernel function that is most appropriate to the data. SVM uses the linear kernel when there are many attributes (more than 100) in the training data, otherwise it uses the Gaussian kernel. In the proposed system we have takes taken the datasets which has the price and days based on the dataset we have made feature list and target list where the target list is price values and feature list is the days. After the analysis of data is done we have fitted both feature list and target list using Python Machine learning SVN Algorithm and predicted the values for 1,30 and 365 days from the last day of the dataset values. Finally we have plotted a graph based on the results from the predicted analysis done with SVN Algorithm.

4.1 Advantages:

SVMs are a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields. Applied to solvency analysis, the common objective of all these ,It has a regularization parameter, which makes the user think about avoiding over-fitting. Secondly it uses the kernel trick, so you can build in expert knowledge about the problem via engineering the kernel. Thirdly an SVM is defined by a convex optimization problem (no local minima) for which there is efficient methods (e.g. SMO). Lastly, it is an approximation to a bound on the test error rate, and there is a substantial body of theory behind it which suggests it should be a good idea. The results that are generated by this algorithm gives more approximate and accurate calculations of the price prediction value compared to the other prediction algorithm for the dataset provided.

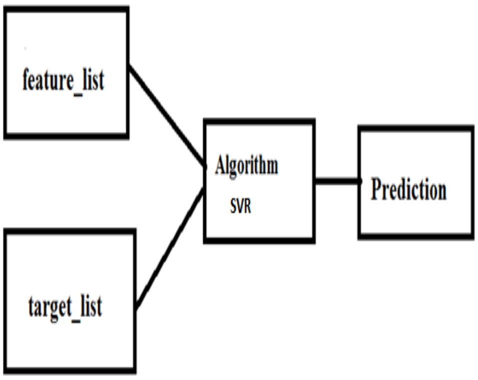

Fig1: proposed Model

4.2 Support Vector Machine:

The Support Vector Machine is one of the most widely used Machine Learning algorithms. The main goal of this algorithm is to find the best data split possible. It is used to solve problems involving classification and regression. It can solve both linear and nonlinear separable data, which is one of its main advantages. The separation line is known as the Hyper plane. Support vectors are the points on which the margins are built. The svm algorithm is depicted in the diagram below.

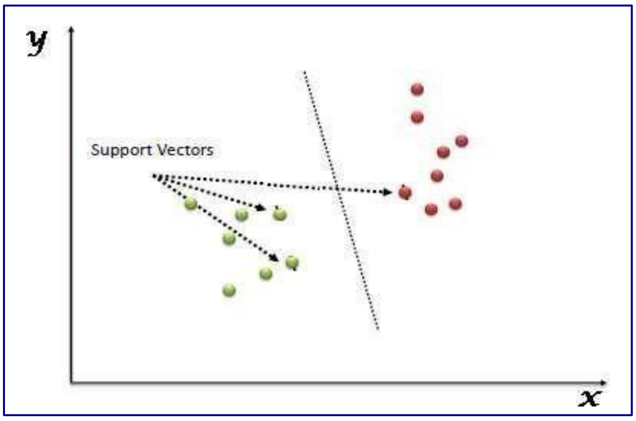

Fig2:- SVM Architure

5 RESULTS AND DISCUSSION

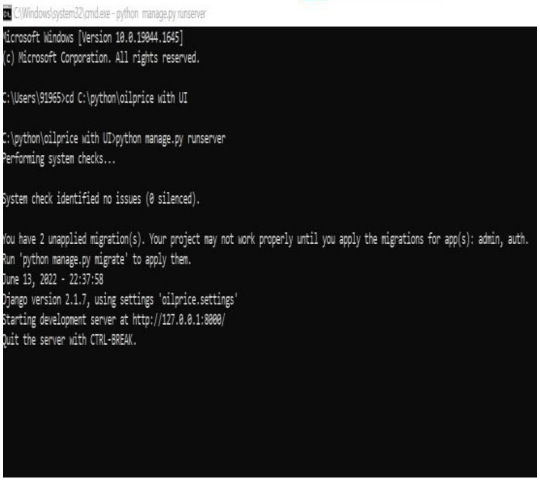

Fig3:executing using flask server

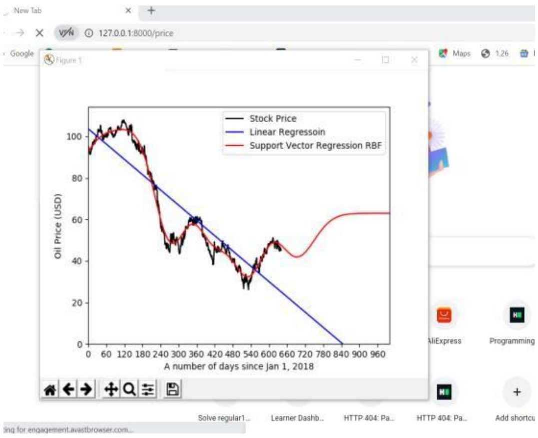

Fig4: prediction graph or price

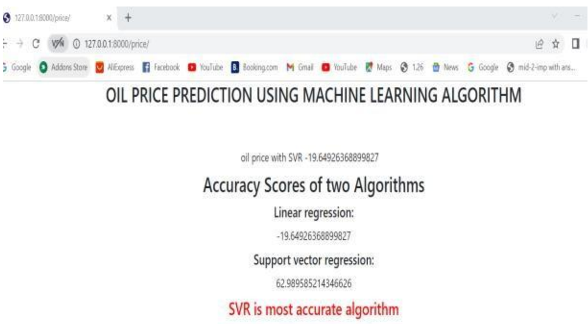

Fig5: accuracy score of two algorithms

6. CONCLUSION AND FUTURE WORK

We have implemented SVR Algorithm (Support Vector Regression) of Machine learning using Python. The predictions are most approximate with SVR Algorithms as they Linear or Gaussian. The algorithm automatically uses the kernel function that is most appropriate to the data. SVM uses the linear kernel when there are many attributes (more than 100) in the training data, otherwise it uses the Gaussian kernel. In the proposed system we have takes taken the datasets which has the price and days based on the dataset we have made feature list and target list where the target_list is price values and feature list is the days. After the analysis of data is done we have fitted both feature list and target list using Python Machine learning SVN Algorithm and predicted the values for 1,30 and 365 days from the last day of the dataset values. Finally we have plotted a graph based on the results from the predicted analysis done with SVN Algorithm.

7. REFRENCES

1. Oglend, A.; Asche, F. Cyclical non-stationarity in commodity prices. Empir. Econ. 2016, 51, 1465–1479. [CrossRef]

2. Lasheras, F.S.; de Cos Juez, F.J.; Sánchez, A.S.; Krzemie ´n, A.; Fernández, P.R. Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resour. Policy 2015, 45, 37–43. [CrossRef]

3. Ebert, L.; Menza, T.L. Chile, copper and resource revenue: A holistic approach to assessing commodity dependence. Resour. Policy 2015, 43, 101–111. [CrossRef]

4. Spilimbergo, A. Copper and the Chilean Economy, 1960–1998. Policy Reform 2002, 5, 115–126. [CrossRef]

5. Alameer, Z.; Abd Elaziz, M.; Ewees, A.A.; Ye, H.; Jianhua, Z. Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. Nat. Resour. Res. 2019, 28, 1385–1401. [CrossRef]

6. Hu, Y.; Ni, J.; Wen, L. A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction. Phys. A Stat. Mech. Appl. 2020, 557, 124907. [CrossRef]

7. García, D.; Kristjanpoller, W. An adaptive

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

forecasting approach for copper price volatility through hybrid and non-hybrid models. Appl. Soft Comput. 2019, 74, 466–478. [CrossRef]

8. Wang, L.; Zhang, Z. Research on Shanghai Copper Futures Price Forecast Based on X12-ARIMA-GARCH Family Models. In Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 17–19 April 2020; pp. 304–308.

9. Rubaszek, M.; Karolak, Z.; Kwas, M. Mean-reversion, non-linearities and the dynamics of industrial metal prices. A forecasting perspective. Resour. Policy 2020, 65, 101538. [CrossRef]

10. Vapnik, V. The Nature of Statistical Learning Theory; Springer: New York, NY, USA, 1995; p. 188. [CrossRef] Vol 12, Issue 08, August/2021 ISSN NO:0377-9254