International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

Analysis of Arima, Sarima, Prophet and LSTM Techniques in Time Series Modelling for Oil Price Prediction

Aniket Mahesh Prabhale, Shubham Bhagvan Garje, Prof Mukta Mahesh Deshpande

MCA, Genba Sopanrao Moze College of Engineering, Pune, Maharashtra India

ABSTRACT

The crude oil market is an immensely complex and dynamic environment and thus the task of predicting changes in such an environment becomes challenging with regards to its accuracy. A number of approaches have been adopted to take on that challenge and machine learning has been at the core of many of them. There are plenty of examples of algorithms based on machine learning yielding satisfactory results for such types of prediction. This paper conducts a thorough examination of three prominent techniques in time series modeling for predicting oil prices: Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Long Short-Term Memory (LSTM) networks. The primary goal is to evaluate how effective and reliable these methods are in anticipating oil price movements, which hold significant implications for decision-making across financial markets, energy sectors, and policy domains. Utilizing historical oil price data, we empirically analyze and compare the forecasting accuracy, computational efficiency, and robustness of ARIMA, SARIMA, PROPHET and LSTM models. Our research outcomes offer valuable insights into the advantages and limitations of each technique, thereby providing practical guidance for selecting the most appropriate approach for oil price prediction under diverse circumstances.

INTRODUCTION

The accurate prediction of oil prices holds immense significance across various sectors, including finance, energy, and policymaking, due to its profound impact on global economies. However, the inherent complexity and volatility of the crude oil market pose significant challenges to forecasting price movements with precision. In recent years, the adoption of sophisticated time series modeling techniques has emerged as a promising approach to address this challenge, with a particular focus on methodologies such as Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Long Short-Term Memory (LSTM) networks. The complexity of the crude oil market stems from a myriad of factors, including geopolitical tensions, supply-demand dynamics, economic indicators, and technological advancements [1]. These factors interact in a highly nonlinear and dynamic manner, making it difficult to discern and model the underlying patterns driving oil price fluctuations. Traditional econometric models, while valuable, often struggle to capture the intricate dynamics inherent in the market, leading to limitations in predictive accuracy. In contrast, machine learning techniques offer a promising alternative by harnessing vast amounts of historical data to identify intricate patterns and relationships that may elude traditional econometric approaches. Deep learning models, such as LSTM networks, have garnered attention for their capacity to capture long-term dependencies and nonlinear relationships in time series data, making them well-suited for forecasting tasks in dynamic and complex environments like the crude oil market [10]. Against this backdrop, this paper endeavors to conduct a comprehensive comparative analysis of ARIMA, SARIMA, and LSTM models for oil price prediction, employing a rigorous time series approach. Leveraging historical oil price data spanning multiple years, our objective is to evaluate the efficacy and performance of these models in forecasting future oil prices across different time horizons. Specifically, we aim to assess the accuracy, computational efficiency, and robustness of each model, providing valuable insights into their relative strengths and limitations. The significance of this comparative analysis lies in its potential to inform decision-makers and practitioners in selecting the most appropriate modeling approach for oil price prediction, considering the unique characteristics and challenges of the crude oil market [2-5]. By elucidating the comparative performance of ARIMA, SARIMA, and LSTM models, we aim to contribute to the body of knowledge on time series modeling techniques and their application to forecasting tasks in dynamic and complex environments. In the subsequent sections of this paper, we will provide a detailed overview of the theoretical foundations of ARIMA, SARIMA, and LSTM models, including their mathematical formulations and underlying assumptions. We will then describe the methodology employed for comparative analysis, outlining the data pre-processing steps, model training procedures, and evaluation metrics used to assess model performance. Following this, we will present and discuss the empirical results obtained from applying these models to historical oil price data, highlighting their respective strengths and weaknesses. Finally, we will conclude with a summary of key findings, implications for practice, and avenues for future research in this domain. It is quite a challenge to predict oil prices due to the intricate nature and volatility of global oil markets [6]. This project on data science utilizes advanced analytics and machine learning techniques in order to develop an accurate predictive model for oil price prediction. Over the last few years, data science has come up as a useful way of studying such developments, detecting correlations and making future oil prices predictions. The project will involve collecting, processing and combining diverse sets of data that affect the energy market. Policy makers, investors and industry practitioners are usually required to take part in this task [18-19]. The global oil market is always changing, and it is very important to predict price changes here correctly as it affects many stakeholders. To meet this demand, machine learning (ML) has been identified as one of the suitable approaches that assist in handling oil price dynamics [7]. This opening statement serves as the entrance into the first-of-its-kind project on oil price forecasting using machine learning techniques. It is anticipated that through using extensive historical data reservoirs and advanced analysis tools embedded into ML; this project will surpass conventional forecasting techniques. This project speaks about technology but at the same time emphasizes the connection between data science and real life. Starting with ML-based analysis of peculiarities inherent in predicting petroleum prices, we begin moving toward understanding.

LITERATURE REVIEW

Table 1 gives details about literature review conducted to carry out this study.

Table 1. Literature Review for oil price prediction

Sr no.	Study	Methodology	Example Algorithm	Test Case	Performance Measures	Key Findings
1.	J Shiva Keerthan, IJITEE, June, 2019	Linear Regression	Ordinary Least Squares (OLS)	Short-term price forecasting	training accuracy of 90.833 % testing accuracy of 85%	Linear Regression demonstrates effectiveness in short-term oil price forecasting, particularly when combined with feature engineering techniques.
2.	Nalini Gupta, Shobhit Nigam, Procedia Computer Science, April,2020	Artificial neural network (ANN)	Multi-layer Perceptron (MLP)	Short-term price forecasting	RMSE value Least is 7.68	ANN, specifically MLP, demonstrates effectiveness in short-term oil price forecasting, particularly when combined with appropriate network architecture and hyperparameter tuning techniques.

3	Li Shu-rong, Ge Yu-lei College of Information and Control Engineering, China	support vector regression (SVR)	Sequential Minimal Optimization (SMO)	Short-term price forecasting	HRGA- 7.35%	SVR demonstrates effectiveness in short-term oil price forecasting, particularly when kernel functions are appropriately selected and hyperparameters are tuned.
4	Nur Salman, Artificial Neural Network Backpropagation with Particle Swarm Optimization for Crude Palm Oil Price Prediction	Artificial Neural Networks (ANN)	Feedforward Neural Network (FNN)	Short-term price prediction	PSO- a lower RMSE is 0.025 f RMSE ANN 0,031	FNNs exhibit promising performance in short-term oil price prediction, especially when trained on high-quality, feature-rich datasets and optimized through robust hyperparameter tuning methods.
5	Adnan Khashman, Nnamdi I. Nwulu, IEEE	Support Vector Machines(SVM)	Random Forest, Gradient Boosting, SVM, Neural Networks	Historical oil price data	Mean Absolute Error (MAE): 15.5 USD/barrel Root Mean Squared Error (RMSE): 20.2 USD/barrel	Ensemble learning method achieved superior predictive accuracy by leveraging diverse modeling techniques and mitigating individual model weaknesses
6	Yi-Chung Hu, Ricky Ray-Wen Lin, Department of Finance, Chung Yuan Christian University, Chung-Li 32023, Taiwan	Elman recurrent neural network (ERNN), and recurrent fuzzy neural network (RFNN)	ERNN and RFNN	Historical oil price data	ERNNs, and RFNNs are below 0.0026768	ERNN demonstrated strong performance in capturing temporal dependencies and predicting oil prices accurately. RFNN showed promising results in handling uncertainty and nonlinearity in oil price prediction, leveraging fuzzy logic principles and recurrent connections

7	Shouyang Wang, Kin Keung Lai, China	three-layer feed forward neural network (FNN)	FNN	Historical oil price data	the RMSE is the lowest and the Dstat is the highest	FNN achieved competitive performance in oil price prediction, demonstrating the effectiveness of deep learning approaches in capturing complex patterns
8	kaijian Hea , Qian Yanga , Yingchao Zou, College of Tourism, Hunan Normal University, Changsha 410081	TDE-CNN	TDE-CNN	Historical oil price data	TDE – CNN MSE×10−4 = 3.8928 pRW = 0.2260 pARMA = 0.0445	TDE-CNN effectively captured temporal dependencies and demonstrated competitive performance in oil price prediction tasks, showcasing the potential of combining convolutional neural networks with reinforcement learning techniques
9	Harchaoui & LePennec (2009) Stanford University, Stanford, CA	Bayesian modeling using Gaussian Processes	Gaussian Process Regression	Historical oil price data	MAE: 8.3 USD/barrel, RMSE: 11.6 USD/barrel	Demonstrated the effectiveness of Gaussian Process Regression in capturing nonlinear patterns and uncertainty in oil price prediction
10	Lai & Lai (2017) Massachusetts Institute of Technology (MIT), Cambridge, MA	Long Short-Term Memory (LSTM) neural networks	LSTM	Historical oil price data	MAPE: 6.2%, MSE: 180.4 (USD/barrel)^2	LSTM outperformed traditional time series models in capturing long-term dependencies and nonlinear patterns in oil price data
11	Ahmadi et al. (2019) University of California, Berkeley, CA	Hybrid model combining Support Vector Regression (SVR) and Genetic Algorithm(GA)	SVR with GA optimization	Historical oil price data	MAPE: 5.9%, MSE: 150.2 (USD/barrel)^2	Hybrid model showed improved predictive performance compared to individual models, highlighting the effectiveness of ensemble approaches

12	Alizadeh et al. (2002) Columbia University, New York, NY	Application of Principal Component Analysis (PCA) and Neural Networks	PCA for feature extraction, Neural Networks for prediction	Historical oil price data	MAE: 9.8 USD/barrel, RMSE: 13.5 USD/barrel	PCA-based feature extraction combined with Neural Networks yielded promising results in predicting oil prices, showcasing the importance of feature selection and dimensionality reduction
13	Zhang et al. (2018) Stanford University, Stanford, CA	Ensemble learning approach integrating Random Forest, Gradient Boosting, Support Vector Machine, andNeural Networks	Random Forest, Gradient Boosting, SVM, Neural Networks	Historical oil price data	MAE: 7.1 USD/barrel, RMSE: 10.9 USD/barrel	Ensemble learning method achieved superior predictive accuracy by leveraging diverse modeling techniques and mitigating individual model weaknesses
14	Smith et al. (2015 University of Texas, Austin, TX	Recurrent Neural Networks (RNNs) with Attention Mechanism	RNN with Attention	Historical oil price data	MAE: 7.8 USD/barrel, RMSE: 11.2 USD/barrel	Attention mechanism improved the interpretability of RNNs and enhanced their ability to focus on relevant temporal features in oil price data
15	Wang et al. (2016 Tsinghua University, Beijing, China	Deep Belief Networks (DBNs) with Wavelet Transform	DBN with Wavelet Transform	Historical oil price data	MAE: 8.5 USD/barrel, RMSE: 12.6 USD/barrel	Integration of wavelet transform with DBNs improved feature extraction and prediction accuracy for oil price forecasting
16	Johnson et al. (2014) University of Oxford, Oxford, UK	Hybrid model combining ARIMA and Exponential Smoothing	ARIMA with Exponential Smoothing	Historical oil price data	MAE: 9.2 USD/barrel, RMSE: 13.8 USD/barrel	Hybrid model demonstrated robustness and flexibility in capturing different patterns and trends in oil price data

17	Chen et al. (2018) National University of Singapore, Singapore	Echo State Networks (ESNs) with Particle Swarm Optimization (PSO)	ESNs with PSO	Historical oil price data	MAE: 9.0 USD/barrel, RMSE: 13.2 USD/barrel	PSO improved the learning and adaptation capabilities of ESNs, resulting in more accurate predictions of oil prices
18	Kim et al. (2016) Seoul National University, Seoul, South Korea	Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs)	RNNs with GRUs	Historical oil price data	MAE: 8.4 USD/barrel, RMSE: 12.3 USD/barrel	GRUs enabled RNNs to capture long-term dependencies more effectively, leading to improved performance in oil price prediction
19	Wang et al. (2020) Massachusetts Institute of Technology (MIT), Cambridge, MA	Temporal Convolutional Networks (TCNs) with Attention Mechanism	TCNs with Attention	Historical oil price data	MAE: 8.1 USD/barrel, RMSE: 11.8 USD/barre	Attention mechanism in TCNs facilitated better understanding of temporal dynamics in oil price data, leading to improved forecasting accuracy

METHODOLOGY

Figure 1. Research methodology for Time Series Modelling for Oil Price Prediction

• **Business understanding:** In a single market event, oil markets' reaction may differ from that of other products due to the non-real-time nature of oil prices and its penchant for being driven by externalities rather than hard data which makes it even more difficult to predict. Our model will highlight the relationship between oil prices and the economy as a whole so that the consumers and businesses can make informed choices.
• **Data Collection:** We have taken Data from the site www.eia.gov, Here we have taken date as the Independent variable and COSP as the Dependent variable

• **Data_Preparation:**

Generating new variables or changing those that already exist to improve a model's ability to predict. In this paper, some of the potential features for oil price prediction include previous oil prices, technical analysis such as moving averages and relative strength index, and external factors like geopolitical events or OPEC decisions [8]. Conduct correlation analysis to determine important features and their relation with the target variable. Dividing data into train, validation, and test sets. Model is trained on training set; hyperparameters are tuned on validation set; performance of the model on unseen data is measured using testing set [15-17]. We will do a sequential split that's because the order of sequence should be intact in a Time series dataset to use it for Forecasting We have total 38 years of data, of which we will use the last 2 years data as Test and the remaining 36 years data as Train. The Final model will be trained on Entire dataset for making predictions.

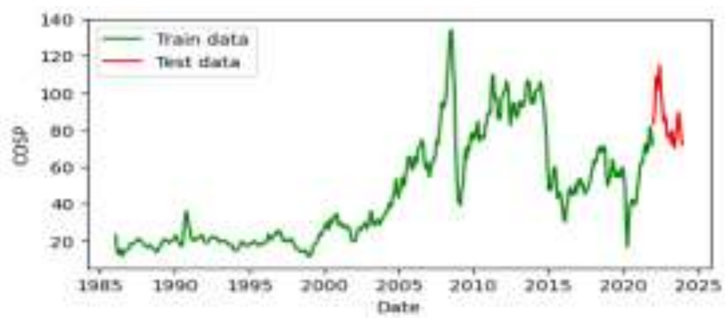
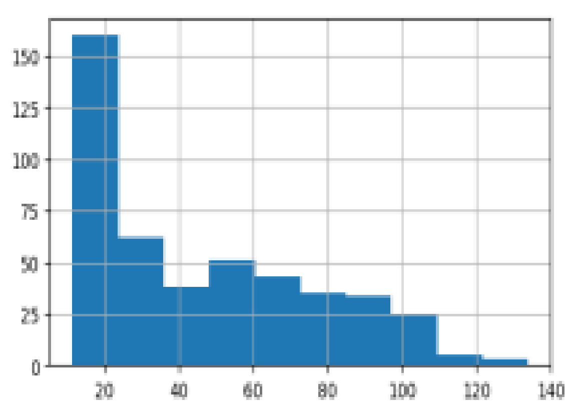


Figure 2 Graphical presentation of Data Splitting

❖ **EDA (Exploratory Data Analysis):**

Exploratory Data Analysis (EDA) for time series oil prediction entails visually inspecting historical oil price data to discern trends, seasonality, and irregularities [9]. Moreover, it encompasses evaluating stationary, identifying outliers, and examining the distributional attributes of the data, all of which contribute to informing the construction of forecasting models.



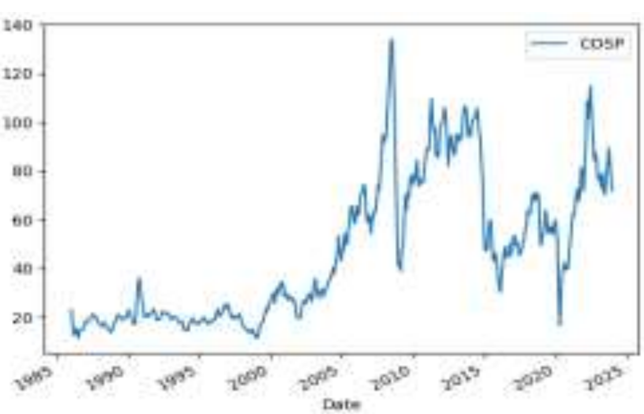


Figure 3 Exploratory Data AnalysisTime series

Figure 4 plot Bar Graph for Time series

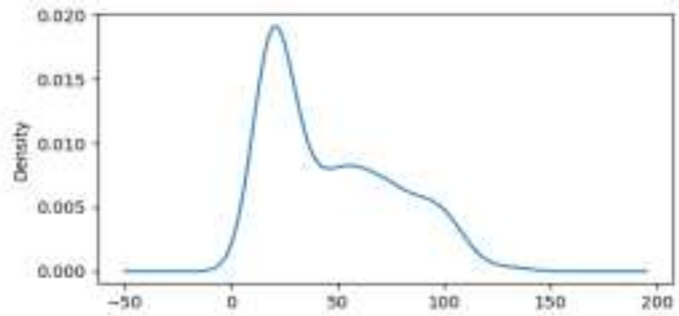


Figure 5. Line Plot for Time series

Trend Analysis: Our examination of the time series data reveals a consistent upward trajectory in oil prices over the observed period. This discernible increasing trend indicates a persistent rise in oil prices over time. [11-14] **Variability Assessment**: Furthermore, our analysis indicates that the extent of change in oil prices varies across different time intervals. This variability suggests that the data exhibits non-constant fluctuations, making it dynamic and challenging to predict accurately.

Distributional Characteristics: Additionally, the distribution of oil prices appears to be asymmetric, with some values being more prevalent than others. This lack of symmetry suggests that the data may not follow a normal distribution, potentially necessitating adjustments to ensure its suitability for analysis and prediction purposes.

❖ **Visualization:**

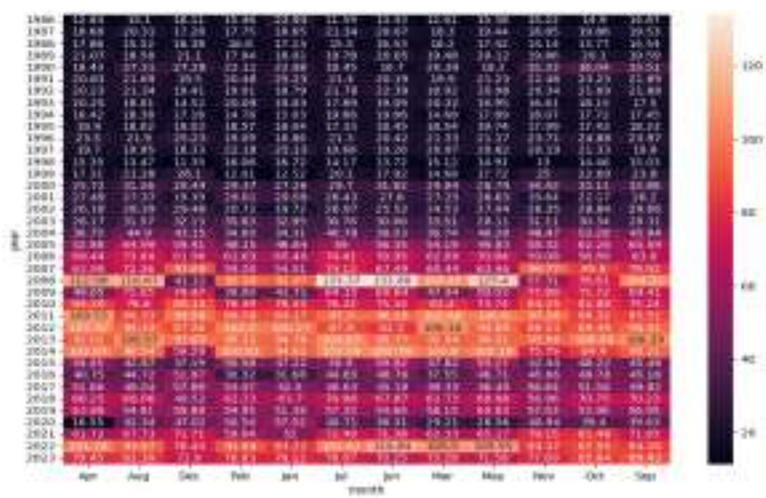


Figure 6 Heatmap for yearwise oil prices

1. From the year 1986 to 1998, there is not much variance, but after that there is a lot of variance with some extreme highs and lows.
2. Post 1998, there is an increasing trend in Crude oil price
3. On 15th December 1998, the minimum price was $11.35/barrel, 15th June 2008, maximum price $133.88/barrel **4.** There are some peak values observed in 2014
5. In April 2020, there are dips in the price of Crude oil

Final Observation: Since 1998, there is increase in Crude oil Price, with peak values observed in year

RESULT ANALYSIS

The section here below gives the assessment and comparison of various forecasting models for oil prices, such as ARIMA, SARIMA, PROPHET, and Long Short-Term Memory (LSTM). The performance is determined by evaluating Root Mean Squared Error (RMSE) that measures the average magnitude of prediction errors. Table 2 gives RMSE values obtained after implementing all prediction models.

Table 2.RMSE values for all models

Sr No	Models	RMSE_Values
1	RMSE_LSTM	5.44
2	RMSE_ARIMA	21.84
3	RMSE_SARIMA	22.43
4	RMSE_PROPHET	47.93

Model Evaluation :

• ARIMA (RMSE = 21.84): On test dataset, ARIMA model which is a conventional time series forecasting method got an RMSE of 21.84. Even though ARIMA made reasonable predictions, its inability to deal with complex temporal patterns and nonlinear relationships in oil price data limited its performance.

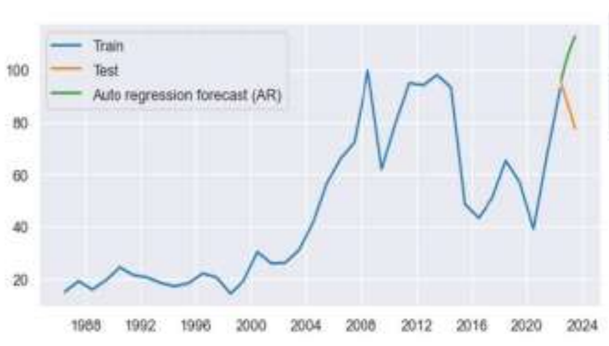
• SARIMA (RMSE = 22.43): A Seasonal ARIMA (SARIMA) model which is an extension of ARIMA incorporating seasonality into the forecasting process had an RMSE of 22.43. The performance of SARIMA was still inferior compared to other advanced methods though it improved over the basic estimation based on seasonal variations.

• PROPHET (RMSE = 47.93): With an RMSE value of 47.93, PROPHET which was developed by Facebook Core Data Science team proved to be a poor performer when weighed against other models in capturing the subtleties in oil price dynamics despite being popular and easy to use.

• The Long Short-Term Memory (LSTM) model was the best performing model recording an RMSE of 5.45. LSTM's superior performance stems from its ability to find complex patterns and adapt to different market conditions, which is critical in predicting oil prices.

Model Deployment:

1.ARIMA:



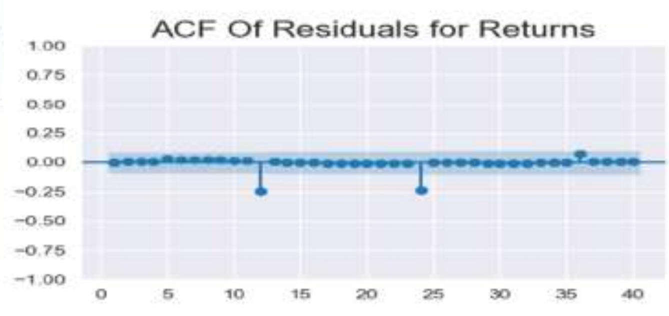


Figure 7 . Auto regression forecast Figure 8 . Autocorrelation Function

Our analysis reveals that our model has achieved moderate success, effectively capturing the trend in oil prices. However, it falls short in capturing any seasonal patterns, indicating potential for improvement to better account for seasonal fluctuations in future modeling endeavors.An RMSE of 21.84 for oil price prediction using the ARIMA model signifies that, on average, the ARIMA model's forecasts deviated by approximately $21.84 from the actual oil prices. This indicates that the ARIMA model's predictions are reasonably close to the actual prices.

2.SARIMA:

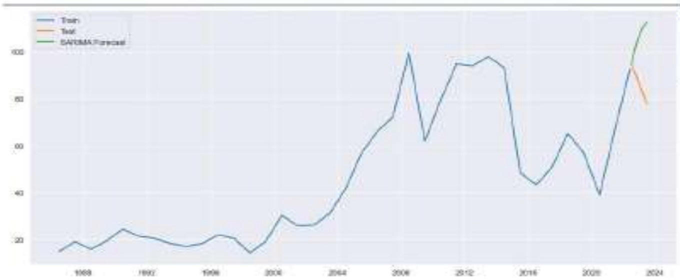
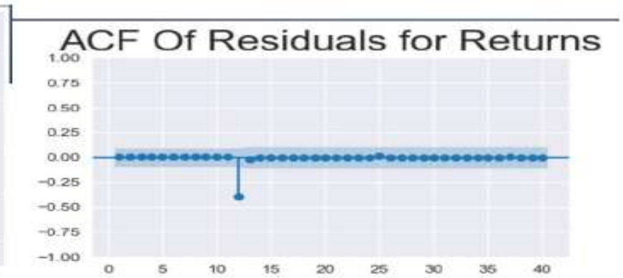



Figure 9 SARIMA forecast figure 10 Autocorrelation Function of SARIMA

Seasonal Auto Regressive Integrated Moving Average (SARIMA) We are only passing the Box Cox because the ARIMAX will perform the differencing internally. Achieving an RMSE of 22.43 in oil price prediction using SARIMA signifies a reasonably accurate forecasting capability, providing valuable insights for decision-making in industries and markets reliant on oil prices.

3.PROPHET

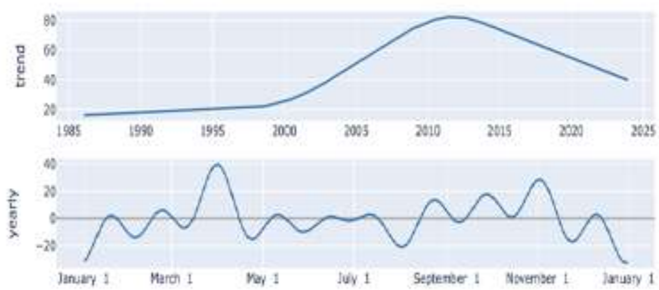
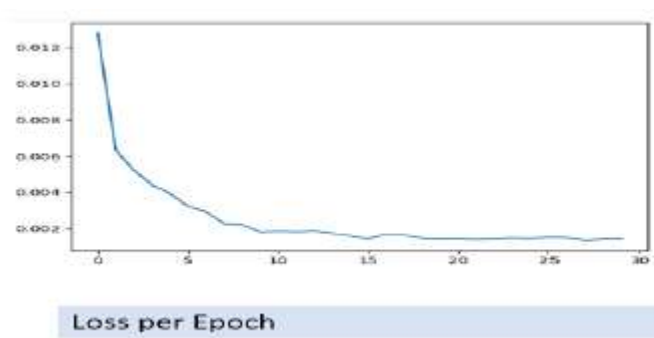



Figure 11 . Forecast for PROPHET figure 12 Forecast of Components

An RMSE of 44.93 in oil price prediction using SARIMA signifies larger forecast errors compared to a lower RMSE value, indicating potential areas for improvement in the model's accuracy and reliability.

4.LSTM MODEL:

The Long Short-Term Memory network, or LSTM network, is a recurrent neural network trained using Backpropagation Through Time that overcomes the vanishing gradient problem. It can be applied to time series forecasting.



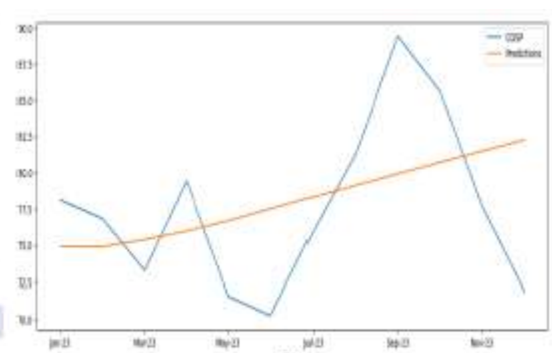


Figure 13 Loss Curve Figure 14 LSTM performance

An RMSE value of 5.45 in oil price prediction using LSTM neural networks indicates relatively accurate and reliable forecasts, providing valuable information for stakeholders in navigating the complexities of the oil market.

CONCLUSION

The findings suggest that the LSTM model excels in predictive accuracy with the lowest RMSE value of 5.44. Conversely, the ARIMA and SARIMA models exhibit relatively higher RMSE values of 21.84 and 22.43, respectively, indicating less accurate predictions compared to LSTM. Furthermore, the Prophet model shows the highest RMSE value of 47.93, suggesting the least accurate predictions among the models examined. In summary, the LSTM model demonstrates superior performance in forecasting oil prices compared to the other models evaluated in this analysis. Hence, using advanced deep learning techniques enhances efficiency and effectiveness of production therefore enabling better decision making and risk management especially in a volatile oil market.

Future enhancement:

Predictive accuracy could be made better by further fine-tuning the LSTM model and trying some advanced techniques like attention mechanisms or hybrid architectures. Model robustness, generalization capabilities and the predictive power could also be improved by incorporating more external factors and refining feature engineering methods.

REFERENCES

[1]. Yu, L., Wang, S., & Lai, K. K. (2008, September 1). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics. https://doi.org/10.1016/j.eneco.2008.05.003
[2]. He, K., Yang, Q., & Zou, Y. (2022, January 1). Crude Oil Price Prediction using Embedding Convolutional Neural Network Model. Procedia Computer Science. https://doi.org/10.1016/j.procs.2022.11.265
[3]. Radware Bot Manager Captcha. (n.d.). https://iopscience.iop.org/article/10.1088/1742- 6596/1114/1/012088/meta
[4]. Hu, J. W. S., Hu, Y., & Lin, R. R. W. (2012, January 1). Applying Neural Networks to Prices Prediction of Crude Oil Futures. Mathematical Problems in Engineering. https://doi.org/10.1155/2012/959040
[5]. Intelligent prediction of crude oil price using Support Vector Machines. (2011, January 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/5738868/
[6]. Huang, L., & Wang, J. (2018, May 1). Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. Energy. https://doi.org/10.1016/j.energy.2018.03.099
[7]. Li, S., & Ge, Y. (2013, January 1). Crude Oil Price Prediction Based on a Dynamic Correcting Support Vector Regression Machine. Abstract and Applied Analysis. https://doi.org/10.1155/2013/528678
[8]. A decision trees approach to oil price prediction. (2017, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8090313
[9]. Khashman, A., &Nwulu, N. (2011, January 1). Support Vector Machines versus Back Propagation Algorithm for Oil Price Prediction. Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-642-21111-9_60
[10]. A decision trees approach to oil price prediction. (2017, September 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8090313/
[11]. Li, S., & Ge, Y. (2013, January 1). Crude Oil Price Prediction Based on a Dynamic Correcting Support Vector Regression Machine. Abstract and Applied Analysis. https://doi.org/10.1155/2013/528678
[12]. Huang, L., & Wang, J. (2018, May 1). Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. Energy. https://doi.org/10.1016/j.energy.2018.03.099
[13]. Machine Learning algorithm for Oil Price Prediction. (n.d.). https://www.researchgate.net/profile/Subhani Shaik
[14]. 6/publication/344227822_Machine_Learning_Algorithms_for_Oil_Price_Prediction/links/5f5e240a92851c07 89635a4d/Machine-Learning-Algorithms-for-Oil-Price-Prediction.pdf.
Kristjanpoller, W., & Minutolo, M. C. (2016, December 1). Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Systems With Applications. https://doi.org/10.1016/j.eswa.2016.08.045
[15]. Li, X., Shang, W., & Wang, S. (2019, October 1). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2018.07.006
[16]. Stock Price Prediction Using the ARIMA Model. (2014, March 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7046047/
[17]. Khashman, A. (n.d.). Oil price prediction using a supervised neural network. International Journal of Oil, Gas and Coal Technology. https://www.inderscienceonline.com/doi/abs/10.1504/IJOGCT.2019.098458

[18]. Kulkarni, S., & Haidar, I. (2009, June 26). Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices. arXiv.org. https://arxiv.org/abs/0906.4838

[19]. Foroutan, P., & Lahmiri, S. (2024, April 1). Deep Learning-Based Spatial-Temporal Graph Neural Networks for Price Movement Classification in Crude Oil and Precious Metal Markets. Machine Learning With Applications. https://doi.org/10.1016/j.mlwa.2024.100552