

Crude Oil Trend Prediction During COVID-19: Machine Learning with Randomized Search and Bayesian Optimization

Seyyed Soroosh Firoozabadi¹, Mehdi Ansari^{2,*}, and Farhad Vasheghanifarahani³

ABSTRACT

This study delves into an innovative research framework aimed at enhancing the precision of crude oil return rate predictions. The study, which holds significant implications for financial institutions, investors, central banks, and corporations operating in volatile markets, rigorously evaluates the performance of three advanced machine learning models—LSTM, XGBoost, and SVM. Leveraging optimization and cross-validation techniques, the research particularly focuses on refining forecasting accuracy amidst the challenges posed by the COVID-19 epidemic. This study explores randomized search and Bayesian optimization, providing a comprehensive understanding of their application in the context of improving model performance and decision-making in the dynamic crude oil market. The findings indicate the accuracy of models with different evaluation metrics and reveal that the SVM demonstrates superior accuracy in regression analysis during the pandemic.

Keywords: COVID-19, Machine learning, Optimization, WTI prediction.

Submitted: November 24, 2023

Published: May 20, 2024

 10.24018/ejbm.2024.9.3.2221

¹Financial Engineering Group, Industrial-land Systems Engineering Faculty, Tarbiat Modares University, Iran.

²Department of Industrial Engineering and Management, College of Engineering, Oklahoma State University, United States.

³Department of Economics, College of Arts and Sciences, University of South Florida, United States.

*Corresponding Author:
e-mail: meansar@okstate.edu

1. INTRODUCTION

Forecasting in the financial time series (FTS) is a crucial area in finance and economics that predicts potential risks, market trends, and investment timing in financial markets. This subject has been studied for decades due to the uncertain and noisy nature of the financial environment. In order to analyze and predict financial market behavior, fundamental and technical analysis are commonly used.

Fundamental and technical analysis was widely used to analyze and forecast financial market behavior prior to the introduction of natural language processing (NLP) models. Fundamental analysis entails examining financial metrics and indicators to determine a company's financial health and growth potential, whereas technical analysis employs historical price and volume data to identify trends and forecast future market behavior. Both methods have benefits and drawbacks and can be used together or separately, depending on an investor's goals and investment style. Predicting financial market behavior, on the other hand, is difficult and is influenced by a number of factors, such as economic indicators, geopolitical events, and investor sentiment. As a result, when making investment

decisions, investors should consider multiple sources of information and analysis.

Crude oil prices, specifically West Texas Intermediate (WTI), have historically experienced significant volatility due to various economic, geopolitical, and environmental factors. One of the notable periods of price shocks was the global financial crisis of 2008–2009, which had a profound impact on the oil market. During this period, the financial crisis led to reduced oil demand due to economic slowdowns worldwide, resulting in a sharp decline in WTI crude oil prices. In addition to the 2008–2009 financial crisis, other price shocks have affected crude oil prices, such as the price collapse in 2014–2015. This collapse was driven by an oversupply of crude oil in the global market, with increased production, particularly from shale oil in the United States and non-OPEC countries. The decision by OPEC not to cut production levels further exacerbated the oversupply situation, leading to a significant drop in oil prices.

Furthermore, the year 2020 presented another major challenge for the crude oil market due to the unprecedented COVID-19 pandemic. Lockdowns, travel restrictions, and economic slowdowns worldwide drastically reduced oil demand. The price war between OPEC



and Russia, along with the supply glut, added further pressure on crude oil prices during this time.

Considering the recurring price shocks, accurate forecasting models are important for understanding crude oil price behavior. The objective of this project is to develop a predictive model for forecasting Texas crude oil prices using advanced machine learning algorithms. We aim to leverage historical data to create accurate and reliable predictions. The project aims to assist stakeholders in the oil industry, financial markets, governments, and energy-dependent sectors to make informed decisions and develop strategies in response to oil price shocks in recent years during the pandemic.

This study uses three machine learning models, including Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Long-Short-Term Memory (LSTM). The performance of the models is evaluated using various metrics to determine the effectiveness of the approaches in predicting the behavior of the WTI. The following is how the paper is structured. The related work is described in the following section. The proposed methodology, evaluation metric, and data set are all covered in Section 3. The experimental setup and results are shown in Section 4. Section 5 finally exposes our conclusions and future work.

2. RELATED WORKS

Machine learning (ML) is a category of data science models that have the ability to learn and enhance their performance through data analysis. The origins of ML can be traced back to the scientific community's fascination in the 1950s and 1960s with simulating human learning using computer programs. In this context, ML involves extracting knowledge from data, which can then be applied for prediction and generating new insights. This knowledge reduces uncertainty by providing guidance on solving specific problems. ML is especially valuable for tasks that can't be explicitly addressed using analytical solutions, such as tasks involving image and voice processing, pattern recognition, or intricate classification problems. Machine learning (ML) has found extensive applications in economic and financial analyses of energy markets, including tasks like price prediction and risk management. When comparing ML to traditional econometric models like ARIMA and GARCH, one person can identify some key factors contributing to the growing adoption of ML in energy economics. A significant advantage of ML methods over classical statistical and econometric approaches is their ability to handle vast amounts of structured and unstructured data, enabling swift decision-making and forecasting (Ghoddusi et al., 2019).

Ghoddusi et al. (2019) reviewed the papers, including the methodologies and findings, of over 130 articles released from 2005 to 2018, stating that the most prevalent applications in energy economics and finance involve forecasting crude oil and power prices. They state that in terms of methodologies, artificial neural networks (ANN) have historically been extensively employed, with recent attention shifting toward support vector machines (SVM).

Moshiri and Foroutan (2006) focus on the challenge of forecasting daily crude oil futures prices from 1983 to 2003, listed on NYMEX. The authors recognize the complexity of oil price movements and the limitations of traditional linear models. They explore the possibility of nonlinear data-generating processes underlying crude oil futures prices and apply various tests for nonlinearity and chaos to confirm this hypothesis. Their tests suggest that crude oil futures prices follow a complex nonlinear dynamic process. Given the nonlinearity and complexity of the data, the authors introduce a flexible Artificial Neural Network (ANN) model for forecasting. They compare the performance of the ANN model with traditional ARIMA and GARCH models and find that the ANN model outperforms them. Abdidizaji and Pakizeh (2021) tried to find statistical arbitrage opportunities in the stock market to predict hidden arbitrage in the prices of oil companies and other big ones using statistical methods.

Yu et al. (2008) present an innovative approach for forecasting world crude oil spot prices. The authors employ Empirical Mode Decomposition (EMD) to decompose the original crude oil spot price series into Intrinsic Mode Functions (IMFs), followed by using a three-layer feed-forward neural network model to forecast each IMF. These individual IMF predictions are then combined using an adaptive linear neural network (ALNN) to form an ensemble output for the original oil series. The study tests this methodology on two main crude oil price series: West Texas Intermediate (WTI) crude oil spot price and Brent crude oil spot price. The results indicate that the EMD-based neural network ensemble model outperforms other forecasting methods, including ARIMA models, feed-forward neural networks (FNN), and various ensemble strategies. The evaluation metrics RMSE (Root Mean Square Error) and Dstat reveal that the EMD-based neural network ensemble forecasting model excels in forecasting accuracy. This study suggests that this approach could serve as a viable alternative for predicting crude oil prices, benefiting investment managers and business practitioners.

Jammazi and Aloui (2012) tackle the formidable challenge of predicting oil prices, a task made intricate by the inherent complexity of the oil market. With the backdrop of the European financial crisis and oil shocks reigniting discussions on understanding oil price behavior, their study introduced the HTWMPNN model. This innovative approach fused multilayer backpropagation neural networks with the Harr A trous wavelet decomposition, resulting in more accurate crude oil price predictions. Distinct from a narrow focus on comparing various neural architectures and decomposition techniques to determine the best forecasting model, their work honed in on a crucial aspect: the selection of transfer functions to ensure the robustness of simulations. They explored three activation function variants—sigmoid, bipolar sigmoid, and hyperbolic tangent—to enhance model flexibility. By experimenting with various input-hidden node setups, HTW-MBPNN demonstrated its superiority over conventional BPNN models.

Sun et al. (2023) tackle the complex challenge of forecasting crude oil prices, a matter of immense interest to both investment firms and governmental bodies. The study

introduces an innovative hybrid forecasting model tailored for predicting trends in crude oil.

3. METHODOLOGY AND DATASET

Machine learning offers a wide array of methods for solving complex problems. In supervised learning, techniques like linear regression provide a foundational approach to predicting numerical outcomes, while logistic regression is employed when dealing with classification tasks. Decision trees and random forests are versatile algorithms used for both regression and classification, and they offer interpretability through their tree-like structures. Support Vector Machines (SVMs) are powerful tools for classification and regression, particularly when dealing with high-dimensional data. k-Nearest Neighbors (k-NN) is a simple yet effective algorithm for classification and regression, relying on the similarity between data points. Naive Bayes methods are great for text classification and other tasks involving probabilities.

Unsupervised learning methods include Principal Component Analysis (PCA), which is used for dimensionality reduction and data visualization, and k-means clustering, which helps group data into clusters based on similarity. Hierarchical clustering methods, such as Agglomerative and Divisive clustering, organize data hierarchically. Apriori and Eclat are employed in association rule learning to uncover patterns in data. Lastly, in reinforcement learning, Q-learning and Deep Q Networks (DQNs) are used to train agents to make sequential decisions and learn optimal strategies through trial and error. These machine learning methods provide the toolkit for a diverse range of applications across various domains.

Machine learning and deep learning are subsets of artificial intelligence. Machine learning encompasses various techniques where a system is trained to learn and make predictions from data without being explicitly programmed. It includes algorithms like decision trees, random forests, and support vector machines. Deep learning, on the other hand, is a more specialized form of machine learning that focuses on neural networks with multiple layers, known as deep neural networks.

Deep learning models, including Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs), are neural networks that can automatically learn and make predictions from data. ANNs are inspired by the human brain and are versatile for tasks like image and speech recognition. CNNs are designed for processing grid-like data, such as images, making them essential for computer vision. RNNs excel with sequence data, making them vital for tasks like time series forecasting and natural language processing. Among RNNs, LSTMs stand out for their ability to capture long-range dependencies, facilitating sequence prediction, while GRUs offer computational efficiency compared to LSTMs while maintaining strong performance on sequence-based tasks. These neural network models collectively enable deep learning's wide applicability in complex data analysis and pattern recognition, often outperforming traditional machine learning

techniques for tasks involving unstructured data like images, text, and audio.

Ensemble methods combine the predictions of multiple machine learning models to improve accuracy and robustness, including techniques like bagging and boosting. Reinforcement learning is a type of machine learning where agents learn how to make sequences of decisions by interacting with an environment and receiving rewards. XGBoost (Extreme Gradient Boosting) is primarily categorized as an ensemble method. It belongs to the ensemble learning category because it combines the predictions of multiple decision trees to create a stronger predictive model. In the context of our literature review, the performance of Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and XGBoost, three distinct predictive models, will be compared to determine which one provides more accurate predictions for crude oil returns during the pandemic-induced market shock.

3.1. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a powerful machine learning algorithm introduced by Vapnik (1999), designed to classify data points into different categories. SVM is particularly suited for both classification and regression tasks, making it versatile for various applications, including stock price prediction. Unlike traditional feed-forward neural networks, SVM is a supervised learning method that doesn't rely on complex neural architectures. SVM operates by identifying a decision boundary, known as a hyperplane, that maximizes the margin between data points of different classes. This margin, called the "support vector," acts as the backbone of the model. By selecting the support vectors, SVM effectively focuses on the most informative data points to make accurate predictions. The ability to find a high-dimensional hyperplane and handle non-linear data through kernel functions is a significant strength of SVM (Schölkopf & Smola, 2002).

One of the key advantages of SVM is its generalization power, which allows it to make predictions on unseen data accurately. SVM can efficiently handle high-dimensional data, and with appropriate parameter tuning, it's less prone to overfitting (Hsu et al., 2003). Furthermore, SVM can deal with non-linear relationships in the data by transforming it into a higher-dimensional space. However, SVM also has limitations, such as the need for careful selection of the kernel function, potential sensitivity to hyperparameters, and challenges with interpreting complex decision boundaries. Moreover, SVM might not perform optimally when the dataset is exceptionally large or noisy (Cristianini & Shawe-Taylor, 2000).

3.2. Long Short-Term Memory (LSTM)

Long-Short-Term Memory method (Hochreiter & Schmidhuber, 1997) is a type of recurrent neural network that is capable of modeling complex temporal patterns in sequential data; it is used mainly in NLP tasks such as next-word prediction (Norouzi et al., 2019). Unlike regular feed-forward neural networks, which only consider the current input, information in the RNN travels in

loops from layer to layer, preserving the context based on previous inputs and outputs (Elman, 1990). However, RNNs have some limitations, such as slow computation time and difficulty retaining information over long periods (Bengio et al., 1994). LSTM overcomes these shortcomings by using a cell to remember information over time intervals and three gates to regulate the flow of information into and out of the cell. The capacity to capture long-term dependencies, versatility in handling different forecasting jobs, and handling missing variables are all advantages of LSTM in forecasting. Disadvantages include complexity in training and optimization, difficulty in interpreting results, sensitivity to hyperparameters, and potential for overfitting (Goodfellow et al., 2016). The specific advantages and disadvantages may vary depending on the use case and dataset.

3.3. Extreme Gradient Boosting (XGBoost)

XGBoost, or Extreme Gradient Boosting, is a powerful ensemble learning algorithm introduced by Chen and Guestrin (2016) that has gained prominence for its exceptional performance in various machine learning competitions and predictive modeling tasks. XGBoost is particularly suitable for both classification and regression tasks, making it a valuable tool for stock price prediction and many other applications.

Extreme Gradient Boosting belongs to the gradient boosting family of algorithms. It builds predictive models by training an ensemble of decision trees sequentially. Each new tree corrects the errors made by the existing ensemble, allowing XGBoost to iteratively improve its predictive accuracy. XGBoost's key strengths include its ability to handle complex non-linear relationships in data, efficiently manage missing values, and reduce the risk of overfitting through regularization techniques (Chen & Guestrin, 2016). One of the distinct advantages of XGBoost is its speed and scalability. By optimizing for computational efficiency, XGBoost processes data faster than many other algorithms, which is essential for handling large datasets. Additionally, XGBoost provides feature importance scores, aiding in the interpretation of model predictions. However, like many machine learning methods, XGBoost also has limitations, including the need for fine-tuning of hyperparameters, sensitivity to outliers, and a potential increase in complexity when dealing with numerous features (Chen & Guestrin, 2016).

3.4. Tune Hyperparameters

Hyperparameters are the parameters of a given machine learning that cannot be tuned by the learning process. Instead, they must be adjusted before training. The accuracy of the machine learning algorithms highly depends on the hyperparameters, and tuning them is a critical step.

1) *Randomized Search*: Among the hyperparameters tuning techniques, randomized search has shown an excellent performance. From a specific distribution, the hyperparameters of a given algorithm are tuned randomly (Bergstra & Bengio, 2012). All steps of the randomized search can be summarized as follows:

- Specifying the hyperparameters and their corresponding distribution and ranges.

- Setting the number of iterations.
- Evaluate the accuracy of the model by the selected hyperparameters.
- Repeat the previous steps until the best model is found.

2) *Bayesian Optimization*: LSTM can perform better when its hyperparameters are optimized using the Bayesian optimization technique. It selects the most promising collection of hyperparameters to examine using a probabilistic model and iteratively changes its model as new data come in (Bergstra et al., 2011). Bayesian optimization can be used to adjust LSTM hyperparameters, including batch size, learning rate, learning rate per layer, and number of LSTM layers, among others. Bayesian optimization can help boost the LSTM model's accuracy and generalization performance in a specific forecasting task by identifying the best hyperparameters.

3.5. Cross-Validation

Cross-validation is a statistical technique used in machine learning and statistical modeling to assess how well a model generalizes to an independent dataset. The primary purpose is to detect and mitigate issues like overfitting, which occurs when a model performs well on the training data but poorly on new, unseen data. The general idea is to divide the dataset into multiple subsets, train the model on some of these subsets (training set), and then evaluate the model's performance on the remaining subsets (testing set). This process is repeated multiple times, and the average performance across all iterations is used as an estimate of how well the model is expected to perform on new data.

The distinction between cross-sectional data cross-validation and time series cross-validation is pivotal. cross-sectional data cross-validation, tailored for independent and identically distributed samples, encapsulates observations taken at a singular temporal point, treating each observation as an autonomous entity. Employing established techniques like k-fold cross-validation, this approach randomly partitions the dataset into subsets, enabling model training and testing iterations. In contrast, Time Series Cross-Validation meticulously upholds the chronological integrity of observations. The latter becomes particularly imperative in gauging a model's aptitude for generalization to forthcoming data points, acknowledging the potential evolution of data characteristics over time. There are two popular cross-validations for Time Series datasets:

1. *R Rolling basis Cross-Validation*: When analyzing time series data, where observations are arranged chronologically and may be connected with one another, common cross-validation procedures assume that data points are independently and identically distributed (Hyndman & Athanasopoulos, 2018). Utilizing a rolling window is one popular method for time series cross-validation, where the training set is made up of all data up to a specific time point, and the test set is made up of the following data window. Using rolling window cross-validation enables the model to consider how the



Fig. 1. Rolling basis time series data with three cross validations.

underlying patterns of the data change over time, which is crucial for time series forecasting. The cross-validation on a rolling basis starts from a small subset (see Fig. 1) or window of the train data and forecasts the next data point(s). The part of the forecasted data point(s) is transferred to the next window as a new train data to forecast the new data point.

2. *Split Cross-Validation*: Time Series Split Cross-Validation is a crucial method for evaluating the performance of predictive models when dealing with time-ordered data, like stock prices, weather data, or even text data over time. Unlike traditional cross-validation methods that randomly shuffle data, Time Series Split maintains the chronological sequence of data points.

It's vital because, in real-world applications, data points collected at one time often depend on, or at least correlate with, previous points. As a result, models need to be trained and tested with this temporal order in mind. This method simulates the real-world scenario of making predictions about the future based on historical data. By iteratively moving a testing window forward in time, it helps ensure that the model can genuinely generalize and make accurate predictions for unseen future data.

Time Series Split can be thought of as a sliding window approach. Imagine your dataset as a timeline from left to right, and you want to perform several iterations of model evaluation. In each iteration, the model is trained on past data and then tested on a slice of future data, ensuring it can handle making predictions as if it were in real time (Fig. 2). This approach prevents data leakage and produces more reliable performance estimates for time-dependent models. Furthermore, it aids in detecting issues like overfitting or underfitting, allowing model developers to refine their approaches to achieve better predictive accuracy and robustness. It's a fundamental technique when building machine learning models for applications involving sequential data.

3.6. Data Set

The dataset provided a historical time series of the West Texas Intermediate (WTI) crude oil price, a fundamental

and influential indicator in the global energy markets. The data spans over a substantial period, commencing on January 2, 1986, and extending until July 14, 2023, constituting 9456 data points. This long-term dataset is exceptionally valuable for understanding the dynamics of the energy sector and for predicting future trends, making it of particular interest to the financial, economic, and energy communities because covering significant economic events like the European financial crisis of 2008-2010 and oil market fluctuations in 2014 and 2015, this dataset becomes particularly remarkable when scrutinizing the unique and unprecedented impact of the COVID-19 pandemic on oil prices, as clearly illustrated in Fig. 3. Date and closing price are two significant. Closing daily price of WTI crude oil for each day serves as an indicator of the daily market sentiment and is essential for monitoring market fluctuations and trends. The "return rate" measures the rate of daily price change, which is a key metric for traders and investors looking to understand the day-to-day fluctuations in oil prices and potentially identify patterns and trends. The inclusion of this "return" metric, expressed as a percentage, is pivotal for advanced financial analysis and modeling. This dataset's extensive temporal coverage and inclusion of price and return rate information render it a valuable resource for in-depth research, econometric analysis, and financial modeling of the energy market dynamics, which are essential for academic studies and the broader financial community.

3.7. Evaluation Metrics

The evaluation of models, evaluation metrics are used to assess the performance and effectiveness of a model in making predictions or classifications. These metrics provide insights into how well a model generalizes to new, unseen data and can guide the selection of the most suitable algorithm for a specific task. In the context of predicting the behavior of the oil return rate during the COVID-19 pandemic, three evaluation metrics are used in this study:

- *Mean Squared Error (MSE)*: The Mean Squared Error (Alpaydin, 2020) is a fundamental metric for quantifying the average of the squared differences between predicted and actual values in a regression model. It assesses the precision of predictions, with



Fig. 2. Splitting basis time series data with four splits.

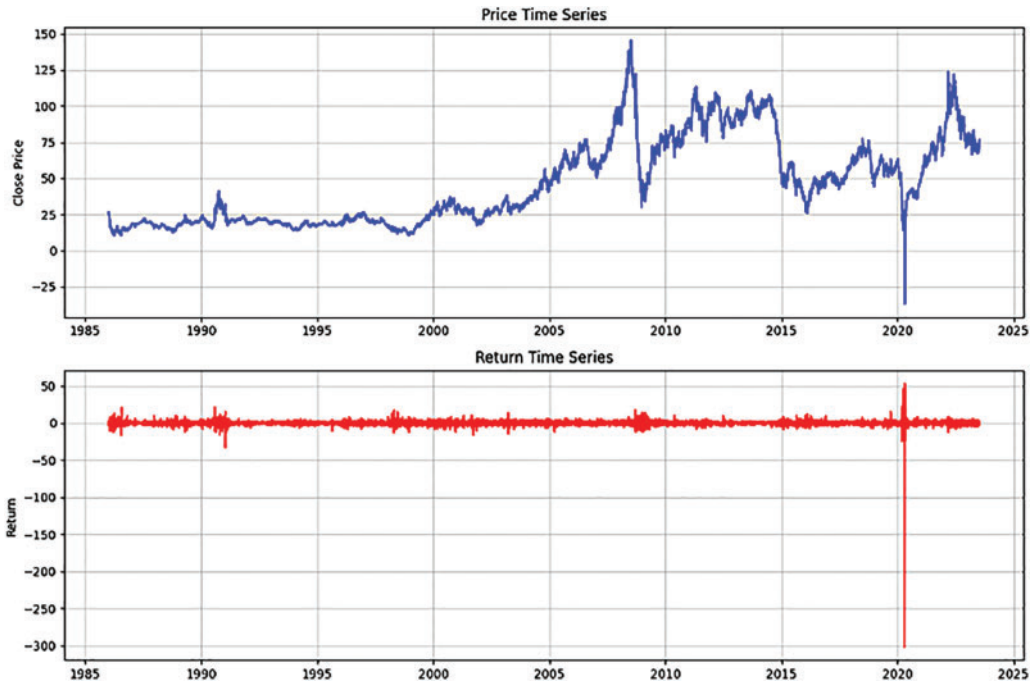


Fig. 3. Historical data of the West Texas Intermediate (WTI) crude oil.

lower MSE values indicating better model accuracy. Mathematically, MSE is calculated as the sum of squared residuals divided by the number of data points. The Root Mean Squared Error (RMSE) is derived from the MSE and is particularly useful for interpreting prediction errors in the original unit of the target variable. It's the square root of the MSE and provides a more interpretable measure of prediction accuracy (Friedman, 2001). The equations are as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (1)$$

$$RMSE = \sqrt{MSE} \quad (2)$$

where y_i represents actual values, \hat{y}_i is predicted values and \hat{y}_i is the number of data points.

- **Mean Absolute Error (MAE):** The Mean Absolute Error (James et al., 2013) evaluates model accuracy by calculating the average of the absolute differences between predicted and actual values. MAE is less sensitive to outliers compared to MSE and

is often used when outlier impact needs to be minimized. Mathematically, (3) shows the MAE for predicted values \hat{y}_i and actual values y_i :

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

- **Directional Statistic (Dstat):** Directional Statistic is a metric for assessing the accuracy of directional predictions, particularly in finance and time series analysis. It measures the proportion of correct directional predictions (up or down) relative to the total predictions, providing insights into the model's ability to predict trends (Makridakis & Hibon, 2000).

$$Dstat = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

4. RESULT

The dataset used in our study includes historical data WTI covers a range of economic conditions, including the aftermath of the global financial crisis and the ongoing COVID-19 pandemic. In order to evaluate the impact of the COVID-19 pandemic on the WTI with rolling basis Cross-Validation, three different observations based on spikes of COVID-19 are created. The first one covers a stable economy before COVID-19, the second includes the pandemic's early stages, and the third extends the analysis to a longer time period impacted by the pandemic Fig. 4.

After data transformation, the next step involves applying machine learning to unveil hidden patterns among data points. Hyperparameters of LSTM, XGBoost, and



Fig. 4. Three rolling basis cross-validation for WTI from January 2, 1986, to July 14, 2023.

TABLE I: RESULTS OF ROLLING BASIS CROSS-VALIDATION

	LSTM	XGBoost	SVM	Train and Test Size
MSE	6.24	6.26	6.22	Train 1-Test 1
	6.20	6.18	6.19	Train 2-Test 2
	6.81	6.86	6.80	Train 3-Test 3
RMSE	2.50	2.50	2.49	Train 1-Test 1
	2.49	2.49	2.49	Train 2-Test 2
	2.62	2.62	2.61	Train 3-Test 3
MAE	1.87	1.88	1.86	Train 1-Test 1
	1.87	1.87	1.87	Train 2-Test 2
	2.04	2.04	2.03	Train 3-Test 3
Dstat	0.1	0.5	0.1	Train 1-Test 1
	0.4	0.3	0.1	Train 2-Test 2
	0.1	0.3	0.1	Train 3-Test 3
	0.2	0.8	0.1	Average

TABLE II: RESULTS OF SPLITTING BASIS CROSS-VALIDATION

Model	Evaluation metrics	Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Split 8	Split 9	Split 10	Average
LSTM	RMSE	2.91	1.58	2.53	2.55	2.32	2.95	1.78	2.36	1.93	12	3.3
	MSE	8.48	2.51	6.39	6.48	6.56	8.68	3.17	5.58	3.72	144.0	19.4
	MAE	1.77	1.17	1.76	1.85	1.71	2.02	1.3	1.65	1.36	2.87	1.7
XGBoost	RMSE	3.42	1.61	2.53	2.94	2.82	2.99	1.81	2.36	1.93	12.01	3.5
	MSE	11.73	2.6	6.39	8.63	7.95	8.97	3.27	5.56	3.73	144.3	20.4
	MAE	2.3	1.19	1.76	2.23	2.22	2.09	1.33	1.66	1.37	2.89	1.9
SVM	RMSE	2.92	1.58	2.52	2.55	2.34	2.95	1.78	2.36	1.93	11.9	3.3
	MSE	8.5	2.5	6.37	6.5	5.46	8.67	3.17	5.58	3.72	141.6	19.4
	MAE	1.78	1.17	1.76	1.87	1.74	2.02	1.3	1.66	1.36	2.87	1.8
Training period		86-01 to 89-05	86-01 to 96-02	86-01 to 96-02	86-01 to 99-07	86-01 to 03-01	86-01 to 06-06	86-01 to 09-11	86-01 to 13-04	86-01 to 16-09	86-01 to 20-02	–
Testing period		89-05 to 92-09	92-10 to 96-02	96-03 to 99-07	99-07 to 03-01	03-01 to 06-06	06-06 to 09-11	09-11 to 13-04	13-04 to 16-09	16-09 to 20-02	20-02 to 23-07	–

SVM are meticulously tuned. Utilizing a rolling basis cross-validation from January 1986 to July 2023, we assess the forecasting performance of these models. Results in Table I show SVM's superior accuracy, although distinctions between models are not highly significant. The Dstat metric highlights forecasting direction. XGBoost excels in classification, accurately predicting WTI return rates, while SVM leads in regression, minimizing errors in return rate forecasts. These findings underscore the adaptability of machine learning in energy forecasting.

When employing splitting basis cross-validation to predict WTI return rate movements using a 10-fold time series cross-validation technique. The results (Table II) reveal significant insights into the performance of each model across different training and testing periods. LSTM, an increasingly popular deep learning model, presents comparable performance with an average RMSE of 3.3, underlining its effectiveness for sequential data analysis.

SVM, although competitive with an average RMSE of 3.3, exhibits slightly lower errors and marginally increased predictive capability compared to the ensemble based XGBoost and LSTM. These findings help practitioners select the most appropriate model for forecasting crude oil price in various market conditions and underline the importance of dynamic model selection based on the specific training and testing periods.

5. CONCLUSION

This study indicates the precision of predictions for crude oil return rates of West Texas Intermediate (WTI) during the COVID pandemic, which holds significant implications for decision-makers in the dynamic energy market. Through a comprehensive assessment of three advanced machine learning models—LSTM, XGBoost, and SVM—this research not only evaluates their forecasting accuracy but also underscores the importance of hyperparameter optimization.

This analysis spans different economic periods, including pre-COVID-19 stability, early pandemic stages, and an extended crisis-impacted timeframe. The outcomes reveal that SVM demonstrates superior accuracy across various metrics, highlighting its competence in regression analysis. However, the differences in performance among the three models are nuanced. The introduction of the Dstat metric provides further insights into forecasting direction, showcasing XGBoost's prowess in classification analysis and SVM's leadership in regression analysis.

These findings underscore the adaptability and versatility of machine learning models in addressing the complex task of forecasting in the energy sector. They provide valuable guidance to stakeholders in the financial and corporate sectors, and governments, enabling more informed decisions in turbulent markets. Furthermore, the research

underscores the significance of hyperparameter tuning in achieving optimal forecasting performance.

The choice of machine learning model should be influenced by the specific focus of forecasting, be it classification or regression. SVM is recommended for precise return rate predictions, while XGBoost excels in directional forecasting.

These results set the stage for further exploration, including hybrid models or the integration of additional variables, potentially yielding advanced predictive tools that enhance financial forecasting accuracy and investment strategies.

CONFLICT OF INTEREST

The authors of this article do not have any conflict of interest.

REFERENCES

- Abdidizaji, S., & Pakizeh, K. (2021). Momentum strategies, statistical arbitrage and the market efficiency the case of Tehran stock exchange. <http://dx.doi.org/10.2139/ssrn.3943891>.
- Alpaydin, E. (2020). *Introduction to Machine Learning*. MIT Press. <https://doi.org/10.1007/s11517-020-02148-2>.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, 24, 2546–2554. <https://doi.org/10.1016/j.bbr.2022.114201>.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(2), 281–305.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press. <https://doi.org/10.22038/IJBMS.2023.68487.14937>.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(2), 1189–1232.
- Ghodduzi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81, 709–727.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. vol. 112, pp. 18. New York: Springer.
- Jammazi, R., & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, 34(3), 828–841.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476.
- Moshiri, S., & Foroutan, F. (2006). Forecasting nonlinear crude oil futures prices. *The Energy Journal*, 27(4), 81–96.
- Norouzi, S. S., Akbari, A., & NaserSharif, B. (2019, October). Language modeling using part-of-speech and long short-term memory networks. *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE, 182–187.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- Sun, C., Min, J., Sun, J., & Gong, X. (2023). The role of China's crude oil futures in world oil futures market and China's financial market. *Energy Economics*, 120, 106619. <https://doi.org/10.1016/j.eneco.2023.106619>.
- Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer Science & Business Media.
- Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), 2623–2635.