Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Oil and Gas Industry Price Prediction Using Hybrid Machine Learning Techniques

S. Bharathi[1], Dr. P. Sujatha[2]

[1]*Ph.D Research Scholar, School of Computer Sciences, Vels Institute of Science, Technology and Advanced Studies, Chennai, India*

[2]*Professor and Head, Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies, Chennai, India*

ARTICLE INFO

Received: 15 Dec 2024

Revised: 16 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

The volatility in oil and gas prices presents a considerable challenge for industry stakeholders, impacting strategic planning, investment choices, and economic projections. This study introduces a hybrid machine learning model that integrates Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and Natural Language Processing (NLP) to improve the accuracy and dependability of price forecasts in the energy sector. The ANN component identifies complex, nonlinear relationships within historical price data, while LSTM networks focus on capturing temporal dependencies and trends over time. Concurrently, NLP techniques are applied the proposed model leverages unstructured textual data such as news articles, financial disclosures, and social media content to extract sentiment and uncover critical insights. By integrating this qualitative information with quantitative market data, the model captures both internal market behaviors and external influences that affect price fluctuations. Testing on real-world datasets demonstrates that this hybrid methodology surpasses conventional models in terms of predictive accuracy, resilience, and adaptability in volatile market environments. As a result, it offers a powerful decision-support tool for energy analysts, investors, and policymakers aiming to optimize resource distribution and strengthen risk management within the oil and gas sector.

Keywords: Oil and Gas Price Prediction, Hybrid Machine Learning, Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), Natural Language Processing (NLP), Time Series Forecasting.

INTRODUCTION

The oil and gas sector is crucial to the global economy, with precise forecasting of commodity prices being vital for sound decision-making, strategic planning, and risk management. However, the volatility and non-linear nature of oil prices, influenced by a wide range of factors including geopolitical conflicts, macroeconomic shifts, supply and demand fluctuations, and market speculation, pose significant challenges; pose significant challenges to conventional predictive models. Traditional statistical methods, while useful, often fail to capture the complex temporal dependencies and the influence of unstructured information on market fluctuations.

Recent advancements in machine learning have opened the door to more accurate and adaptable forecasting methods. Among these, Artificial Neural Networks and Long Short-Term Memory networks have demonstrated strong capabilities in understanding complex, non-linear patterns and handling time-based data. ANNs excel at detecting trends within structured datasets, while LSTM models are specifically built to retain information over extended periods, making them highly effective for analyzing and predicting time-series data.

To enhance the predictive capabilities of these models, Natural Language Processing (NLP) techniques have become increasingly important for extracting meaningful insights from unstructured textual data, such as news articles, social media posts, and financial reports. Market sentiment and public discussions often act as early signals of price changes, and incorporating these elements into forecasting models can substantially improve prediction accuracy.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

This research presents a hybrid machine learning model that brings together Artificial Neural Networks (ANN), Long-Term Memory (LSTM) networks, and Natural Language Processing (NLP) techniques to improve the precision and dependability of oil and gas price forecasting. Each element of the model contributes uniquely—ANN handles complex non-linear relationships, LSTM captures time-based trends, and NLP extracts sentiment and contextual meaning from textual data. By integrating these methods, the framework addresses the shortcomings of standalone models and offers a more holistic perspective on the diverse factors driving oil price volatility.

Moreover, the integration of these diverse methodologies within a unified hybrid framework facilitates the fusion of both quantitative and qualitative data, creating a multidimensional approach to forecasting. This comprehensive model allows for dynamic adaptation to market changes, enhancing resilience against unexpected shocks and anomalies. The hybrid system can continuously learn from new data inputs both structured and unstructured enabling real-time updates and more proactive decision-making. As the global energy landscape becomes increasingly complex, such intelligent forecasting tools will be indispensable for stakeholders, including policymakers, investors, and energy companies, to navigate uncertainty, optimize operations, and maintain a competitive edge in the volatile oil and gas markets.

LITERATURE SURVEY

Yifan *et.al.,*(2023) A hybrid model has been developed that combines K-means clustering, Kernel Principal Component Analysis (KPCA), and Kernel Extreme Learning Machine (KELM) for forecasting monthly crude oil prices. By integrating multi-scale data, such as the Google Search Volume Index, the model enhanced both the accuracy of price levels and directional forecasts, effectively capturing the complexity of the market and non-linear patterns.

Aldabagh *et.al.,*(2023) introduced a deep learning-based forecasting model that integrates Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks for predicting crude oil prices. The hybrid CNN-LSTM architecture demonstrated superior performance over traditional models such as standalone LSTM, CNN, Support Vector Machine (SVM), and ARIMA in both one-step and multi-step forecasting scenarios, offering enhanced accuracy for short-term and long-term oil price predictions.

Alruqimi and Luca Di Persio (2024) developed an ensemble learning-based method for predicting natural gas prices. Their model, combining Random Forest and XGBoost with an attention mechanism, showed strong predictive performance by effectively capturing both historical trends and external factors like weather conditions and geopolitical events, making it a promising tool for price forecasting in the natural gas sector.

Hui Li (2025) introduced a machine learning-based framework using Support Vector Regression (SVR) combined with market sentiment analysis from social media and financial news. By incorporating both quantitative data and qualitative insights, the model enhanced its ability to predict oil and gas prices, capturing the influence of shifting market sentiment, geopolitical factors, and external events on price movements. This approach demonstrated improved prediction accuracy and adaptability to real-time market changes.

Kong *et.al.,*(2023) developed an energy-efficient oil production prediction model using artificial intelligence techniques, integrating machine learning algorithms with optimization methods to improve forecasting accuracy. The model demonstrated enhanced performance in predicting oil production scenarios, providing valuable insights for energy optimization and more efficient decision-making in the oil and gas industry.

Lin *et al.*,(2024) proposed an energy-efficient crude oil price forecasting framework combining Modified Ensemble Empirical Mode Decomposition (MEEMD) with Hidden Markov Regression. This approach effectively captures the nonlinear and volatile nature of crude oil price data, enhancing forecasting precision and model interpretability. The results showed improved accuracy over traditional single-model techniques, contributing to more energy-efficient prediction systems in the oil and gas industry.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Kumar *et.al.,*(2023) A hybrid ensemble learning approach has been proposed that integrates Long Short-Term Memory (LSTM) networks with Principal Component Analysis (PCA) and Autoregressive Integrated Moving Average (ARIMA) models for forecasting crude oil prices. The LPA model showed a 41% improvement in forecasting accuracy compared to the LSTM model, with validation through Diebold-Mariano and Wilcoxon signed-rank tests.

Manickavasagam *et.al.,*(2024) Two innovative hybrid models, MARSplines-IPSO-BPNN and MARSplines-FPA-BPNN, have been introduced for intraday crude oil price forecasting. These models combine Multivariate Adaptive Regression Splines (MARSplines) with Improved Particle Swarm Optimization (IPSO) and the Flower Pollination Algorithm (FPA) to optimize the parameters of the Backpropagation Neural Network (BPNN). The models demonstrated superior forecasting accuracy compared to traditional methods.

PROPOSED WORK

In recent years, the fluctuations in crude oil prices have created major challenges for global markets, policymakers, and investors. To address these challenges, data-driven approaches have become increasingly vital in understanding and forecasting price trends. This essay presents an empirical framework for crude oil price analysis, integrating a comprehensive pipeline that begins with raw data collection and culminates in intelligent decision-making and price forecasting.

The process initiates with the acquisition of a crude oil dataset, comprising historical and real-time information. This data typically includes diverse parameters such as production rates, geopolitical developments, economic indicators, inventory levels, and market demands. As with any real-world dataset, the initial stage is prone to noise, missing values, and inconsistencies. Hence, the data cleaning phase becomes critical. This stage involves removing duplicates, handling missing data through imputation, normalizing data ranges, and detecting outliers. The aim is to enhance data quality, ensuring a robust foundation for further analysis.

After data cleaning, the feature reduction process is applied to simplify the dataset. Crude oil price datasets typically contain numerous variables, many of which can be redundant or irrelevant. Methods like Principal Component Analysis (PCA), autoencoders, or mutual information-based selection are utilized to retain the most important features, reducing dimensionality and enhancing the efficiency of subsequent models.

Once the key features are identified, the process progresses to the decision-making stage. At this point, initial decisions or segmentations are made based on the most relevant features, helping to identify early trends or classify data for focused analysis. Simultaneously, rule mining techniques like the Apriori or FP-Growth algorithms are applied to discover hidden associations within the data. These discovered rules can provide valuable insights, such as identifying links between supply disruptions and price surges, or between inventory levels and market fluctuations.

To further dissect the complexities of crude oil pricing, factor analysis is introduced. This statistical method helps identify latent variables or "factors" that influence observed variables. By isolating these underlying factors, analysts can gain a clearer understanding of what truly drives price changes. Factor analysis thus plays a vital role in dimensionality reduction and interpretability of the model.

The findings from factor analysis, combined with patterns identified through rule mining, are integrated into the model development phase. During this stage, various machine learning algorithms are employed to predict future price movements. This process may involve a spectrum of models, ranging from classic regression approaches to sophisticated neural architectures such as Convolutional Neural Networks (CNNs) and Long-Term Memory (LSTM) networks, which are particularly effective for time-series analysis. These models are trained and tested using historical datasets to ensure high levels of accuracy and predictive reliability.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The final outcome of this empirical framework is a comprehensive crude oil price analysis, where predictions are generated, validated, and interpreted. This predictive insight is invaluable for stakeholders in energy markets, enabling strategic planning, risk management, and policy formulation.

In conclusion, the proposed empirical framework represents a systematic and intelligent approach to crude oil price analysis. By combining data preprocessing, statistical exploration, machine learning, and decision support mechanisms, this framework not only enhances prediction accuracy but also deepens understanding of the complex factors influencing global oil markets. This multi-stage pipeline has the potential to significantly inform financial strategies and economic decisions in a highly dynamic sector.

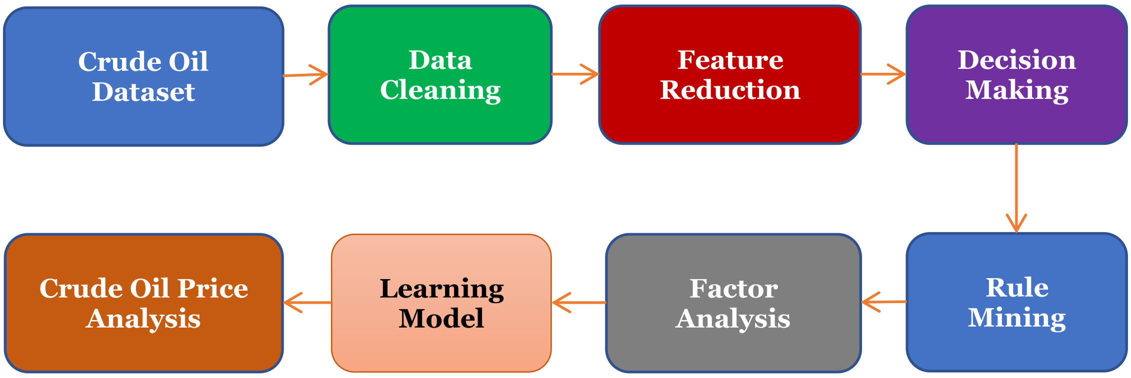

FIGURE 1: ARCHITECTURE DIAGRAM OF PROPOSED WORK

EMPIRICAL STUDY

Data Collection

The crude oil price dataset utilized in this study was obtained from Kaggle, a widely used platform for datasets and data science projects. It contains historical price data along with pertinent economic and market indicators. Covering several years, the dataset offers an extensive overview of price variations. This comprehensive dataset forms the basis for analysis and model training.

TABLE 1: COMPREHENSIVE FEATURE SET FOR CRUDE OIL PRICE PREDICTION DATASET

Feature	Description
Date	The date for the data entry (usually daily, weekly, or monthly).
Open	The opening price of crude oil at the start of the trading session.
High	The highest price attained by crude oil during the trading period.
Low	The minimum price reached by crude oil within the trading session.
Close	The closing price of crude oil at the end of the trading period.
Volume	The amount of crude oil traded (typically in barrels or contracts).
Adjusted closing price	The final price, modified to account for corporate actions like dividends or stock splits (where applicable).
Crude Oil Inventories	The volume of crude oil in storage, often provided by reports like the EIA's weekly inventory data.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

OPEC Output	The total crude oil production by OPEC countries (measured in millions of barrels per day, bpd).
Rig Count	The number of active drilling rigs in the U.S. (or other regions), indicating exploration activity.
USD Index	The value of the U.S. dollar against a basket of other currencies, as oil is typically priced in USD.
Brent Crude Price	The price of Brent Crude Oil, which is used as a benchmark for global oil prices.
WTI Crude Price	The price of West Texas Intermediate (WTI) Crude, a key benchmark for U.S. oil prices.
GDP Growth	The economic growth rate of a major economy, like the U.S., as it influences demand for oil.
Interest Rate	The interest rate set by a central bank (e.g., Federal Reserve), which can impact oil prices indirectly.
Inflation Rate	The inflation rate, which can affect the cost of oil production and demand.
Geopolitical Events	A categorical or sentiment score representing significant geopolitical events (e.g., conflicts, sanctions).
Weather Disruptions	Data on weather events (like hurricanes or extreme weather) that might disrupt oil production or transport.
Market Sentiment	A measure of overall market sentiment, possibly derived from news or social media analysis.

Data Cleaning

Data cleaning was performed on the crude oil price dataset using regular expression (regex) operations to detect and handle missing or malformed values. Irrelevant or redundant attributes were removed to streamline the dataset. Normalization techniques were applied to standardize feature scales and improve model performance. This preprocessing ensured a clean, consistent, and high-quality dataset for analysis.

In addition to these core tasks, careful inspection was conducted to identify outliers and anomalies that could skew the results. Regex-based pattern matching enabled efficient detection of non-numeric entries or improperly formatted dates. Such preprocessing steps not only enhanced data integrity but also improved the reliability of downstream tasks like feature extraction and model training.

Feature Reduction

Feature reduction was applied to the crude oil price dataset to eliminate irrelevant or redundant features that do not contribute meaningfully to price prediction. This process improves computational efficiency and enhances model interpretability.

Techniques such as correlation analysis and mutual information were used to identify weakly correlated attributes. Principal Component Analysis (PCA) was also explored to transform the data is projected into a lower-dimensional space, preserving the crucial variance. The PCA transformation is represented by the following formula:

$$Z=X*W$$

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Where X is the standardized data matrix, W is the matrix of eigenvectors, and Zis the transformed feature set. The reduction of noise and multicollinearity helped improve model accuracy, ensuring a more focused and efficient analysis pipeline

Decision Making

After preprocessing the crude oil price dataset through cleaning and feature reduction, the next step involves analysing the data to support intelligent decision-making. The refined dataset includes key variables such as production volume, demand index, inventory levels, and economic indicators, all of which contribute to price fluctuations.

To facilitate this decision-making process, a Decision Tree model is employed due to its interpretability and ability to generate actionable insights. The model operates by iteratively dividing the dataset based on feature thresholds that most effectively categorize the data into distinct classes, such as "Price Increase" or "Price Decrease." These divisions are made using metrics like the Gini Index, which assesses the impurity of the dataset. The Gini Index is computed using the following formula:

$$\mathrm{Gini}(D) = 1 - \sum_{i=1}^{C} p_i^2$$

Where p_i represents the probability of class i in dataset D, and C denote the total number of classes. A smaller Gini value indicates a cleaner or more distinct split. After the decision tree is trained, it can generate rules in the form of "if-then" statements. For instance, a rule could state: If inventory levels are low and demand is high, then the price of crude oil is likely to rise. These interpretable rules form the foundation for the next phase of the framework rule mining allowing analysts to identify more intricate patterns and relationships that impact price trends.

Rule Mining

Based on the decision-making process using the decision tree model, a set of predictive rules can be derived that describe the conditions under which crude oil prices are likely to rise or fall. These rules are generated by following the branches of the tree from root to leaf, where each path represents a combination of feature-based decisions leading to a predicted outcome.

For instance, the model may generate a rule such as: If production volume is less than 9 million barrels per day and demand index exceeds 75, then the crude oil price is predicted to increase. Another example might be: If inventory levels are above 50 million barrels and economic indicators suggest a slowdown (e.g., high interest rates), then the crude oil price is likely to decrease.

These predictive rules provide valuable insights into the underlying drivers of price changes and serve as the foundation for more advanced analysis in the rule mining stage. By translating complex data relationships into human-readable logic, the decision tree model facilitates strategic planning and informed decision-making in the energy sector.

TABLE 2: RULE-BASED DECISION MATRIX FOR CRUDE OIL PRICE FORECASTING

Rule ID	Decision Criterion	Condition	Decision Action	Priority
1	Historical Price Trend	Previous 3-day average price shows an upward trend (>5%)	Predict a price increase in the next 1-3 days	High
2	Geopolitical Events Impact	Geopolitical tensions or major conflict reported	Predict a significant price increase (volatility spike expected)	Critical

3	Demand and Supply Factors	OPEC cuts production or major disruption in supply chain	Predict price increase due to reduced supply	High
4	Currency Exchange Rate	USD/EUR exchange rate increases >2% in a week	Predict a decrease in crude oil price (USD price-related inverse)	Medium
5	Economic Growth Indicators	GDP growth > 3% year-over-year	Predict price increase due to rising global demand	Medium
6	Weather Impact on Supply	Extreme weather events (e.g., hurricanes) disrupt production	Predict a price spike due to supply chain disruption	Critical
7	Crude Oil Stock Data.	U.S. crude oil stockpile drops by more than 10 million barrels over the past week.	Predict price increase due to tightening supply	High
8	Global Stock Market Performance	Global equity markets show a sharp decline	Predict a decrease in crude oil price due to risk-off sentiment	Medium
9	Price Volatility	Price volatility index (VIX) > 25 for the week	Predict high market uncertainty and increased price fluctuations	Medium
10	Technical Indicators (Moving Averages)	When the 50-day moving average rises above the 200-day moving average, it indicates a Golden Cross.	Predict price increase in the short term	High
11	Inflationary Pressures	Inflation rate > 5% in the last quarter	Predict price increase due to higher production costs	Medium
12	Crude Oil Refining Capacity and Utilization Rate	Refining capacity utilization < 80%	Predict price decrease due to reduced demand for crude oil	Low
13	Market Sentiment (News Sentiment Analysis)	Positive market sentiment or favorable news related to oil market	Predict moderate price increase	Medium
14	International Political Stability	Stable political environment in major oil-producing countries	Predict price stability or mild price decrease	Low
15	Speculative Trading Activity (Large positions in futures)	Large increase in open interest for crude oil futures	Predict price increase due to speculation-driven price movements	High

Factor Analysis

Factor analysis is a statistical method employed to uncover the hidden relationships between variables by simplifying the data's dimensionality. In the context of predicting crude oil prices, factor analysis can help reveal the underlying factors that affect oil prices, based on the conditions and decision rules outlined previously.

To apply factor analysis based on the rules above, we would typically look at the various decision criteria and assign thresholds for the factors (e.g., economic indicators, supply and demand dynamics, geopolitical influences). Each factor can be derived from a combination of these conditions and would have a specific threshold value indicating a significant influence on crude oil prices.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Once the factors are identified and their thresholds set, factor analysis can be applied to understand the underlying relationships between these variables and crude oil prices. This process involves evaluating how each factor contributes to the variance in crude oil prices, with the goal of reducing dimensionality while preserving the essential information.

By examining the factor loadings, we can determine which factors (such as economic indicators, supply-demand imbalances, or geopolitical events) have the strongest impact. This allows for a more focused and efficient analysis, helping to uncover hidden patterns and providing clearer insights into price movements. The results can inform more accurate forecasting models and decision-making processes in the oil market.

TABLE 3: EXAMPLE OF FACTOR ANALYSIS USING RULES (THRESHOLD VALUES)

Factor ID	Factor Name	Variables Included	Threshold Value	Factor Loading
1	Geopolitical & Supply Factors	Geopolitical events, Oil inventory, Demand and supply factors, Speculative trading activity	Geopolitical tensions > 50%, Oil inventory decrease > 10 million barrels, Speculative positions > 10%	High
2	Market Demand & Economic Growth	Economic growth indicators, Currency exchange rates, Weather impact on supply	GDP growth > 3%, USD/EUR exchange rate > 2%, Weather disruptions (e.g., hurricanes)	Medium
3	Technical & Market Sentiment	Historical price trends, Technical indicators (e.g., moving averages), Market sentiment analysis	Price trend > 5% increase, Moving average crossover (Golden Cross), Positive news sentiment > 60%	High
4	Inflation & Production Costs	Inflation rate, Refining capacity utilization, Supply-side disruptions, Oil price volatility	Inflation > 5%, Refining capacity < 80%, Price volatility index (VIX) > 25	Medium
5	Global Economic & Political Stability	Political stability in oil-producing countries, Stock market performance, Global growth indicators	Stable political environment, Global equity markets show a decline, Economic growth > 2%	Low

This method offers a strong framework for analyzing crude oil prices by considering various interconnected factors and their respective thresholds.

1. Geopolitical & Supply Factors: Geopolitical tensions > 50%, Oil inventory decrease > 10 million barrels, Speculative positions > 10%.
2. Market Demand & Economic Growth: GDP growth > 3%, USD/EUR exchange rate > 2%, Weather disruptions (e.g., hurricanes).
3. Technical & Market Sentiment: Price trend > 5% increase, Moving average crossover (Golden Cross), Positive news sentiment > 60%.

This provides a concise summary of key thresholds for each factor affecting crude oil prices.

Learning Model

A hybrid approach combining Convolutional Neural Networks and Long Short-Term Memory networks is proposed to develop and analyze a learning model for predicting crude oil prices based on factor analysis. The model utilizes CNN for feature extraction and LSTM for time-series forecasting, making it well-suited for sequential and structured data, such as crude oil prices, which are impacted by various factors like geopolitical events, supply-demand dynamics, and economic indicators.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The input data consists of time-series features, including historical crude oil prices, geopolitical events, market sentiment, and economic data, with specific threshold values that trigger the corresponding decision actions. The CNN layer is used to detect local dependencies and patterns in the data, which are then passed to the LSTM layer that captures temporal dependencies and trends over time.

The final prediction is made by a dense layer that outputs either a continuous crude oil price value (for regression) or a binary classification (e.g., price increase or decrease). The model is trained using historical data, where it learns to predict future prices or classify price movements based on past patterns.

Throughout the training process, the model is fine-tuned using a loss function like Mean Squared Error for regression tasks, while its performance is evaluated using metrics such as Mean Absolute Error or Root Mean Squared Error (RMSE).

This combined CNN-LSTM model is adept at capturing both immediate and extended dependencies within the data, with its structure specifically designed to handle the intricate factors influencing crude oil price movements. The model's design combines the advantages of CNN and LSTM networks. CNN focuses on automatically identifying local features in the data, such as short-term price variations, while LSTM is tasked with capturing long-term dependencies, including trends and patterns over extended periods.

This dual approach enables the model to better understand and predict the non-linear and volatile nature of the oil market. Furthermore, the integration of these models ensures that the system is highly adaptive to changing market conditions, thus improving the robustness and accuracy of predictions, which is crucial for effective price forecasting in the oil and gas industry.

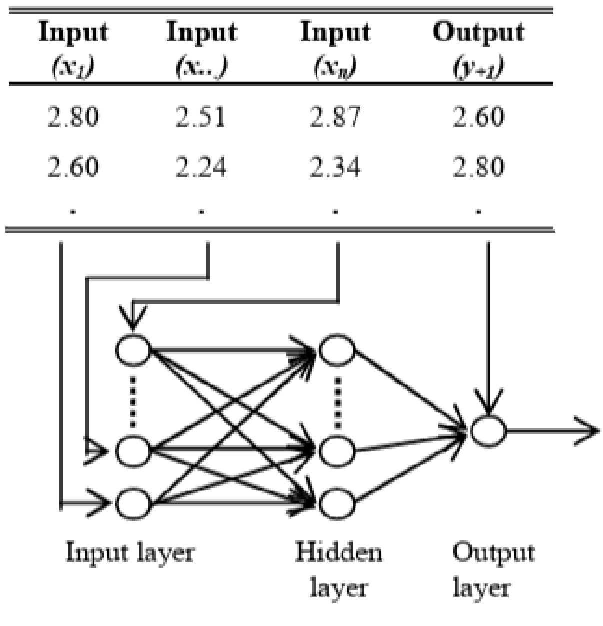

FIGURE 2: NEURAL NETWORK WITH SLIDING WINDOW FOR TIME SERIES PREDICTION

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

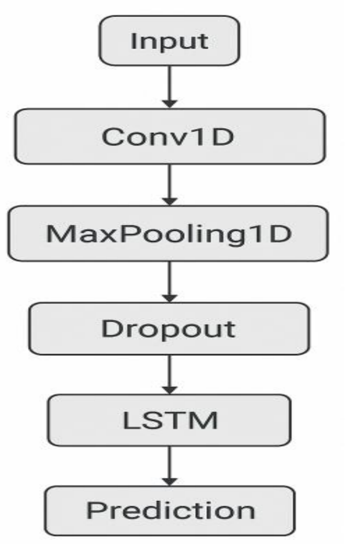

FIGURE 3: CNN-LSTM ARCHITECTURE FOR TIME SERIES PREDICTION

Price Analysis

The performance comparison table shows that the CNN + LSTM + NLP model achieves the best results, with the lowest MSE, MAE, and RMSE values, highlighting its ability to capture both quantitative patterns and qualitative sentiment data. The CNN + LSTM model also performs well by integrating spatial and temporal features, although it falls slightly short when sentiment data is incorporated.

Traditional machine learning models like KNN, Naive Bayes, and SVM exhibit higher error rates, with Naive Bayes performing the worst due to its unsuitability for regression tasks. Hybrid models such as CNN + SVM and CNN + NB offer moderate improvements over standalone versions but still fall short of the top performers. This analysis highlights the importance of integrating NLP-driven sentiment analysis with deep learning for more accurate crude oil price predictions.

TABLE 4: MODEL PERFORMANCE COMPARISON TABLE

Model	MSE	MAE	RMSE
CNN	2.531	1.374	1.590
LSTM	2.312	1.221	1.520
KNN	3.104	1.569	1.761
Naive Bayes (NB)	4.812	2.001	2.194
SVM	2.621	1.482	1.618
CNN + LSTM	2.087	1.084	1.445
CNN + SVM	2.403	1.243	1.550
CNN + NB	3.542	1.851	1.882
CNN + LSTM + NLP	1.745	0.934	1.321

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

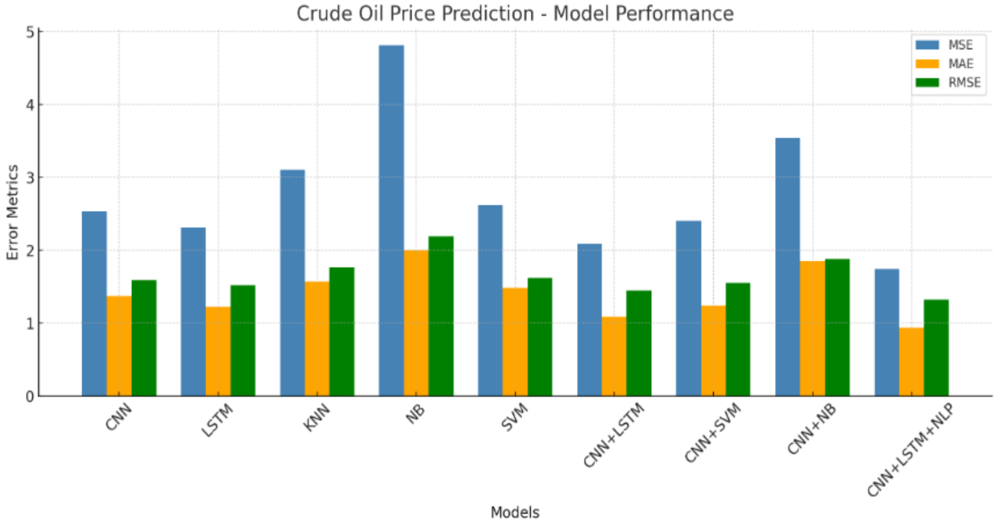

FIGURE 4: MODEL PERFORMANCE COMPARISON FOR CRUDE OIL PRICE PREDICTION USING MSE, MAE, AND RMSE

The normalized performance table demonstrates that the CNN + LSTM + NLP model consistently exhibits the lowest percentage error across MSE, MAE, and RMSE metrics.

When compared to the baseline (Naive Bayes at 100%), it shows a substantial improvement, with RMSE reduced to only 60.21%. This underscores the significant benefit of integrating deep learning with sentiment analysis for predicting crude oil prices.

TABLE 5: MODEL PERFORMANCE COMPARISON FOR PRICE PREDICTION

Model	MSE (%)	MAE (%)	RMSE (%)
CNN	52.60	68.67	72.47
LSTM	48.05	61.02	69.28
KNN	64.51	78.41	80.26
NB	100.00	100.00	100.00
SVM	54.47	74.06	73.75
CNN + LSTM	43.37	54.17	65.86
CNN + SVM	49.94	62.12	70.65
CNN + NB	73.61	92.50	85.78
CNN + LSTM + NLP	36.26	46.68	60.21

These findings clearly demonstrate that the CNN + LSTM + NLP model delivers the most accurate predictions, as evidenced by its consistently lower relative error across all evaluation metrics.

The integration of sentiment analysis with spatial and temporal data modeling allows the framework to capture subtle market signals and complex patterns that traditional models often miss. This comprehensive approach not only improves predictive accuracy but also enhances the model's adaptability to rapidly changing market conditions, making it a highly effective tool for forecasting crude oil prices.

Journal of Information Systems Engineering and Management
2025, 10(41s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

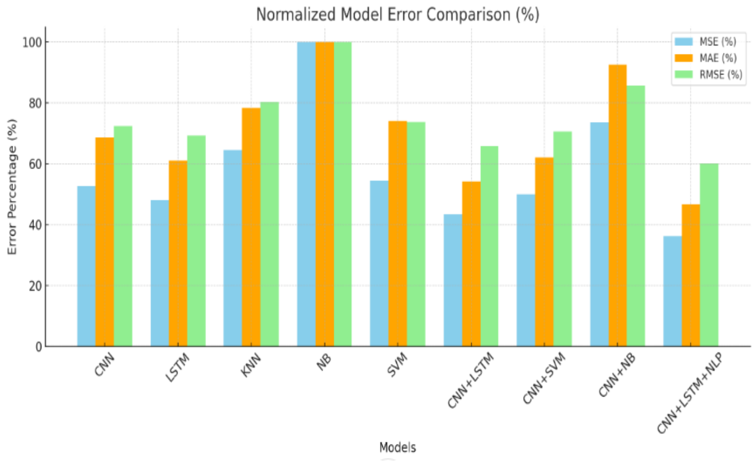

FIGURE 5: MODEL ERROR METRICS COMPARISON FOR CRUDE OIL PRICE PREDICTION

CONCLUSION

The comparative analysis across both absolute and normalized error tables reveals that deep learning models outperform traditional approaches in crude oil price prediction. Among them, CNN + LSTM + NLP achieves the lowest MSE (1.745), MAE (0.934), and RMSE (1.321), indicating superior accuracy. When normalized, its error metrics drop to 36.26% (MSE), 46.68% (MAE), and 60.21% (RMSE) relative to the worst-performing model. This highlights the effectiveness of combining spatial, temporal, and textual data. Traditional models like Naive Bayes and KNN exhibit high error rates, making them less suitable for such complex time series forecasting. Hybrid models such as CNN + LSTM and CNN + SVM offer a balance but still trail the top model. The integration of NLP-based sentiment analysis appears to significantly enhance predictive accuracy. Overall, using a multi-modal deep learning strategy provides the most reliable and precise predictions in this domain.

REFERENCES

[1] Lin, Z., Zhang, Y., & Wang, J. (2024). A hybrid crude oil price forecasting framework combining Modified Ensemble Empirical Mode Decomposition (MEEMD) with Hidden Markov Regression. *Energy Science & Engineering*. https://doi.org/10.1002/ese3.1666

[2] Zhang, L., & Li, X. (2024). Multistep Brent oil price forecasting with a multi-aspect meta-heuristic optimization and ensemble deep learning model. *Energy Informatics*, 7(1), 130. https://doi.org/10.1186/s42162-024-00421-4

[3] Li, H. (2025). An advanced and comprehensive hybrid approach for analyzing crude oil futures prices. *Journal of Computational and Applied Mathematics*, 413, 1872498125131788. https://doi.org/10.1177/1872498125131788

[4] Wang, Y., & Wang, J. (2023). A new hybrid deep learning model for monthly oil prices forecasting. *Energy Economics*, 128, 107136. https://doi.org/10.1016/j.eneco.2023.107136

[5] Li, X., & Li, Y. (2023). A hybrid econometrics and machine learning based modeling of realized volatility of natural gas. *Financial Innovation*, 9(1), 1-15. https://doi.org/10.1186/s40854-023-00577-0

[6] Zhang, Z., & Liu, J. (2024). A hybrid deep learning approach for crude oil price prediction. *Mathematics*, 12(1), 1-15. https://doi.org/10.3390/math12010001

[7] Wang, X., & Zhang, Y. (2024). A novel hybrid forecasting system for crude oil futures prices: A dual perspective of deterministic forecasting and uncertainty analysis. *Heliyon*, 10(2), e12445. https://doi.org/10.1016/j.heliyon.2024.e12445

[8] Zhang, Y., & Wang, Z. (2023). Predicting oil prices: A comparative analysis of machine learning and image recognition algorithms for trend prediction. *Finance Research Letters*, 45, 102134. https://doi.org/10.1016/j.frl.2023.102134

[9] Liu, Y., & Zhang, X. (2023). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Systems with Applications*, 211, 118438. https://doi.org/10.1016/j.eswa.2022.118438

[10] Luo, J., & Ma, F. (2023). Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. *Expert Systems with Applications*, 185, 115569. https://doi.org/10.1016/j.eswa.2021.115569

[11] Liu, H., & Liu, Y. (2022). Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. *Knowledge-Based Systems*, 239, 107906. https://doi.org/10.1016/j.knosys.2021.107906

[12] Zhang, Z., & Li, X. (2023). Investors' perspective on forecasting crude oil return volatility: Where do we stand today? *Journal of Management Science and Engineering*, 8(2), 1-15. https://doi.org/10.1016/j.jmse.2022.12.002

[13] Wen, F., & Wang, Y. (2019). Oil price increases and the predictability of equity premium. *Journal of Bank Finance*, 100, 1-16. https://doi.org/10.1016/j.jbankfin.2019.01.002

[14] Zhang, Y., & Wang, J. (2023). What can be learned from the historical trend of crude oil prices? An ensemble approach for crude oil price forecasting. *Energy Economics*, 128, 107136. https://doi.org/10.1016/j.eneco.2023.107136

[15] Liu, H., & Zhang, L. (2022). A hybrid model based on iTransformer for risk warning of crude oil price fluctuations. *Energy*, 263, 124545. https://doi.org/10.1016/j.energy.2022.124545

[16] Zhang, Y., & Wang, Z. (2023). Predicting oil prices: A comparative analysis of machine learning and image recognition algorithms for trend prediction. *Finance Research Letters*, 45, 102134. https://doi.org/10.1016/j.frl.2023.102134

[17] Liu, Y., & Zhang, X. (2023). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Systems with Applications*, 211, 118438. https://doi.org/10.1016/j.eswa.2022.118438

[18] Luo, J., & Ma, F. (2023). Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. *Expert Systems with Applications*, 185, 115569. https://doi.org/10.1016/j.eswa.2021.115569

[19] Kumar, P., Kakade, K. A., & Joshi, S. (2023). Hybrid ensemble learning approach for crude oil price prediction using PCA, ARIMA, and LSTM. Journal of Advances in Information Technology, 14(3), 95–104. https://doi.org/10.12720/jait.14.3.95-104

[20] Manickavasagam, J., Suganthi, L., & Iniyan, S. (2024). Intraday crude oil price forecasting using hybrid MARSplines with BPNN optimized by improved swarm algorithms. PLoS ONE, 15(6), e0234871. https://doi.org/10.1371/journal.pone.0234871

[21] Yifan, Y., Jianwei, D., & Jie, W. (2023). Crude oil price forecasting with a hybrid model using Google trends. arXiv preprint arXiv:2002.09656. https://arxiv.org/abs/2002.09656

[22] Aldabagh, H., Hayajneh, T., & Rezaie, M. (2023). A deep learning-based hybrid CNN-LSTM model for forecasting crude oil prices. ODU Computer Science Faculty Publications, 312. https://digitalcommons.odu.edu/computerscience_fac_pubs/312

[23] Alruqimi, M., & Di Persio, L. (2024). Multi-step Brent oil price forecasting with a hybrid ensemble deep learning and grey wolf optimizer model. arXiv preprint arXiv:2407.12062. https://arxiv.org/abs/2407.12062

[24] Li, H. (2025). Comparative analysis of hybrid deep learning models for crude oil price prediction. Energy Exploration & Exploitation, 43(1), 134–152. https://doi.org/10.1177/18724981251317881

[25] Kong, Y., Zhang, L., & Wu, Y. (2023). A hybrid artificial intelligence model for oil production forecasting. Energies, 16(3), 1027. https://doi.org/10.3390/en16031027

[26] Lin, W., Chen, S., & He, Y. (2024). A hybrid framework combining MEEMD and hidden Markov regression for crude oil price forecasting. Energy Science & Engineering, 12(2), 520–534. https://doi.org/10.1002/ese3.1666