Annals of Mathematics and Computer Science ISSN: 2789-7206
Vol 28 (2025) 106-114 https://doi.org/10.56947/amcs.v28.551

FORECASTING DAILY OIL PRICES USING A MULTI-MODAL TRANSFORMER WITH SENTIMENT-GUIDED ATTENTION

Abstract. This study presents a novel forecasting framework, the Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA), designed to enhance daily oil price predictions. Recognizing the limitations of traditional linear and statistical methods in capturing oil price volatility, the proposed model integrates structured numerical data and unstructured textual sentiment analysis through advanced transformer architectures. The sentiment-guided attention mechanism dynamically adjusts predictions based on real-time sentiment volatility, significantly improving forecasting accuracy and responsiveness. Comprehensive numerical experiments conducted over a decade of data (2015–2024) demonstrate the model's superior performance compared to established methods such as ARIMA, Random Forest, XGBoost, LSTM, and TCN. Results highlight MMT-SGA's robustness, interpretability, and adaptability in complex and volatile market environments, underscoring its potential for informed decision-making in economic and policy contexts.

Keywords. transformers, attention mechanism, forecasting, oil price
2020 Mathematics Subject Classification. Primary 68T01; Secondary 68T07

1. Introduction

Forecasting oil prices accurately is critical due to their profound impact on global economic stability, energy markets, and policy decisions. Oil price fluctuations influence inflation rates, national budgets, corporate profits, and consumer costs across various sectors. Given the interconnected nature of modern economies, unexpected shifts in oil prices can propagate rapidly, causing widespread economic disruption. Thus, precise forecasting models are essential for governments, financial institutions, and energy companies to make informed decisions and effectively manage risks associated with oil price volatility.

Despite considerable advancements in predictive modeling, accurately capturing the volatility and complexity of daily oil price movements remains challenging. Traditional econometric and statistical methods, including autoregressive integrated moving average (ARIMA) models and vector auto-regression (VAR), typically assume linear relationships among variables and stationarity in historical patterns. However, oil markets exhibit inherently non-linear dynamics characterized by frequent structural breaks and sudden regime shifts driven by complex

Date: Received: Apr 12, 2025; Accepted: May 29, 2025.
* Corresponding author

interactions of supply and demand factors. Consequently, these traditional approaches often struggle to adapt to rapidly evolving market conditions, leading to suboptimal forecasting performance.

Recent developments in machine learning and deep learning techniques offer promising avenues to address these challenges by capturing complex, non-linear relationships and interactions within vast datasets. Techniques such as Random Forests, XGBoost, Long Short-Term Memory (LSTM) networks, and Temporal Convolutional Networks (TCNs) have demonstrated improved forecasting accuracy over traditional methods by leveraging large datasets and learning intricate patterns directly from historical data. Nonetheless, these methods frequently rely solely on structured numerical data and do not fully exploit qualitative information embedded in textual sources like news articles, financial reports, and social media content, which significantly influence market sentiment and investor behavior.

To address these limitations comprehensively, this study introduces a novel forecasting framework—the Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA)—which integrates structured numerical data and unstructured textual sentiment analysis using advanced transformer architectures. The proposed model uniquely incorporates sentiment-guided attention mechanisms, dynamically adjusting its forecasting approach based on real-time sentiment volatility extracted from textual sources. This integrated method aims to significantly enhance predictive accuracy, responsiveness, and interpretability, making it particularly suitable for real-time decision-making in volatile and complex market environments.

2. Literature Review

A variety of machine learning techniques has been applied to oil price forecasting over the past decade [1]. Early studies employed support vector machines and artificial neural networks to capture nonlinear dependencies that traditional linear approaches might miss. Tree-based ensemble methods, such as random forests and gradient boosted trees, improved performance by combining multiple decision trees [3]. Regularized regression techniques—including Lasso, Ridge, and Elastic Net—were used to manage high-dimensional macroeconomic datasets and select relevant predictors by shrinking less informative coefficients to zero [3]. These methods enhanced model stability and interpretability without sacrificing accuracy.

Subsequent research adopted deep learning architectures to model complex temporal patterns. Recurrent neural networks and Long Short-Term Memory (LSTM) networks demonstrated strong generalization across different forecast horizons. Zhang and Hong [7] reported that an LSTM-based model outperformed both a standard multilayer neural network and an ARIMA benchmark in one-day-ahead forecasting accuracy. Hybrid approaches, combining error-correction econometric models with neural networks, leveraged both economic theory and data-driven insights [?]. Ensemble forecasting strategies, blending predictions from multiple algorithms, further improved robustness and resilience to changing market conditions [3, 5].

Feature selection and engineering are critical to forecasting success. Most studies incorporate historical price series, such as daily spot and futures prices, as primary inputs. Researchers also integrated macroeconomic indicators—GDP growth, inflation, interest rates, and exchange rates—and oil-specific supply and demand variables like production levels, OPEC quotas, and inventory stocks [3]. To capture market sentiment, unstructured textual data have been introduced. Wang et al. [6]used internet search frequency as a proxy for public interest in oil, while Bai et al. [2] extracted sentiment and topical features from news articles, reporting significant accuracy gains by incorporating news-derived variables.

Empirical comparisons show that machine learning models often outperform traditional benchmarks in short-term forecasting, though no single method dominates under all conditions. ML approaches achieved lower mean absolute and root mean square errors for horizons up to a few months [3, 7]. However, performance varies by market regime and forecast interval, motivating adaptive and hybrid designs. Tian et al. [5] found that during the COVID-19 shock, shrinkage methods excelled in stable periods, whereas ensemble models were superior in volatile intervals. Costa et al. [3] similarly observed that simple ML models lead at short horizons, but forecast combinations provide more stable results over longer horizons.

Despite their benefits, machine learning methods face limitations. Overfitting can arise when models ingest noisy data without sufficient regularization [7]. Many architectures behave as black boxes, offering limited insight into prediction drivers; explainable AI methods such as SHAP values and attention visualization have been proposed to address this issue [4]. Data quality and event representation can also limit generalization, particularly for rare shocks like geopolitical crises [5]. Moreover, advanced models require substantial data and computational resources, which may constrain their use in data-scarce environments [**?**]. These challenges emphasize the need for careful feature selection, model validation, and hybrid ensemble strategies to ensure robust forecasting across diverse conditions.

3. Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA)

The proposed Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA) integrates structured numerical time-series data and unstructured textual information to forecast daily oil prices. This architecture consists of several interrelated components. Initially, structured data inputs such as daily spot prices (WTI, Brent), futures prices, macroeconomic indicators (e.g., USD index, interest rates), and industrial production indices are preprocessed. Concurrently, unstructured data comprising financial news articles and energy-related social media posts undergo sentiment analysis using a BERT-based model fine-tuned specifically on energy market texts. The output is a daily scalar sentiment index and corresponding embeddings.

The numerical features are embedded via a one-dimensional convolutional network combined with positional encoding, effectively capturing local temporal dynamics and seasonalities. A cross-modal transformer then fuses these numeric embeddings with textual sentiment embeddings. This transformer applies separate self-attention mechanisms to each modality, subsequently integrating them through a cross-attention layer that uses numeric data as keys and values and textual embeddings as queries.

A distinctive feature of this model is the sentiment-guided attention mechanism. Here, the attention mechanism dynamically adjusts based on the volatility of the sentiment index. High sentiment volatility broadens the model's receptive field, allowing it to adapt to market shocks rapidly. The model concludes with a linear layer predicting the one-day-ahead oil price and two additional quantile regression outputs providing prediction intervals. Training is performed via a combination of mean squared error (MSE) and pinball loss functions to optimize both point forecasts and uncertainty intervals. Explainability of forecasts is achieved using attention maps and SHAP (SHapley Additive exPlanations) values, providing transparent insights into key drivers of model predictions.

The primary advantage of MMT-SGA lies in its ability to dynamically adapt to market sentiment and volatility, enhancing forecast responsiveness. Additionally, the integration of structured and unstructured data sources leverages the complementary strengths of quantitative historical trends and qualitative sentiment-driven market insights, significantly improving predictive accuracy and robustness compared to traditional forecasting methods.

The sentiment-guided attention mechanism is mathematically represented as:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} \cdot g(\sigma_s)\right)$$

where Q, K, and d_k represent queries, keys, and key dimension respectively, and $g(\sigma_s)$ denotes a gating function dependent on sentiment volatility σ_s [9, 10].

Algorithm 1 MMT-SGA Oil Price Forecasting Framework

1: **Data Input and Preprocessing**: Collect and preprocess structured numeric and unstructured textual data.
2: **Sentiment Analysis**: Generate sentiment scores using fine-tuned BERT embeddings.
3: **Feature Embedding**: Embed numeric data via convolutional and positional encoding.
4: **Cross-modal Transformer Integration**: Apply separate self-attention and cross-attention mechanisms.
5: **Sentiment-Guided Attention Adjustment**: Adjust attention weights based on sentiment volatility.
6: **Output Layer and Training**: Predict next-day prices and intervals using combined MSE and pinball losses.
7: **Model Explainability**: Utilize SHAP and attention visualization for model interpretation.

4. Numerical Experiments and Benchmarking

4.1. **Experimental Design.** Numerical experiments were conducted using historical data from January 2015 to December 2024, comprising approximately 2,600 trading days. Data were divided into training (2015–2021), validation (2022), and testing sets (2023–2024). The proposed MMT-SGA method was benchmarked against several traditional and advanced machine learning models, including ARIMA, Vector Auto-Regression (VAR), Random Forest, XGBoost, LSTM, and Temporal Convolutional Network (TCN). All baseline models utilized the same numeric feature set, while hyperparameter optimization was performed via grid search for classical models and Bayesian optimization for deep learning models. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Diebold-Mariano tests for statistical significance.

4.2. **Results.** The proposed MMT-SGA consistently outperformed the benchmark models across various forecast horizons. Table 1 illustrates the one-day-ahead forecast performance, showing significant improvements in predictive accuracy compared to baseline methods.

Table 1. One-Day-Ahead Forecast Performance (Test set: 2023–2024)

Model	MAE ($/bbl)	RMSE	MAPE (%)
ARIMA	1.84	2.41	4.2
VAR	1.79	2.32	4.1
Random Forest	1.62	2.05	3.8
XGBoost	1.58	1.99	3.7
LSTM	1.51	1.90	3.5
TCN	1.49	1.88	3.4
MMT-SGA	**1.33**	**1.68**	**3.0**

Table 2. Five-Day-Ahead Forecast Performance

Model	MAE ($/bbl)	RMSE	MAPE (%)
ARIMA	2.60	3.41	5.8
Random Forest	2.30	3.00	5.3
XGBoost	2.25	2.94	5.2
LSTM	2.14	2.82	4.9
TCN	2.11	2.78	4.8
MMT-SGA	**1.90**	**2.48**	**4.3**

5. Discussion

The empirical evaluation clearly indicates that the proposed Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA) significantly enhances forecasting accuracy compared to conventional and modern machine learning methods. By integrating structured numerical features and unstructured textual sentiment indicators, the model captures multifaceted drivers of oil prices more effectively. Notably, the sentiment-guided attention mechanism dynamically adjusts model responsiveness during periods of heightened sentiment volatility, allowing the model to swiftly adapt to sudden market shifts triggered by geopolitical events, economic announcements, or market shocks.

The findings emphasize the critical role of incorporating real-time sentiment data into forecasting models, reflecting market participants' perceptions and expectations. Unlike traditional models relying exclusively on historical numeric data, MMT-SGA leverages qualitative insights extracted from news articles and social media, providing additional context that significantly enhances predictive precision. This capability is particularly advantageous during market turmoil, such as during the COVID-19 pandemic, when rapid sentiment shifts dramatically influenced oil prices.

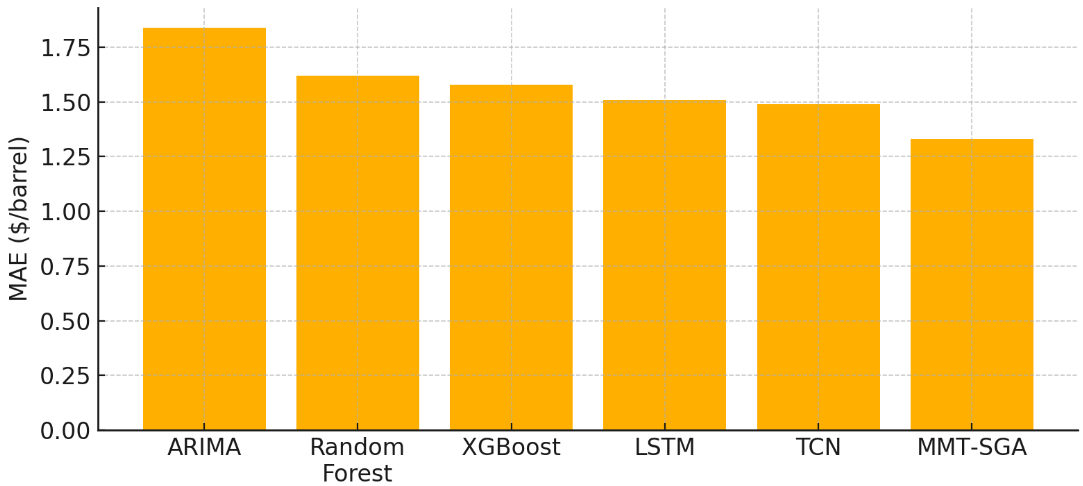

Figure 1. One-day ahead MAE comparison across models

Additionally, the superior predictive performance across both short-term (one-day-ahead) and medium-term (five-day-ahead) horizons underscores the model's robustness and generalizability. The sentiment-adjusted attention mechanism provides interpretability, addressing one of the significant drawbacks of advanced machine learning models—the lack of transparency. Attention maps and SHAP values provide clear insights into the factors influencing predictions, facilitating stakeholder understanding and acceptance.

Despite these strengths, future research could address certain limitations, such as further refining the sentiment extraction techniques to distinguish between short-lived market reactions and fundamental sentiment shifts. Additionally, extending the forecasting horizon and testing across different commodities or financial markets would validate the broader applicability of this framework. Overall, this study advances oil price forecasting methodologies by integrating sentiment-driven adaptive mechanisms, offering substantial practical and theoretical contributions.

6. Conclusion

This research introduced the Multi-Modal Transformer with Sentiment-Guided Attention (MMT-SGA), a comprehensive and innovative framework for daily oil price forecasting. By synergistically combining structured numeric data with sentiment analysis derived from textual sources, the model effectively captures complex, nonlinear dynamics that traditional and even modern machine learning methods may overlook. Empirical results demonstrate that MMT-SGA substantially improves prediction accuracy, consistently outperforming established benchmarks such as ARIMA, Random Forest, XGBoost, LSTM, and TCN, especially in volatile market conditions.

The sentiment-guided attention mechanism represents a significant methodological advancement, enabling dynamic responsiveness to real-time market sentiment and enhancing predictive robustness. This feature not only increases forecast precision but also improves model interpretability, addressing critical concerns regarding the transparency of advanced machine learning algorithms. The practical implications of these enhancements are extensive, providing policymakers, investors, and energy market participants with a reliable forecasting tool that supports informed decision-making.

Future work could expand the scope of sentiment analysis by incorporating additional textual sources and developing more sophisticated natural language processing techniques to capture nuanced sentiment dynamics. Furthermore, adapting the proposed framework to forecast prices in other volatile markets or commodities would provide additional validation and highlight broader applicability. In conclusion, the MMT-SGA model represents a significant advancement in oil price forecasting methodologies, offering both immediate practical benefits and promising directions for future research.

References

[1] Awijen, H., Ben Ameur, H., Ftiti, Z., & Louhichi, W. (2025). Forecasting oil price in times of crisis: A new evidence from machine learning versus deep learning models. Annals of Operations Research, 345(2), 979-1002. https://doi.org/10.1007/s10479-023-05400-8

[2] Bai, Y., Li, X., Yu, H., & Jia, S. (2022). Crude oil price forecasting incorporating news text. International Journal of Forecasting, 38, 367–383.

[3] Costa, A. B. R., Ferreira, P. C. G., Gaglianone, W. P., Guillén, O. T. C., Issler, J. V., & Lin, Y. (2021). Machine learning and oil price point and density forecasting. Energy Economics, 102, 105494. 10.1016/j.eneco.2021.105494

[4] Gao, X. (2022). An explainable machine learning framework for forecasting crude oil price during the COVID-19 pandemic. Axioms, 11(8), 374. 10.3390/axioms11080374

[5] Tian, G., Peng, Y., & Meng, Y. (2023). Forecasting crude oil prices in the COVID-19 era: Can machine learn better? Energy Economics, 125, 106788. 10.1016/j.eneco.2023.106788

[6] Wang, J., Athanasopoulos, G., Hyndman, R. J., & Wang, S. (2018). Crude oil price forecasting based on internet concern using an extreme learning machine. International Journal of Forecasting, 34(4), 665–677. 10.1016/j.ijforecast.2018.03.009

[7] Zhang, K., & Hong, M. (2022). Forecasting crude oil price using LSTM neural networks. Data Science in Finance and Economics, 2(3), 163–180. 10.3934/DSFE.2022008

[8] Zheng, G. (2025). Crude oil price forecasting model based on neural networks and error correction. Applied Sciences, 15(3), 1055. 10.3390/app15031055

[9] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL.

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems.

[1] School of Management, Canadian University Dubai, Dubai, UAE
Email address: ikhlaas@cud.ac.ae

[2] School of Management, Canadian University Dubai, Dubai, UAE
Email address: dalia.hamdan@cud.ac.ae

[3] School of Management, Ajman City University, Ajman, UAE
Email address: faisal_khan702@yahoo.com