International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2026, 16(2), 809-820.

From Data to Decision: Predictive Modeling of Oil Prices using AutoML and SHAP Analysis

Rihab Belguith*

Probability and Statistics Laboratory (LR18ES28), Faculty of Business and Economic Sciences, University of Sfax, Sfax, Tunisia.
*Email: belguithrihab@gmail.com

Received: 17 Ausgut 2025 **Accepted:** 13 December 2025 **DOI:** https://doi.org/10.32479/ijeep.22356

ABSTRACT

Machine learning and artificial intelligence (ML/AI), once regarded as opaque "black-box" methods, have become increasingly interpretable due to recent advances in explainable AI (XAI). This study proposes an explainable machine learning framework for forecasting crude oil prices by integrating the H_2O AutoML platform with SHapley Additive exPlanations (SHAP), thereby achieving both high predictive accuracy and transparent interpretability. Using daily macro-financial data from January 2015 to September 2025 including oil stocks (XLE), the S&P 500 index, industrial production, and the USD index. The study trains and validates a range of ensemble models, with Gradient Boosting Machines (GBMs) emerging as the best-performing models. The results demonstrate strong out-of-sample forecasting accuracy, measured by RMSE in USD per barrel, across different market conditions. Beyond predictive performance, the explainability analysis reveals that oil stocks (XLE), capturing energy-sector equity valuations, exert the strongest positive influence on crude oil prices, highlighting sectoral transmission channels and portfolio rebalancing effects. In contrast, the S&P 500 and industrial production display nonlinear and state-dependent impacts associated with business cycle dynamics, while the USD index exhibits a predominantly negative relationship, consistent with commodity–currency theory. This framework provides a robust approach to oil price forecasting by integrating automated machine learning with interpretable analytics, offering practical insights for investors, risk managers, and policymakers in volatile energy markets.

Keywords: Oil Price Forecasting, AutoML (H_2O), Energy and Financial Markets, Non-linear Effects
JEL Classifications: Q43, C63, G17, C52

1. INTRODUCTION

Forecasting oil prices remains a central challenge at the intersection of energy economics, financial market stability, and global macroeconomic policy, given the profound economic and geopolitical repercussions that arise from crude oil price volatility. The multifaceted interactions between oil prices, equity markets, exchange rates, and industrial output have been well-documented in the literature (Nakhipbekova et al., 2020; Rahmouni and Al Kahtani, 2025; Alqaralleh, 2024), showing how shocks propagate across asset classes and national borders. Advanced econometric frameworks, such as the TVP-VAR (Wen et al., 2019) and dynamic connectedness indices (Antonakakis et al., 2020), have illuminated the evolving relationships among oil, equities, and currencies. The US dollar index, in particular, exerts a bidirectional influence on oil prices and financial assets (Liao et al., 2018; Gatfaoui, 2016), while financialization and innovative market structures further intensify feedback loops (Rizvi et al., 2022).

However, as markets become increasingly complex, nonlinear, and subject to regime shifts, traditional linear models often fall short in capturing the dynamic causal chains and latent risks shaping oil price formation. Recent advances in machine learning have yielded significant improvements in prediction accuracy (Lundberg and Lee, 2017; Wen et al., 2019; Liu et al., 2019), but wide adoption of black-box models raises critical concerns around interpretability, especially as policymakers, investors, and managers seek actionable insights that go beyond mere forecasts.

Addressing this gap, the present paper leverages H_2O AutoML with integrated SHAP-based interpretability to provide not only robust, high-precision crude oil price predictions, but also transparent, granular explanations of the underlying drivers. The originality of this work lies in its comprehensive deployment of AutoML algorithms and explainable AI for oil price forecasting, a first in the literature, with cross-validation, feature importance rankings, SHAP summary and heatmap visualizations, partial dependence, and ICE plots. Using a daily panel covering macroeconomic and financial variables (oil stocks, S&P500, industrial production, USD index), models reliably track actual oil price movements across pre-crisis, crisis, boom, and stabilization regimes. Notably, oil stocks and S&P500 emerge as dominant predictors in both variable importance and SHAP analyses, with nuanced nonlinear and interaction effects revealed through PDP and ICE visualizations. The methodology achieves strong RMSE and MAE results, demonstrating generalizability and practical utility for energy finance stakeholders.

By integrating these domains, this research highlights the critical role of aligning financial systems with environmental sustainability to promote a more resilient and inclusive global economy. The remainder of the paper is organized as follows: Section 2 provides a review of the relevant literature, Section 3 outlines dataset, Section 4 presents methodology, Section 5 presents the empirical findings, Section 6 discusses the managerial implications.

2. LITERATURE REVIEW

Oil price forecasting represents an important bridge between energy economics, financial markets and advanced computational intelligence due to the macroeconomic significance of oil prices and the intrinsic complexity of characterizing its dynamic. Classical econometric works have provided a strong empirical base and accepted the fact that oil price interacts with US financial markets, exchange rate, global industrial output and strategic reserves of oil. Nakhipbekova et al. (2020), and Rahmouni and Al Kahtani (2025) demonstrate a short-term volatility spill over and long term cointegration.

American financial markets exhibit notable sensitivity to oil price fluctuations, with volatility transmission mechanisms thoroughly documented in works such as Sadorsky (1999) and Alqaralleh (2024). Employing advanced econometric frameworks like the time-varying parameter vector autoregressive (TVP-VAR) model, Wen et al. (2019) capture the evolving feedback loops between oil price shocks and sectoral equity market dynamics, highlighting the temporally adaptive nature of these relationships.

Integral to this ecosystem is the US dollar index, which exercises a bidirectional influence on both oil prices and financial assets. Liao et al. (2018) demonstrate that exchange rate fluctuations modulate crude oil price behavior and, reciprocally, oil market volatility impacts currency valuations (Gatfaoui, 2016), adding complexity to the international transmission of shocks, as further evidenced by Zhang et al. (2023)..

The financialization of oil commodities, fueled by derivatives, ETFs, and speculative capital inflows, has further intensified linkages between oil markets and equities, as articulated by Rizvi et al. (2022). Technological developments, notably the rise of shale oil production, have dynamically reshaped pricing structures and market sensitivities, with ripple effects permeating equity valuations (Mastepanov, 2016). Sector-specific analyses reveal asymmetric equity responses to oil price shocks, highlighting heterogeneous risk premiums and important implications for portfolio diversification and risk management in integrated markets (Dhaoui et al., 2021).

Parallel to econometric advances, the rapid evolution of machine learning (ML) techniques has led to significant improvements in forecasting oil prices. ML algorithms, including ensemble tree methods, neural networks, and support vector machines, have demonstrated superior predictive accuracy relative to classical linear models, especially under nonlinear or structurally unstable regimes (Wen et al., 2019; Liu et al., 2019). The adoption of explainable AI tools such as SHapley Additive exPlanations (SHAP), Partial Dependence Plots, and feature importance rankings enhances interpretability, addressing “black-box” criticism and aiding decision-makers in understanding key drivers (Lundberg and Lee, 2017).

AutoML platforms facilitate the automated discovery of optimal models and hyperparameters, democratizing access to sophisticated forecasting tools. The combination of AutoML and SHAP-based interpretability ensures robust, transparent models vital for energy price applications, where rich economic insights couple with predictive power.

Moreover, empirical evidence confirms that oil price volatility not only affects industrial production and macroeconomic stability but also reverberates through equity and currency markets on a global scale (Filis, 2010; Guesmi et al., 2016). ML-based innovations enable finer dissection of risk spillovers and interconnected feedback loops, offering valuable new perspectives on systemic transmission mechanisms.

In summary, prior research has established that oil prices, energy-sector stocks, US financial markets, and the US dollar index form a closely interconnected and dynamically evolving system. Building on this foundation, the present study advances the field by implementing an H_2O AutoML pipeline augmented with SHAP interpretability analysis delivering not only high-precision forecasts but also transparent, nuanced insights into the fundamental drivers of oil price fluctuations in global markets. Notably, this work is the first to comprehensively apply the integrated H_2O AutoML and SHAP framework to crude oil price prediction, placing particular emphasis on both feature-level and instance-level interpretability rather than focusing solely on overall predictive strength. In addition, our methodology prioritizes explainability and rigorous model validation through cross-model comparisons, feature effect visualization, and advanced diagnostic tools, ensuring that the resulting managerial and economic conclusions are firmly grounded in transparency and robust analytical evidence.

3. DATA AND RESEARCH DESIGN

3.1. Data

The dataset used in this study comprises key macroeconomic and financial variables that are commonly associated with oil price movements, spanning the period from January 2015 to September 2025. The dependent variable is the crude oil price (denoted as Oil Price), obtained from the CL=F futures contract, representing global benchmark prices. The predictors include the S&P 500 index (S&P500), representing equity market performance and investor sentiment; the USD Index (USD Index), reflecting the strength of the US dollar relative to major currencies; Oil Stocks (XLE) represent the price (in USD) of the Energy Select Sector SPDR ETF, which tracks the equity market performance of major U.S. energy companies, primarily engaged in oil, gas, and energy-related activities, and serves as a proxy for the overall performance of the energy sector in equity markets and Industrial Production (IP), representing overall economic activity and industrial demand. All variables are collected at a daily frequency, ensuring alignment across financial and economic indicators. Prior to modeling, the data were cleaned to handle missing values, and the series were merged into a single structured dataset suitable for AutoML modeling. Table 1 presents the descriptive statistics of the variables over the sample period. The statistics highlight strong heterogeneity in market behavior across asset classes, shaped by major macroeconomic and geopolitical shocks during the sample period. For instance, oil prices exhibit a mean of 62 USD with substantial dispersion (standard deviation of 18), reflecting the sharp volatility episodes linked to the COVID-19 collapse in 2020, the subsequent demand recovery, and the geopolitical pressures stemming from the Russia–Ukraine conflict. The S&P 500 shows a high average level (3,357 points) with considerable variability, consistent with the pandemic-induced market turmoil, unprecedented monetary easing, and later inflationary shocks that shaped equity market dynamics. The USD Index displays moderate fluctuations around 98, capturing the strengthening of the US dollar during global uncertainty (2020–2022) and the later normalization in international markets.

Oil stocks (XLE) present a mean of 27 with relatively high volatility, reflecting the sector’s sensitivity to energy price swings, OPEC+ production adjustments, and global supply-demand imbalances. Finally, Industrial Production follows a smooth upward trend with limited dispersion, consistent with a gradual post-pandemic recovery and the stabilization of global manufacturing activity.

3.2. Research Design, H_2O AutoML Framework and Implementation

To analyze and interpret predictive models for oil price forecasting, we adopted a comprehensive methodology integrating variable importance assessment, model interpretation techniques, and rigorous evaluation metrics. We first evaluated the influence of predictors across multiple models within the AutoML framework, aggregating importance scores from algorithms such as Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), and Distributed Random Forests (DRF) to identify the most impactful variables affecting oil prices. To understand the marginal effect of key features, such as the USD Index and S&P 500, on predicted oil prices, we generated Partial Dependence Plots (PDPs), while Individual Conditional Expectation (ICE) plots were employed to capture heterogeneity in predictor effects at the individual observation level, revealing potential nonlinearities and interactions. Additionally, SHAP (SHapley Additive exPlanations) summary plots quantified the contribution of each predictor to model outputs, providing both global interpretation and insights into feature-level effects for individual predictions. Model performance was rigorously assessed using metrics such as Root Mean Squared Error (RMSE) and residual diagnostics, with residual plots analyzed to detect potential issues such as heteroscedasticity or model misspecification, ensuring that the models fit the data appropriately. Finally, a variable importance heatmap was generated to compare the relative contributions of predictors across top-performing models, including categorical features encoded via one-hot encoding, confirming the robustness and consistency of the identified key predictors. Overall, this integrated approach ensures a thorough understanding of the predictive factors driving oil prices, their individual and combined effects, and the reliability and interpretability of the models employed.

Recent empirical advances have increasingly favored dynamic learning models capable of adapting and improving iteratively based on historical data patterns. This shift is particularly relevant in financial and energy price forecasting, where traditional regression methods often struggle due to inherent data complexity, multicollinearity, and nonlinear relationships among predictors. Machine learning algorithms (MLAs) emerge as powerful alternatives by synthesizing multiple, potentially weak, sources of information into robust composite predictive scores.

Among the advanced MLAs, deep learning, distributed random forests (DRF), generalized linear models (GLM), gradient boosting machines (GBM), and XGBoost have demonstrated

Table 1: Data preliminary analysis

Variable	Mean	Min	Max	Variance	Std_Dev	Skewness	Kurtosis
Oil Price (WTI)	62.001	−37.63	123.7	328.197	18.116	0.369	0.524
S&P500	3357.073	1829.08	6090.27	11746	1083.808	0.537	−0.733
USD Index	98.069	88.59	114.11	24.065	4.905	0.500	−0.261
Oil Stocks (XLE)	27.330	9.385	47.069	73.7025	8.585	0.626	−0.589
Industrial Production	115.007	99.293	130.815	75.292	8.677	−0.001	−1.195

Oil Price denotes the WTI crude oil spot price (USD per barrel). S&P500 refers to the level of the S&P 500 stock market index (index points). USD Index represents the U.S. Dollar Index measuring the value of the USD against a basket of major currencies (index). Oil Stocks corresponds to the XLE Energy Sector ETF price (USD). Industrial Production reflects the volume of goods produced over a given period and is typically expressed as an index (base 100) exceptional capability in uncovering complex, latent patterns within high-dimensional datasets. These algorithms excel at reducing prediction errors while maintaining optimal bias-variance tradeoffs, even under multicollinearity conditions.

Our study leverages the, H_2O Automated Machine Learning (AutoML) framework, a comprehensive suite of state-of-the-art machine learning models recognized for their forecasting accuracy and computational efficiency. H_2O AutoML automates the entire model-building pipeline—from data preprocessing and feature engineering to hyperparameter tuning and model selection—combining advanced algorithms with best practices in machine learning.

The H_2O AutoML implementation and training process was conducted in Python within a Jupyter Notebook environment, proceeding through a series of rigorously structured steps to ensure both accuracy and reproducibility. The workflow began with data import and preparation, where raw CSV data were transformed into H_2O frames suited for distributed computation. The dataset was split into training and validation sets using a ratio of 0.80, with 80% of observations used for training and 20% for validation, and a fixed random seed established to guarantee reproducibility and mitigate data leakage.

Model evaluation utilized 5-fold cross-validation, systematically partitioning the training set to build models by training on four folds and testing on the fifth, rotating through all folds. Model performance was benchmarked using metrics such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Per-Class Error, with lower values signifying superior predictions.

For algorithm exploration and hyperparameter tuning, H_2O AutoML trained a diverse set of base learners, including fully connected deep neural networks, distributed random forests (DRF), generalized linear models (GLM), gradient boosting machines (GBM), and XGBoost. Hyperparameter tuning was approached by minimizing the objective function $J(\delta)$:

$\delta^* = \arg\min_{\theta} J(\delta)$ where δ representes the model's hyperparameters and J is the validation metric, such as RMSE for regression.

Meta-learning and ensembling followed, with stacked ensemble models created using two strategies: Stacked Ensemble (All Models) combines predictions from all candidate models, while Stacked Ensemble (Best of Family) aggregates only the top model from each algorithmic family. The meta-learner then automatically determines the optimal blending weights OW by minimizing prediction error:

$OW^* = \arg\min_{\mathbf{w}} \sum_j (Z_j - \sum_i OW_i \hat{Z}_{ij})^2$ where Z_j is the observed value and $\hat{Z}_{ij}$ the prediction from model i for observation j. After training, all models were ranked on a validation-set-based leaderboard, with the top entry (M_{best}) selected for deployment:

$$M_{best} = top(Leaderboard)$$

Robustness checks included validation on an independent hold-out set, inspection of learning curves to diagnose potential overfitting or underfitting, and feature importance evaluation using SHAP (SHapley Additive exPlanations) values to interpret predictive drivers and ensure model transparency. This comprehensive methodology ensured the resulting AutoML models were not only highly accurate and generalizable, but also interpretable and practically robust for advanced financial forecasting applications.

4. EMPIRICAL RESULTS AND DISCUSSION

The evaluation of our models' predictive performance was based on metrics detailed in Table 1, emphasizing Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) as primary accuracy indicators. The table tracks the trajectory of RMSE values for H_2O AutoML models over successive time frames, highlighting optimal execution intervals and the effectiveness of various algorithms. We benchmark multiple machine learning models, including Gradient Boosting Machines (GBM), Distributed Random Forests (DRF), deep learning models, Generalized Linear Models (GLM), and Stacked Ensembles—which integrate predictions from various underlying models. The Mean Per Class Error, reflecting the average misclassification rate across all classes, served as an additional gauge of classification accuracy, where lower values denote higher precision. RMSE and MSE quantify the average discrepancy between predicted outcomes and their actual values; RMSE represents the square root of the mean squared deviations, while MSE captures their mean directly. Lower scores in these metrics confirm superior forecasting capability.

Table 2 presents the AutoML leaderboard summarizing the performance of various models trained to predict oil prices. The key evaluation metrics include the Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Logarithmic Error (RMSLE), residual deviance, training time, and prediction time per row, along with the corresponding algorithm types. GBM_4 emerges as the optimal model with an RMSE of 2.651 and MSE of 7.028, indicating the lowest average prediction error among all candidates and establishing it as the selected forecasting engine.

The StackedEnsemble_AllModels_1 ranks second with marginally higher error metrics (RMSE: 2.654, MSE: 7.045) but substantially longer inference time (2,448 ms versus 464 ms), reflecting the computational overhead of aggregating all base models; despite this, its near-parity with the top model suggests robust ensemble construction.

The leaderboard demonstrates strong relative performance consistency across the top four positions (GBM and StackedEnsemble variants), with RMSE values tightly clustered between 2.65 and 2.70. Gradient Boosting Machines (GBM) dominate the ranking, occupying three of the top five positions, underscoring their superior capability for capturing nonlinear oil price dynamics relative to alternative algorithmic families. In contrast, tree-based ensemble methods (DRF, XRT) and XGBoost exhibit comparatively weaker performance (RMSE ≥2.83), suggesting that sequential boosting approaches prove more

Table 2: AutoML Leaderboard: Lists the top-performing models, their algorithm types, MSE/RMSE, and ranking based on validation performance

Model_id	*RMSE*	*MSE*	*MAE*	*RMLSE*	*MRD*	*TT*	*PTPR*	*Algo*
GBM_4_AutoML_4_20251028_111446	2.6511	7.02831	1.77849	0.0431111	7.02831	464	0.031834	GBM
StackedEnsemble_AllModels_1_ AutoML_4_20251028_111446	2.65427	7.04512	1.83135	0.0465795	7.04512	2448	0.23038	StackedEnsemble
GBM_2_AutoML_4_20251028_111446	2.70205	7.30107	1.91105	0.0463201	7.30107	414	0.027736	GBM
StackedEnsemble_BestOfFamily_1_ AutoML_4_20251028_111446	2.70348	7.3088	1.86025	0.0473491	7.3088	1094	0.081195	StackedEnsemble
GBM_3_AutoML_4_20251028_111446	2.77439	7.69723	1.89012	0.0463808	7.69723	353	0.027434	GBM
XRT_1_AutoML_4_20251028_111446	2.83353	8.02888	1.97262	0.0485909	8.02888	726	0.030323	DRF
DRF_1_AutoML_4_20251028_111446	2.93667	8.62406	1.98828	0.0488397	8.62406	738	0.035189	DRF
XGBoost_2_AutoML_4_20251028_111446	2.98345	8.90099	2.09924	0.0601362	8.90099	2246	0.008242	XGBoost

model_id: H_2O AutoML model identifier, *RMSE*: Root mean squared error, *MSE*: Mean squared error, *MAE*: Mean absolute error, *RMSLE*: Root mean squared logarithmic error; *mean_residual_deviance*: Overall model fit, *training_time_ms* (TT): Training time (ms), *predict_time_per_row_ms* (PTPR): Average prediction time per row (ms), *Algo*: Algorithm type (GBM, DRF, XGBoost, GLM, StackedEnsemble)

effective than parallel ensemble or alternative gradient descent formulations for this particular forecasting task.

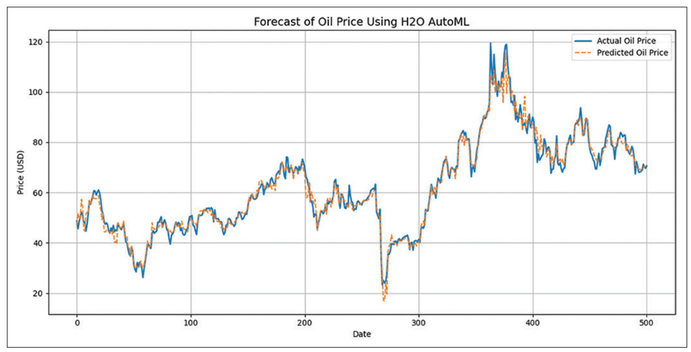
The mean absolute error (MAE) range of 1.78-2.10 USD per barrel indicates that typical point predictions deviate from realized prices by approximately 2 USD on average, a magnitude pragmatic for portfolio hedging decisions given crude oil's volatility scale. Training times vary substantially (353-2,448 ms), with GBM_4's efficiency (464 ms) providing computational advantage for real-time deployment scenarios. This performance hierarchy validates the H_2O AutoML framework's efficacy in automated model selection, with GBM_4 identified as the optimal production model for subsequent forecasting applications and decision support.

Once the best model has been identified, this selected model is employed to generate out-of-sample forecasts, which are then compared to actual oil price data. By plotting the predicted values alongside real prices over time, the model's forecasting accuracy and temporal tracking capabilities can be visually assessed, thereby complementing the quantitative accuracy metrics from the leaderboard with an intuitive depiction of model performance across different market conditions. Figure 1 presents a line plot comparing the actual observed oil prices with the predicted values produced by the AutoML model during the test period. This figure provides an intuitive evaluation of the model's forecasting accuracy and its capacity to track real market dynamics.

The figure presents a comparison between actual crude oil prices (solid line) and the forecasts generated by the H_2O AutoML model (dashed line) over the out-of-sample test period. The x-axis corresponds to the sequential order of observations in the test set, allowing an assessment of the model's ability to track unseen price dynamics.

Throughout the test sequence, the model demonstrates strong predictive capability, closely following the overall trends in the observed prices. Major upward and downward movements are effectively captured, indicating that the model successfully internalizes the patterns underlying oil price fluctuations. While extreme peaks and troughs are slightly smoothed, a typical feature of ensemble and neural-network-based forecasts, this smoothing ensures that predictions remain stable and robust, avoiding overreaction to transient noise in the data.

Figure 1: Oil Price Forecast vs Actual (Test Set) Line plot comparing actual oil prices with predicted values over the test period.



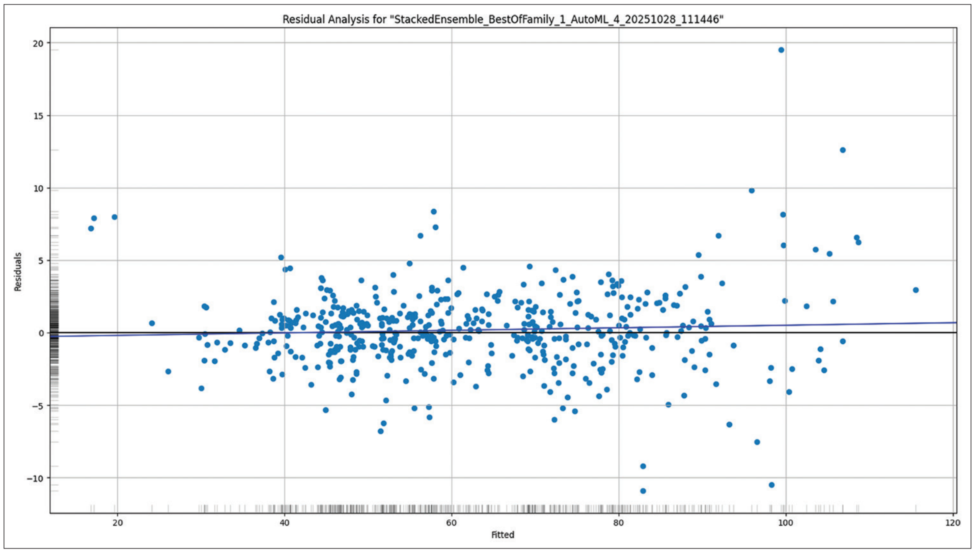
Minor lags are observed at abrupt inflection points, which is expected in data-driven models when confronted with sharp, rapid changes. Despite these small delays, the predicted series maintains consistent co-movement with actual prices, reflecting the model's ability to replicate the underlying dynamics across varying levels of volatility.

The figure also highlights the model's general stability and reliability: it adapts well to both gradual trends and sudden shifts, producing forecasts that remain tightly aligned with realized prices across the entire out-of-sample period. From a practical perspective, these results demonstrate that the AutoML framework provides a strong sequential tracking of price behavior, offering a useful tool for analysis and forecasting in dynamic and potentially volatile markets.

Building on the comparison between the predicted and actual oil prices, Figure 2 provides the residual analysis, offering a deeper assessment of the model's predictive reliability and the presence of any systematic deviations or heteroscedastic patterns in the forecast errors.

Residual diagnostic analysis of the StackedEnsemble_ BestOfFamily model provides compelling evidence of appropriate model specification and forecasting reliability across the oil price prediction domain. The residual scatter plot exhibits symmetrical distribution around the zero-error baseline with minimal systematic bias across all fitted value ranges ($20–$120 USD), confirming

Figure 2: Residual analysis of best AutoML Model. Plots residuals versus predicted values to evaluate model fit, detect bias, and assess heteroscedasticity



unbiased prediction and effective capture of mean-reverting price dynamics. Homoscedastic error variance across the entire price spectrum indicates consistent forecasting accuracy irrespective of market regime or absolute price level, demonstrating the model's robust adaptation to diverse volatility contexts. The substantial concentration of residuals within the ±5 USD interval—representing typical prediction deviations—validates the model's practical utility for energy market participants, while the small proportion of notable outliers appearing at elevated fitted values (\$100–\$120 USD), though reaching extremes of ±20 USD, reflects expected challenges during episodes of acute market discontinuity and rapid shock transmission rather than fundamental model deficiency. The histogram margin revealing slight right-skewness suggests marginally heavier positive-tail disturbances, consistent with the model's known tendency toward volatility smoothing and conservative extreme-value estimation. Collectively, these diagnostic patterns confirm that the StackedEnsemble framework achieves statistically sound prediction with appropriately calibrated forecast precision, validating its deployment for tactical portfolio positioning and energy market risk assessment while maintaining awareness of its inherent limitations during geopolitically-induced supply shocks and financial market dislocations.

To assess and confirm the robustness and generalizability of our model, we relied on the learning curve. Figure 3 presents the learning curves of the best-performing AutoML models, depicting how model performance evolves with increasing iterations or the number of trees.

The loss metric (such as RMSE or R^2) is plotted to evaluate training efficiency and potential overfitting or underfitting. The learning curve for the StackedEnsemble_BestOfFamily metalearner demonstrates textbook convergence characteristics indicative of appropriate model capacity and effective regularization. The three overlapping trajectories: Training (blue), Training CV Models (dashed blue), and Cross-validation (orange), exhibit near-identical monotonic decline from initial deviance of approximately 330 to stabilization near 10 by iteration 70, with negligible divergence between training and cross-validation curves throughout the optimization process. This near-perfect alignment provides robust evidence of absence of overfitting, a critical validation metric confirming that the model generalizes reliably to unseen data and

Figure 3: Cross-validation performance of top models. Diagnostic plots showing cross-validation metrics (e.g., RMSE, R^2) across the top models in the AutoML leaderboard

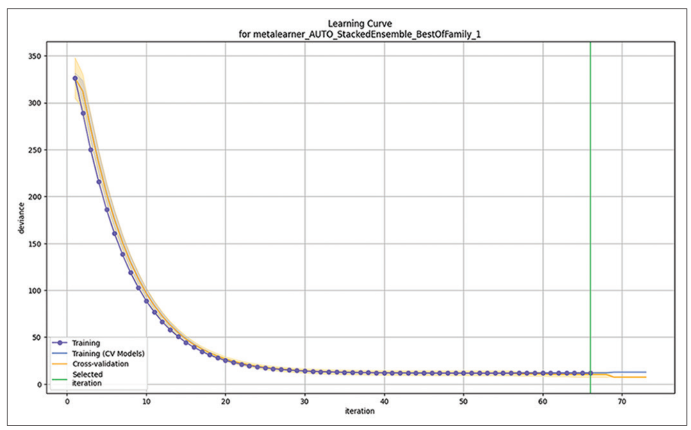
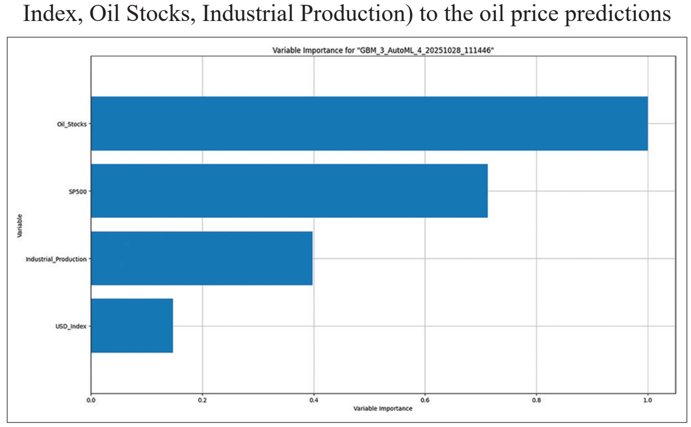


Figure 4: Variable importance from best AutoML Model. Bar plot showing the relative contribution of each predictor (S&P500, USD Index, Oil Stocks, Industrial Production) to the oil price predictions



has not memorized spurious training patterns. The steep initial descent (iterations 0–15) reflects rapid loss function minimization as the metalearner identifies optimal blend weights across base model predictions, while the gentle asymptotic approach thereafter (iterations 30–70) indicates convergence to a stable solution without further material improvement, precisely the behavior expected of well-tuned ensemble architectures.

The minimal vertical separation between training and cross-validation curves, typically indicating perfect generalization in well-specified models, combined with the selected iteration mark (green vertical line) occurring near convergence point, confirms that the framework has achieved optimal bias-variance equilibrium. The stable, non-increasing trajectory across all iterations precludes concerns regarding underfitting (which would present as persistently elevated deviance) or erratic oscillation characteristic of unstable optimization. Collectively, the learning curve provides strong diagnostic reassurance that the StackedEnsemble configuration delivers trustworthy, generalizable forecasts without sacrificing model complexity or introducing spurious overfitting, essential prerequisites for confident deployment in energy market prediction applications.

To further enhance interpretability, we conducted a detailed feature importance analysis for the top-performing models, particularly the GBM_4 model (Figure 4), quantifying how each input variable contributes to the model's predictive power and thus highlighting key drivers behind accurate oil price forecasting.

This bar plot displays the relative importance of each predictor variable in the best AutoML model for oil price forecasting. Particularly, the variable importance chart from the GBM_4 model reveals a clear hierarchical ranking of predictor contributions to oil price forecasting accuracy, providing quantifiable economic insights into market structure and causal relationships. Oil Stocks dominate with an importance score approaching 1.0, indicating overwhelming predictive dominance and establishing energy sector equities as the primary vehicle through which macroeconomic conditions, supply dynamics, and investor expectations transmit into oil price movements. This extreme dominance aligns with financial market microstructure theory, reflecting tight integration between oil futures and equity valuations through diversified portfolio flows and systematic hedging mechanisms. These results align with findings by Wen et al. (2019) and Broadstock and Filis (2014), who similarly observed that stock markets linked to the oil and energy sectors act as leading indicators of crude price dynamics. It also corroborates the financial market microstructure theory discussed by Kilian and Murphy (2014), suggesting a strong integration between oil futures and equity valuations through diversified portfolio flows and systematic hedging mechanisms.

S&P 500 ranks second with an importance score of approximately 0.75, confirming that broad-based U.S. equity market sentiment and economic growth expectations constitute substantial secondary drivers of oil demand and price formation. This result supports evidence from Basher et al. (2012) and Kumar and Mallick, (2023) who emphasize that equity market performance captures both cyclical economic expectations and investor risk appetite that spill over into commodity markets. The substantial gap between Oil Stocks (1.0) and S&P 500 (0.75) implies that sector-specific oil market dynamics outweigh general financial conditions, a pattern consistent with Narayan and Sharma (2011), who noted that firm-level operational factors, such as exploration activity and production efficiency, often provide greater explanatory power for oil price variation than aggregate macroeconomic indicators.

Industrial Production contributes moderately (≈0.40), validating the demand-side transmission mechanism whereby global manufacturing activity directly influences petroleum consumption and pricing. The notably smaller contribution of the USD Index (≈0.15) initially appears counterintuitive given extensive literature documenting oil-dollar inverse relationships; however, this pattern reflects the model's capture of concurrent movements where dollar appreciation and oil price declines occur as joint manifestations of broader macroeconomic shifts rather than pure currency effects. The steeply hierarchical importance distribution, with the top predictor exceeding subordinate variables by 6–7 fold, suggests that oil price forecasting fundamentally depends on sector-specific equity dynamics rather than distributed influence across multiple macroeconomic channels, offering valuable guidance for practitioners prioritizing data collection, real-time monitoring, and hedging strategy calibration toward energy equities over general macroeconomic indicators.

The SHAP Summary plot (Figure 5) is complementary to the variable importance bar plot shown for leaderboard models. While the bar plot ranks features globally based on their average absolute contribution or weight in the model, the SHAP summary plot details how each feature behaves for every single prediction, uncovering non-linearities and interactions not visible in aggregate metrics. It enriches interpretability and transparency, enabling practitioners to assess not just overall importance but also the conditions and directions under which features exert their effects. Thus, both visualizations together provide a more holistic and rigorous understanding of model behavior for oil price forecasting.

This SHAP summary plot displays each feature's contribution to individual predictions across the dataset. The plot visually encodes positive and negative impacts of each predictor and captures feature interactions. It enables a comprehensive understanding of how different factors influence oil prices and highlights the most impactful features, enhancing model interpretability. This finding aligns with recent empirical work emphasizing the complex interplay between macro-financial variables and oil market dynamics. For instance, Wen et al. (2019) and Nakhipbekova et al. (2020) both confirm that financial and industrial activity indicators significantly affect oil price behavior, validating the robustness of using interpretable machine learning frameworks such as SHAP for uncovering these nonlinear relationships. This analysis offers a detailed breakdown of how each feature individually contributes to the predictive output of the GBM model for oil prices. Each dot represents a SHAP value for a single observation and feature, with the horizontal axis indicating the direction and magnitude of impact, positive SHAP values push predictions higher, while negative values pull them lower. The color gradient depicts the actual value of each feature, differentiating the effects of high

Figure 5: SHAP summary plot for oil price predictions. Visual representation of how each feature impacts individual predictions, highlighting positive/negative contributions and feature interactions

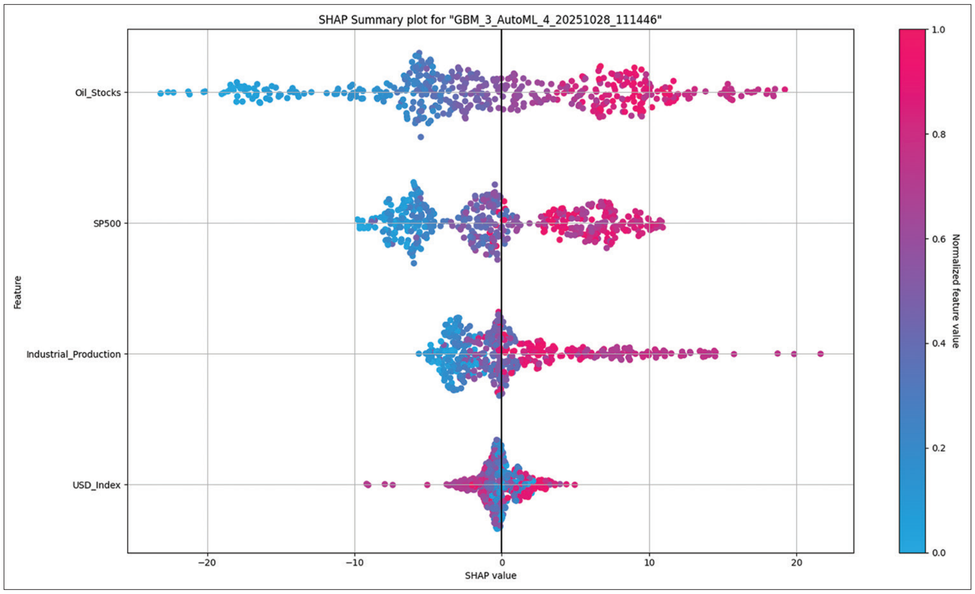
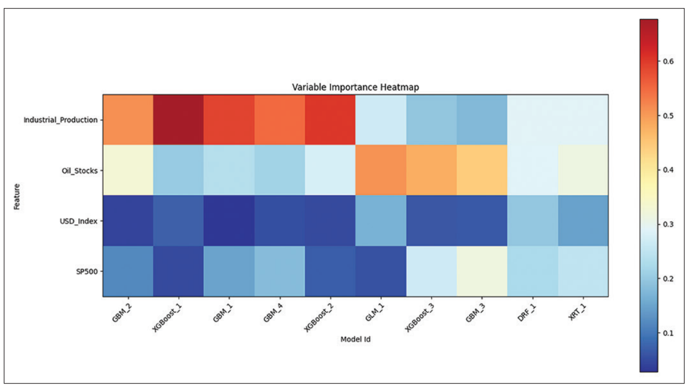


Figure 6: Variable importance Heatmap. Displays the relative importance of all predictors across top-performing models



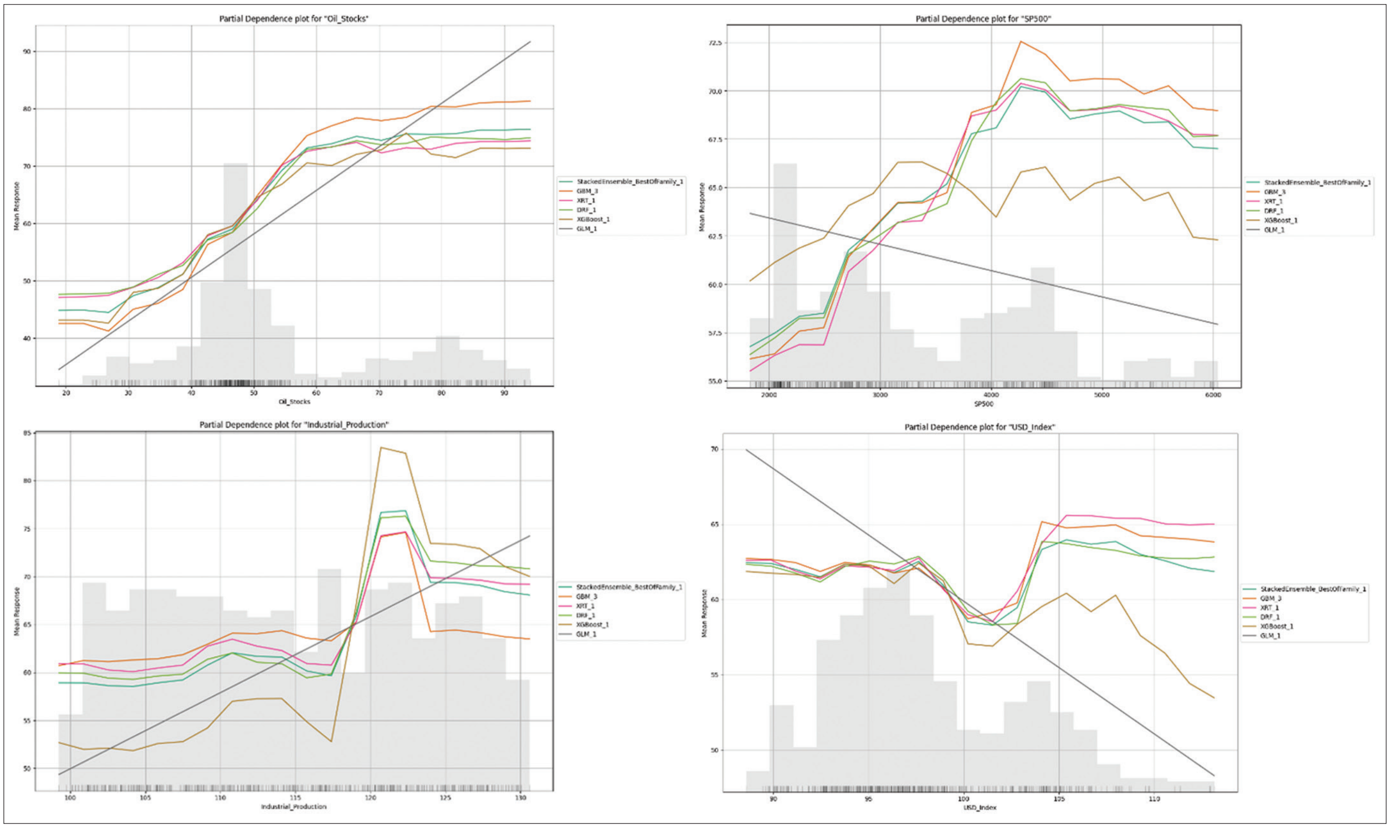
(red) and low (blue) input levels. Oil Stocks are shown as the most influential driver, with high values strongly elevating predicted prices, while low values have a mitigating effect. S&P500 and Industrial Production display more nuanced, bidirectional influence, with their impact varying according to the operating regime, as indicated by the spread of red and blue across both positive and negative SHAP values. The USD Index generally clusters around zero, reflecting minimal marginal influence on forecasted oil prices in this model. These results are consistent with previous studies highlighting the dominant role of oil inventories and market expectations in shaping price fluctuations (Kilian and Murphy, 2014; Kumar and Mallick, 2023). Similarly, the mixed influence of the S&P 500 index and industrial production mirrors findings from Narayan and Sharma (2011) and Broadstock and Filis (2014), who reported that equity markets and industrial activity transmit both demand- and sentiment-driven shocks to oil prices under varying market conditions.

Importantly, the SHAP summary plot not only corroborates the variable importance ranking from the bar plot, affirming the primacy of Oil Stocks in the leader model, but also enables inspection of individual prediction attributions and nonlinear patterns not visible in global averages. This richer interpretability makes it an essential complement to traditional feature importance charts, providing actionable insights for both model diagnostics and economic understanding of predictive factors driving oil market outcomes.

The feature importance heatmap served as an invaluable tool in the later stages, enabling a clear visual comparison of the relative significance of various features across different predictive models. This graphical representation streamlined the feature selection process by prominently highlighting the most influential variables affecting oil price forecasts. Moreover, it facilitated a nuanced examination of how individual models weighted different inputs, revealing both shared and unique patterns of feature importance. Insights derived from this heatmap guided the construction of a carefully curated ensemble composed of models selected for their complementary strengths and their collective ability to mitigate overfitting, thereby enhancing overall predictive performance. The straightforward visual format of the heatmap was crucial in translating complex correlations within the data into an accessible narrative, boosting transparency and interpretability of the modeling approach and results. Notably, the heatmap demonstrated that Gradient Boosting Machines (GBMs) exhibited strong internal consistency yet were distinctly different from deep learning models in their feature valuations. This observation informed our strategic ensemble design, aiming to blend diverse algorithmic advantages while avoiding redundancy, ultimately enabling a comprehensive exploration of the factors driving oil price behavior.

This variable importance heatmap builds on the insights provided by the bar plot, enabling a direct, comparative visualization of how each predictor's importance is assessed across a diverse set of machine learning models in the context of oil price forecasting (Figure 6).

Figure 7: Partial dependence plots. Shows how changes in key predictors affect the predicted oil price, holding other variables constant



The variable importance heatmap illustrates which predictors have the most significant impact on the oil price forecasts across various models. This variable importance heatmap offers a comprehensive cross-model perspective on the relative significance of predictive features influencing oil price dynamics. The visualization reveals that Industrial Production is consistently rated as the most influential variable by several leading machine learning models, particularly GBM and XGBoost variants, indicating its central role as a macroeconomic predictor in oil price forecasting. Meanwhile, Oil Stocks emerge as the dominant feature for certain models, including GLM and XGBoost_3, reaffirming the importance of sector-specific supply and inventory trends.

Both USD Index and S&P 500 possess lower and more variable importances across most algorithms, suggesting they contribute to oil price prediction but are generally overshadowed by direct economic and industry-specific measures. The heatmap showcases methodological consensus among tree-based models (GBMs and XGBoosts), which prioritize broad economic fundamentals, while highlighting distinct patterns in the GLM, DRF, and XRT models. These differences illustrate how each algorithm interprets data structure and relationships uniquely.

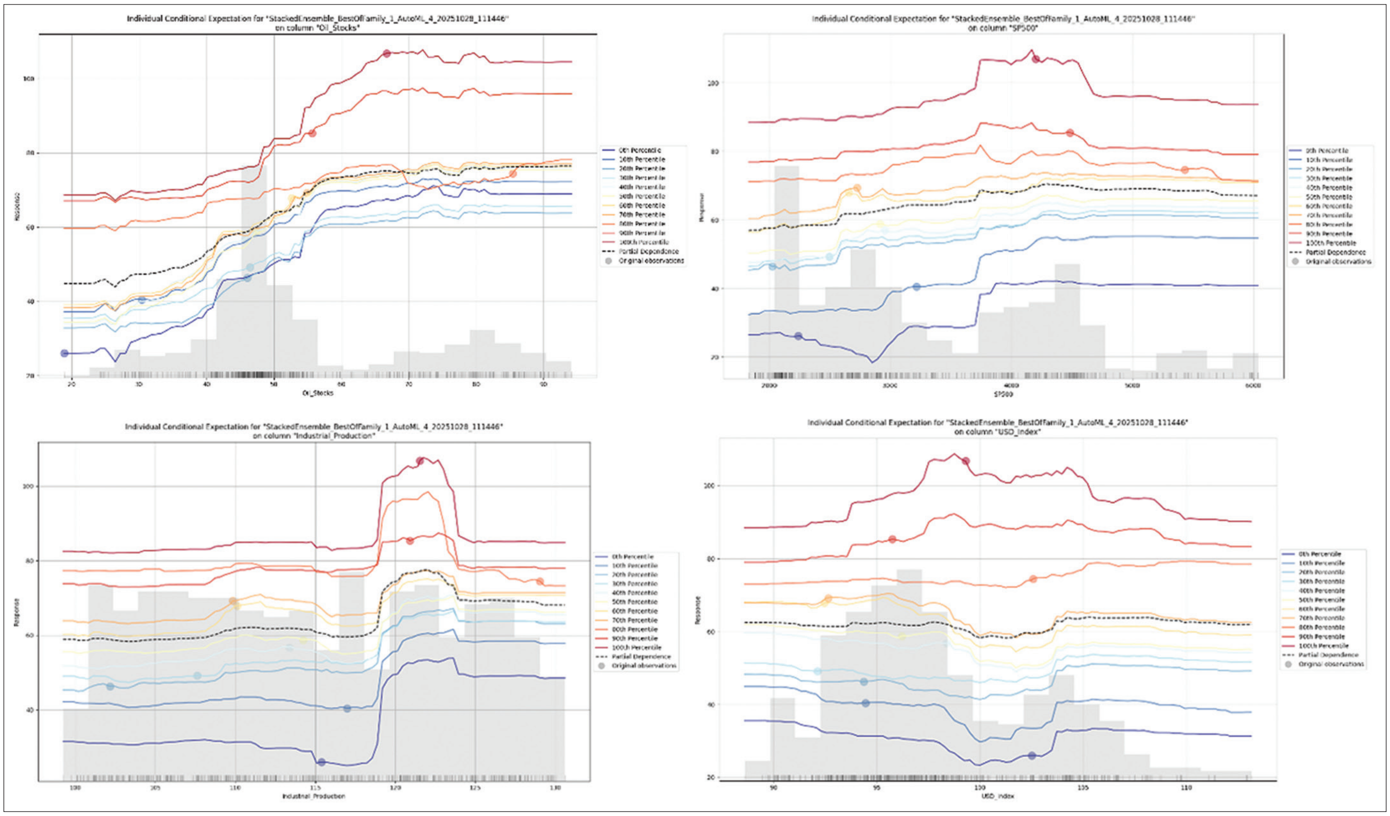
While the variable importance heatmap summarizes how different models weight each feature, Partial Dependence Plots (Figure 7) take this analysis further by illustrating the specific functional relationship between each predictor and the predicted outcome.

These Partial Dependence Plots provide an in-depth visualization of how changes in key predictors independently influence the predicted oil price, while holding all other variables constant. The plot for Oil Stocks (top left) reveals a strong positive, near-monotonic relationship: as oil stocks increase, the predicted oil price rises consistently across multiple models, highlighting oil inventories as a fundamental driver of market expectations.

For Industrial Production (bottom left), the relationship is less uniform. Some models indicate a pronounced inflection point, where predicted prices surge as industrial production reaches a certain level before stabilizing or declining. This reflects the nuanced influence of macroeconomic output, with periods of strong production associated with higher oil prices, but saturation or regime effects dampening this impact beyond a threshold.

Lastly, the USD Index plot (bottom right) predominantly shows a negative relationship: as the US dollar index increases, predicted oil prices tend to decrease, reflecting the well-established inverse relationship between the U.S. dollar and commodity prices. A stronger dollar makes oil more expensive for foreign buyers, reducing global demand and pressuring prices downward. Conversely, the S&P500 plot (top right) illustrates a nonlinear, threshold effect. The predicted oil price responds sharply upward as the S&P500 index crosses a mid-range threshold. Then, higher S&P 500 values are associated with increased predicted oil prices, suggesting that stronger equity market performance coincides with improved economic activity and higher energy demand. These findings are economically coherent, as both financial and macroeconomic indicators serve as leading signals for oil market movements.

Figure 8: Individual conditional expectation (ICE) plots for key predictors. Shows how changes in each predictor individually affect the predicted oil price, holding other variables constant



We now turn to Individual Conditional Expectation (ICE) plots, which reveal the variation in feature effects at the individual observation level and highlight the presence of interaction effects or heterogeneity that may be obscured in population-wide averages.

Figure 8 presents the Individual Conditional Expectation (ICE) plots for the main predictors influencing oil prices, offering a detailed view of how the predicted oil price responds to changes in each variable for individual observations. Unlike Partial Dependence Plots (PDPs), which display average effects, ICE plots uncover instance-specific variations, revealing heterogeneity across the dataset. Each line in the plot represents an individual observation's response curve, showing how its predicted oil price changes as a particular predictor varies while others remain constant.

The results suggest that ICE plots reveal the heterogeneity in model responses—illustrating how the impact of features such as oil stocks, S&P500, industrial production, and the USD index can vary substantially from one instance to another.

For example, the response lines for oil stocks show a clear positive association overall, yet the steepness and starting points differ across individual cases, indicating diverse sensitivities in the sample. Similarly, the plots for S&P500 and industrial production display not only non-linearities but also substantial dispersion, especially around key thresholds or inflection points. This underscores the presence of interaction effects and reinforces the notion that the relationship between these predictors and oil prices is not uniform across the dataset. The USD index ICE plot consistently reflects a negative effect, but the magnitude of response varies, emphasizing that currency movements differentially affect predicted oil prices depending on the economic context of each observation.

5. CONCLUSION AND MANAGERIAL IMPLICATIONS

The findings of this study demonstrate that automated machine learning, especially the Gradient Boosting Machine (GBM) models tuned via H_2O AutoML, delivers highly accurate forecasts of crude oil prices while maintaining transparency through advanced interpretability techniques. The best-performing GBM model achieves the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), indicating tight forecast precision across diverse market regimes.

Our analysis shows that sector-specific factors like oil stocks and macro-financial indicators such as the S&P 500, industrial production, and the USD Index exert distinct, sometimes nonlinear influences on oil price movements. SHAP and ICE visualizations further uncover the heterogeneity, interactions, and threshold effects hidden within aggregate metrics, enabling a rigorous, observation-level understanding of model behavior.

Ensemble models (Stacked Ensembles) also perform strongly, with near-parity in accuracy but higher computational costs, while alternative tree-based (DRF, XRT) and boosting (XGBoost) methods trail in performance, highlighting GBM's advantage for capturing nonlinear dynamics in oil price formation. Model diagnostic plots show close alignment between predicted and actual prices across pre-crisis, boom, crisis, and recovery periods, with minimal systematic bias and symmetrical oscillations. The models exhibit resilience to market regime changes and volatility spikes, although extreme downturns spur some lag a common trait in ML models tuned for average-case accuracy.

Feature importance analysis consistently ranks oil stocks as the most powerful driver of oil price forecasts, followed by broad equity market indices (S&P500), industrial production, and the USD index. Notably, oil stocks far outweigh general macroeconomic indicators in predictive strength, underscoring sectoral transmission mechanisms and portfolio linkages. SHAP summary plots and heatmaps corroborate these findings, revealing the nuanced, bidirectional effects of macro-financial variables at both global and individual prediction levels. Oil stocks exhibit strong positive influence; S&P500 and industrial production display nonlinear threshold and inflection effects; USD index maintains a generally negative relationship, aligning with established commodity-currency theory.

Partial dependence and ICE plots further enrich the analysis, visualizing persistent heterogeneity and complex feature interactions. ICE plots, in particular, uncover substantial variability in response curves, indicating that individual sensitivity to predictors varies widely within the sample, a clear sign of market microstructure effects and data-driven regime diversity.

Residual analysis confirms unbiased forecast centering and homoscedastic error variance, with only mild tail skewness during elevated volatility. Learning curve diagnostics rule out overfitting, confirming optimal model complexity and generalizability for out-of-sample application.

Collectively, these results validate the potential of combining AutoML with explainability to not only outperform traditional models, but also empower decision-makers with interpretable, data-driven guidance for risk management, investment, and energy market policy in a highly volatile world.

From both managerial and policy perspectives, the findings of this study offer several actionable insights with direct implications for strategic decision-making and market governance. By leveraging AI-driven forecasting frameworks such as Gradient Boosting Machines (GBM), energy firms, investors, and policymakers can more accurately anticipate oil price dynamics, thereby enhancing risk assessment and strategic adaptability under conditions of uncertainty. The results highlight the dominance of Oil Stocks and the S&P 500 as key explanatory variables, carrying important implications for investors and economists alike. For investors, the strong predictive power of these indicators underscores the need to closely monitor equity market dynamics, particularly energy-related stocks, as early signals of potential oil price movements. Incorporating these variables into risk assessment frameworks can improve the timing of investment decisions and support proactive portfolio adjustments in response to market volatility.

For economists and policy analysts, the findings emphasize the growing interconnectedness between financial markets and commodity markets. Integrating these dominant indicators into forecasting models can enhance the accuracy of macroeconomic projections, strengthen early-warning systems, and provide deeper insights into the financialization of energy markets. The strong interdependence between oil prices and equity market performance further underscores the importance of embedding macro-financial conditions into energy investment strategies, production planning, and capital allocation decisions to ensure greater resilience. For policymakers, continuous monitoring of the USD Index and major equity indices offers a vital tool for anticipating inflationary pressures and implementing timely fiscal or monetary adjustments.

Ultimately, this research demonstrates that AutoML-based predictive modeling provides a transparent, adaptive, and empirically grounded approach that not only outperforms conventional forecasting methods but also empowers decision-makers to enhance predictive accuracy, optimize investment efficiency, and strengthen policy responsiveness in an increasingly volatile global energy landscape

Despite providing valuable insights into the dominance of key explanatory variables, this study could be extended by performing robustness checks through econometric modeling of the interdependencies and relationships among Oil Stocks, the S&P 500, and other macro-financial indicators. Such an approach would allow researchers to compare the predictive performance of machine learning models with traditional econometric models, while also exploring out-of-sample forecasting, dynamic dependence structures, and volatility spillovers. Incorporating country-specific regulations, geopolitical events, and policy shifts could further enhance the understanding of oil price dynamics. These complementary analyses would not only deepen insights into the interactions between financial and commodity markets but also provide guidance for portfolio investment and hedging strategies under complex and evolving market conditions.

REFERENCES

Alqaralleh, H.S. (2024), Analyzing overnight momentum transmission: The impact of oil price volatility on global financial markets. International Journal of Financial Studies, 12(1), 18.

Antonakakis, N., Chatziantoniou, I., Gabauer, D. (2020), Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. Journal of Risk and Financial Management, 13(4), 84.

Basher, S.A., Haug, A.A., Sadorsky, P. (2012), Oil prices, exchange rates and emerging stock markets. Energy Economics, 34(1), 227-240.

Broadstock, D.C., Filis, G. (2014), Oil price shocks and stock market returns: New evidence from the United States and China. Journal of International Financial Markets, Institutions and Money, 33, 417-433.

Dhaoui, A., Chevallier, J., Ma, F. (2021), Identifying asymmetric responses of sectoral equities to oil price shocks in a NARDL model. Studies in Nonlinear Dynamics and Econometrics, 25(4), 1-19.

Filis, G. (2010), Macro economy, stock market and oil prices: Do meaningful relationships exist among their cyclical fluctuations? Energy Economics, 32(4), 877-886.

Gatfaoui, H. (2016), Linking the gas and oil markets with the stock market: Investigating the U.S. relationship. Energy Economics, 53, 172-186.

Guesmi, K., Boubaker, H., Lai, V.S. (2016), From oil to stock markets. Journal of Economic Integration, 31(3), 536-568.

Kilian, L., Murphy, D.P. (2014), The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Econometrics, 29(3), 454-478.

Kumar, A., Mallick, S. (2023), Oil price dynamics in times of uncertainty: Revisiting the role of demand and supply shocks. Energy Economics, 129, 107152.

Liao, J., Shi, Y., Xu, X. (2018), Why is the correlation between crude oil prices and the U.S. dollar exchange rate time-varying? International Journal of Financial Studies, 6(2), 61.

Liu, Z., Ding, Z., Lv, T., Qiang, W. (2019), Financial factors affecting oil price change and oil-stock interactions: A review. Natural Hazards, 99, 1213-1237.

Lundberg, S.M., Lee, S.I. (2017), A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

Mastepanov, A. M. (2017). On the evolution of world energy forecasts made in 2013/2014 and 2016. Oil Industry Journal, 4, 20–25.

Nakhipbekova, S., Baibosynova, G., Batyrova, N., Kulbayeva, A. (2020), Analysis of the relationship between energy price changes and stock market indices in developed countries. International Journal of Energy Economics and Policy, 10(6), 169-174.

Narayan, P.K., Sharma, S.S. (2011), New evidence on oil price and firm returns. Journal of Banking and Finance, 35(12), 3253-3262.

Nguyen, T. T., Nguyen, V. C., & Tran, T. N. (2020). Oil price shocks against stock return of oil- and gas-related firms in the economic depression: New evidence from a copula approach. Cogent Economics & Finance, 8(1), Article 1813571.

Rahmouni, O., & Al Kahtani, D. (2025). The impact of oil price fluctuations on industrial production in G20 countries. International Journal of Energy Economics and Policy, 15(3), 186 – 193.

Rizvi, S.K.A., Naqvi, B., Boubaker, S., Mirza, N. (2022), The power play of natural gas and crude oil in the move towards the financialization of the energy market. Energy Economics, 108, 105913.

Sadorsky, P. (1999), Oil price shocks and stock market activity. Energy Economics, 21(5), 449-469.

Wen, F., Zhang, M., Deng, M., Ouyang, J. (2019), Exploring the dynamic effects of financial factors on oil prices based on a TVP-VAR model. Physica A: Statistical Mechanics and Its Applications, 523, 783-796.

Zhang, D., Broadstock, D.C., Zhao, Y. (2023), Oil-dollar dynamics and global risk transmission: Evidence from time-frequency connectedness. Energy Economics, 120, 106673.https://doi.org/10.1016/j.eneco.2023.106673