Predicting Crude Oil Price Direction Using Machine learning

Student Researchers

Katie Schuman | Aarambh Shrestha

Faculty Advisor

Dr. Mansi Bhavsar

Why Predicting Oil Price Direction Matter

The oil market is a global infrastructure where crude oil is extract, traded refined and distributed world wide, serving as the backbone of the energy, transport and industrials products. Its importance originated from its roll in powering the world economies making it highly sensitive to geopolitics and financial market.

1990 GULF WAR

2008 FINANCIAL CRISIS

2014 OPEC OVERSUPPLY

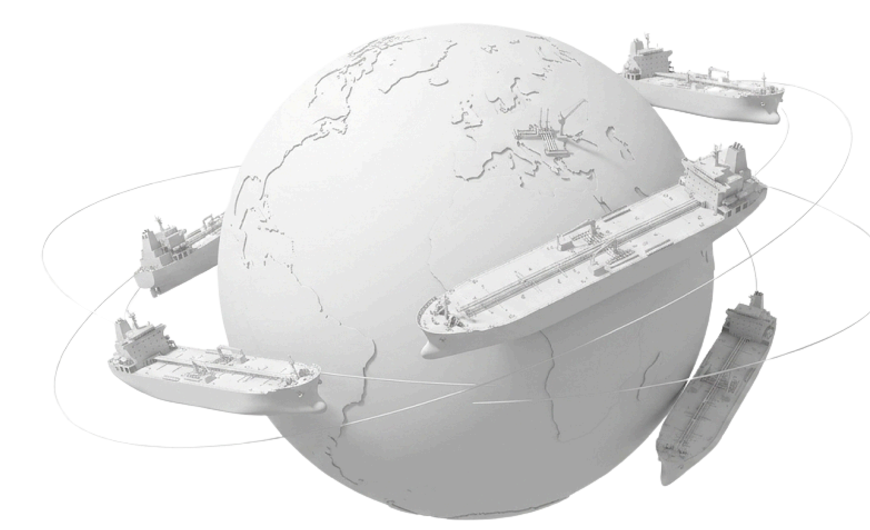
2020 COVID-19

Global Consumption	Global Trade	West Texas Intermediate
100 Million Barrels	6 Billion Barrels	93.38 USD Barrel
DAILY	DAILY	TODAY

Gulf War	Financial Crisis	Covid 19
+130% Per Barrels	-77% Per Barrels	-160% Per Barrels
PRICE INCREASE	PRICE DECREASE	PRICE REACHED NEGATIVE

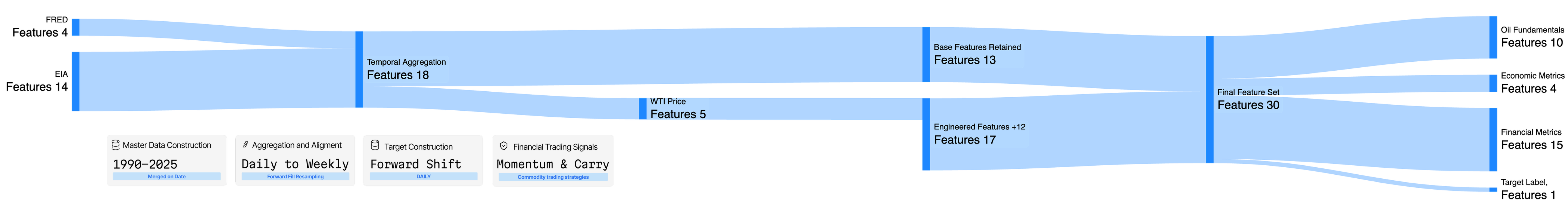
Challenges In Predicting Oil Price Direction

Crude oil price direction is no longer explained by supply and demand alone. Prices are shaped by conditions across the physical oil market, the wider economic and geopolitical environment, and activity in financial markets. Hedging activity and oil-related financial instruments add further complexity by influencing how the market reacts over time. Together, these overlapping dynamics make price direction harder to interpret and forecast clearly.



Data Pipeline Arcitecture

We built a structured data pipeline architecture to address the challenge of forecasting the direction of WTI crude oil prices. We began by constructing a custom dataset from two major sources, enhanced with feature engineering based on three key pillars: oil fundamentals, economic indicators, and financial metrics. Since these inputs varied in frequency daily, weekly, monthly. We unified them through temporal aggregation into a consistent weekly dataset spanning 35 years. From this unified dataset, we developed two streams of model architectures one flowing into machine learning and the other into deep learning. We compare and evaluate forecasts, benchmarking them to drive ongoing improvement.



Master Data Construction	Aggregation and Aligment	Target Construction	Financial Trading Signals
1990–2025	Daily to Weekly	Forward Shift	Momentum & Carry
Merged on Date	Forward Fill Resampling	DAILY	Commodity trading strategies

Machine Learning

	Accuracy	AUC	F1 Score	Precision	Recall
Logistic Regression	58.0%	53.7%	65.6%	59.4%	73.2%
Decision Tree	51.7%	51.2%	64.8%	53.8%	81.3%
Gradient Boosting	55.2%	56.1%	64.3%	57.0%	73.7%
Random Forest	58.6%	60.0%	61.7%	62.4%	61.1%

Feature Selection	Train / Test Split
10 Best Features	Chronological
EXHAUSTIVE FEATURE SELECTION	First 80% → Training | Final 20% → Test

Standardization	Tuning
Z-Score	6 Max Dept
Logistic Regression	Random Forest

We started of with running four different models to compare how each performed on the weekly feature set using the same time-ordered 80/20 split. This allowed us to evaluate them under a consistent setup while keeping the sequence of the data intact. Logistic Regression was built in a standardized pipeline, while the other models were run with fixed configurations so the comparison stayed direct, fair, and easy to interpret.

Deep Learning

	Accuracy	AUC	F1 Score	Precision	Recall
MLP	56.5%	50.6%	72.1%	56.4%	100.0%
RNN	53.6%	54.8%	58.2%	59.0%	57.5%
LSTM	55.3%	49.5%	70.9%	55.9%	97.0%

Hidden Layers	Neurons per Layer
5 layers	128,64,32,32,16
3 LSTM layers + 2 dense layers	LSTM + Dense

Activation Functions	Optimization Setup
ReLU,Sigmoid	Adam,BCE
Hidden / Output	Learning Rate 0.0005

Moving on to deep learning to assess whether a deeper architecture could better capture patterns in the data and how it compared with the earlier models. The final architecture used three stacked LSTM layers, followed by dense layers and a sigmoid output, with ReLU in the dense hidden layers and binary cross-entropy as the loss function. Training was carried out using Adam with a learning rate of 0.0005, batch size 32, and early-stopping controls.

Findings

Tree Based Models are High-Performing

Tree based machine learning models modestly outperform traditional linear baseline and other classical models. Preliminary deep learning models' performances are on par with the highest performing machine learning models.

Derivatives Signals are Strong

Futures-derived and market-based features appear promising, consistent with their forward-looking nature.

Uncaptured Market Complexity

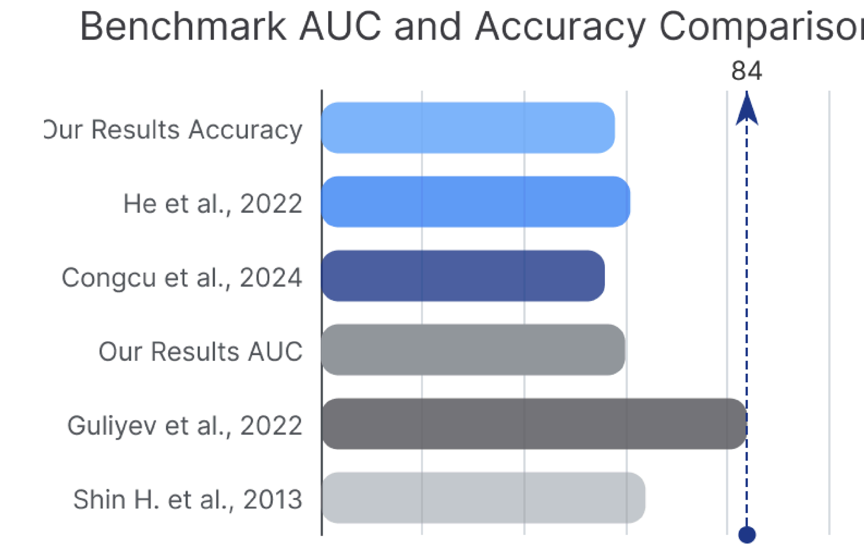
The predictive signal remains limited, suggesting deeper market dynamics are not yet fully incorporated.

Conclusions

While the performance may appear modest at first glance, it should be interpreted in the context of the forecasting task. Predicting WTI crude oil price direction is inherently difficult, shaped by complex and overlapping market forces, so results should not be judged only in absolute terms. In that setting, the findings remain encouraging, as they suggest the models are capturing meaningful signal and moving in a positive direction.

58% ACCURACY OUR RESULT | **56%** ACCURACY Göncü et al., 2024 | **61%** ACCURACY He et al., 2022

60% AUC OUR RESULT | **64%** AUC Shin H. et al., 2013 | **84%** AUC Guliyev et al., 2022 (Reference Paper)



Future Work

Future work will focus on model-specific tuning across both machine learning and deep learning approaches. This includes refining layer structure, activation functions, split configuration, regularization, and optimization settings for each architecture individually. With a common baseline now established, the next step is to improve each model in a more targeted and systematic way.

There is also scope to revisit preprocessing decisions if further testing suggests that certain transformations or engineered features are limiting performance; as well as isolating features or performing a shap analysis to determine feature importance. Geopolitical and other non-linear event-based features could be added to the dataset as well. A longer-term extension is to move beyond directional classification and explore prediction of the magnitude of price change. Overall, the expectation is that more focused tuning and refinement will strengthen the deep learning models and improve their performance relative to the earlier machine learning baseline.