Advanced Topics in DS//Deep Learning

Oil Spot Prices with Transformers

Richie Zhang
Rohan Sanghrajka
Uday Sapra

December, 2024

Executive Summary

Problem Statement: Oil prices are highly volatile, driven by complex, interdependent factors like macroeconomic shifts, market sentiment, and geopolitical events. Traditional forecasting methods fail to capture this complexity, leading to suboptimal decisions and increased risks for industries, investors, and policymakers.

Goal: Develop a predictive model integrating time series analysis, economic indicators, and sentiment data to improve oil price predictions.

Solution

- Transformer-based models (PatchTST, TFT) outperforming traditional benchmarks.
- Optimized predictions through feature engineering and sentiment analysis.

Value: Improved short-term forecasting supports strategic decision-making and risk management.

Motivation

The Problem: Oil prices impact global economics, yet accurate forecasting remains elusive.
The Challenge: Traditional models miss the nuances of complex, non-linear relationships.
The Opportunity: Innovative transformer-based approaches integrating sentiment and macroeconomic data.

Crude Oil Price Trends Over the Past Year

Economic Variables + Sentiments + Engineered Features → Crude Oil Prices

Related Works

J. Stat. Appl. Pro. 11, No. 3, 845-855 (

Forecasting Crude (

Hleil Alrweili[1] and Haitham Fawz

[1] Department of Mathematics, Fac
[2] Department of Applied Statistics

Received: 5 Sep. 2021, Revised: 1 Nov
Published online: 1 Sep. 2022

Abstract: This paper aims to use
Integrated Moving Average (ARI
is flexible enough to capture two
nonlinear that can only model the
Saudi Riyal per Barrel was used d
215 observations are used as trai
Mean Square Error (MSE), Mean
combination ARIMA and ANN
improvement in (MSE, MAPE, an

Keywords: Time Series, ARIMA,

1 Introduction

Time series data can define as a s
these observations. (Every minute
areas produce huge amount of ti
wind speed in the meteorologica
renewable power production and
heartrate, pulse, and electrocardio
domain [2-6].

Time series prediction uses metho
and important to make good pre
determining the routes of cruise sl
considering costs requires the elec
line requires the forecasts of the c
the forecasts of the number of us
area where time series is used. Fo
ago; it is necessary to obtain day-
capital. As a result, a good forec
effective and efficient way for dec

2 Related Works

In the literature, there are a small
diversified areas of economic,
of heating period and non-heatin
meteorological types which are li

*Corresponding author e-mail: haith

CityU
香港城市大學
City University of Hong Kong

Crude oil price prediction using t

He, Kaijian; Zheng, Linyuan; Yan

Published in:
Procedia Computer Science

Published: 01/01/2023

Document Version:
Final Published version, also known as

License:
CC BY-NC-ND

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1016/j.procs.2023.08.070

Publication details:
He, K., Zheng, L., Yang, Q., Wu, C., Yu
transformer model. Procedia Computer

Citing this paper
Please note that where the full-text provided
Manuscript, Peer-reviewed or Author Final v
you check and use the publisher's definitive

General rights
Copyright for the publications made accessi
copyright owners and it is a condition of acc
requirements associated with these rights. U
or commercial gain.

Publisher permission
Permission for previously published items a
RoMEO database. Links to full text versions
allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you belie
remove access to the work immediately and

Download date: 16/12/2024

Annals of Operations Research
https://doi.org/10.1007/s10479-023-05810-8

ORIGINAL RESEARCH

A blending ensemble learning model for crude oil price forecasting

Mahmudul Hasan[1] · Mohammad Zoynul Abedin[2] · Petr Hajek[3] · Kristof Coussement[4] · Md. Nahid Sultan[1] · Brian Lucey[5]

Received: 4 January 2023 / Accepted: 19 December 2023
© Crown 2024

Abstract
To efficiently capture diverse fluctuation profiles in forecasting crude oil prices, we here propose to combine heterogenous predictors for forecasting the prices of crude oil. Specifically, a forecasting model is developed using blended ensemble learning that combines various machine learning methods, including k-nearest neighbor regression, regression trees, linear regression, ridge regression, and support vector regression. Data for Brent and WTI crude oil prices at various time series frequencies are used to validate the proposed blending ensemble learning approach. To show the validity of the proposed model, its performance is further benchmarked against existing individual and ensemble learning methods used for predicting crude oil price, such as lasso regression, bagging lasso regression, boosting, random forest, and support vector regression. We demonstrate that our proposed blending-based model

✉ Mohammad Zoynul Abedin
m.z.abedin@swansea.ac.uk

Mahmudul Hasan
mahmudulmoon123@gmail.com

Petr Hajek
petr.hajek@upce.cz

Kristof Coussement
k.coussement@ieseg.fr

Md. Nahid Sultan
nahid.sultan@hstu.ac.bd

Brian Lucey
brian.lucey@tcd.ie

1 Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur 5200, Bangladesh
2 Department of Accounting and Finance, School of Management, Swansea University, Fabian Way, Bay CampusSwansea, Wales SA1 8EN, UK
3 Science and Research Centre, Faculty of Economics and Administration, University of Pardubice, 532 10 Pardubice, Czech Republic
4 IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, Lille F-59000, France
5 Trinity Business School, Trinity College Dublin, Dublin 2, Ireland

Published online: 25 January 2024

Springer

1. **ARIMA Models**
   - Strength: Simple, interpretable, effective for linear trends.
   - Limitation: Cannot handle non-linear relationships or multivariate data.
2. **Blending Ensemble Models**
   - Strength: Combines multiple algorithms for improved accuracy.
   - Limitation: Struggles in volatile markets; lacks advanced feature integration.
3. **Temporal Fusion Transformer (TFT)**
   - Strength: Excellent for long-term multivariate forecasting with exogenous features.
   - Limitation: Complex, prone to overfitting on small datasets; underperforms in noisy short-term data.

Method: Dataset

30 Total Variables

1. Economic Variables
2. Sentiment Scores
3. Engineered Future Features for Decoder

Method: Dataset

Economic Variables

S&P 500
Reflects overall economic health and investor sentiment.

Interest Rates
Influence on economic growth and capital investments.

VIX/ OVX
Indicators of market volatility and expectations of oil price fluctuations.

USO
Tracks daily WTI crude oil prices and market sentiment.

DXY
Strength of USD inversely affects global oil pricing.

Crude Oil
Historical spot prices provide essential data for identifying trends and forecasting.

Method: Dataset

Sentiment Scores

1. Scraped oil related headlines from 6 news sources
2. Preprocessed text data
3. Obtain sentiment scores using pretrained BERT model

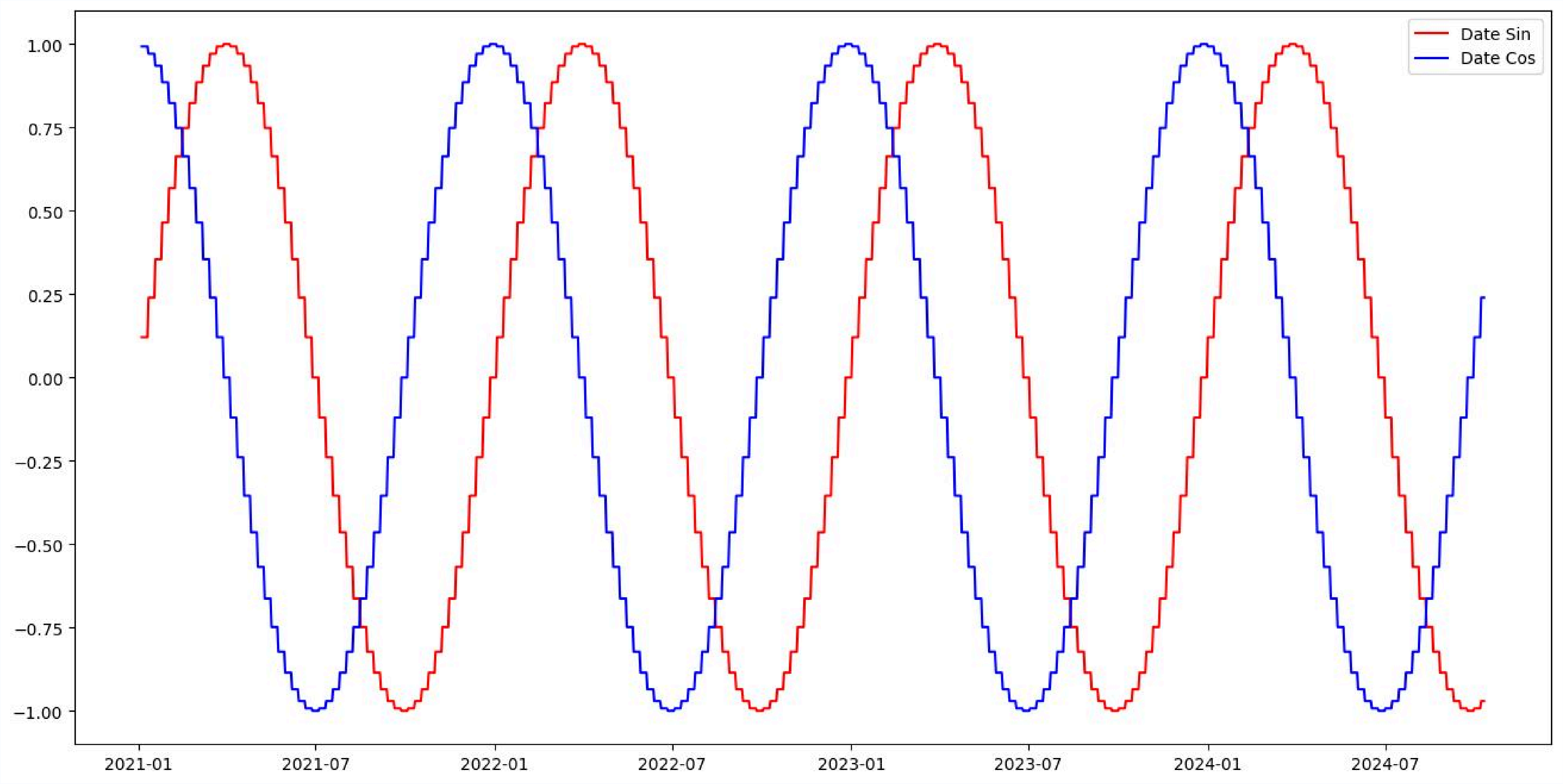
Method: Dataset

Engineered Future Features

Date variables modeled as functions of Sin and Cos

Lagged target variables:

- TargetLag(7)
- TargetLag(30)
- TargetLag(365)



Teaching Models Seasonality

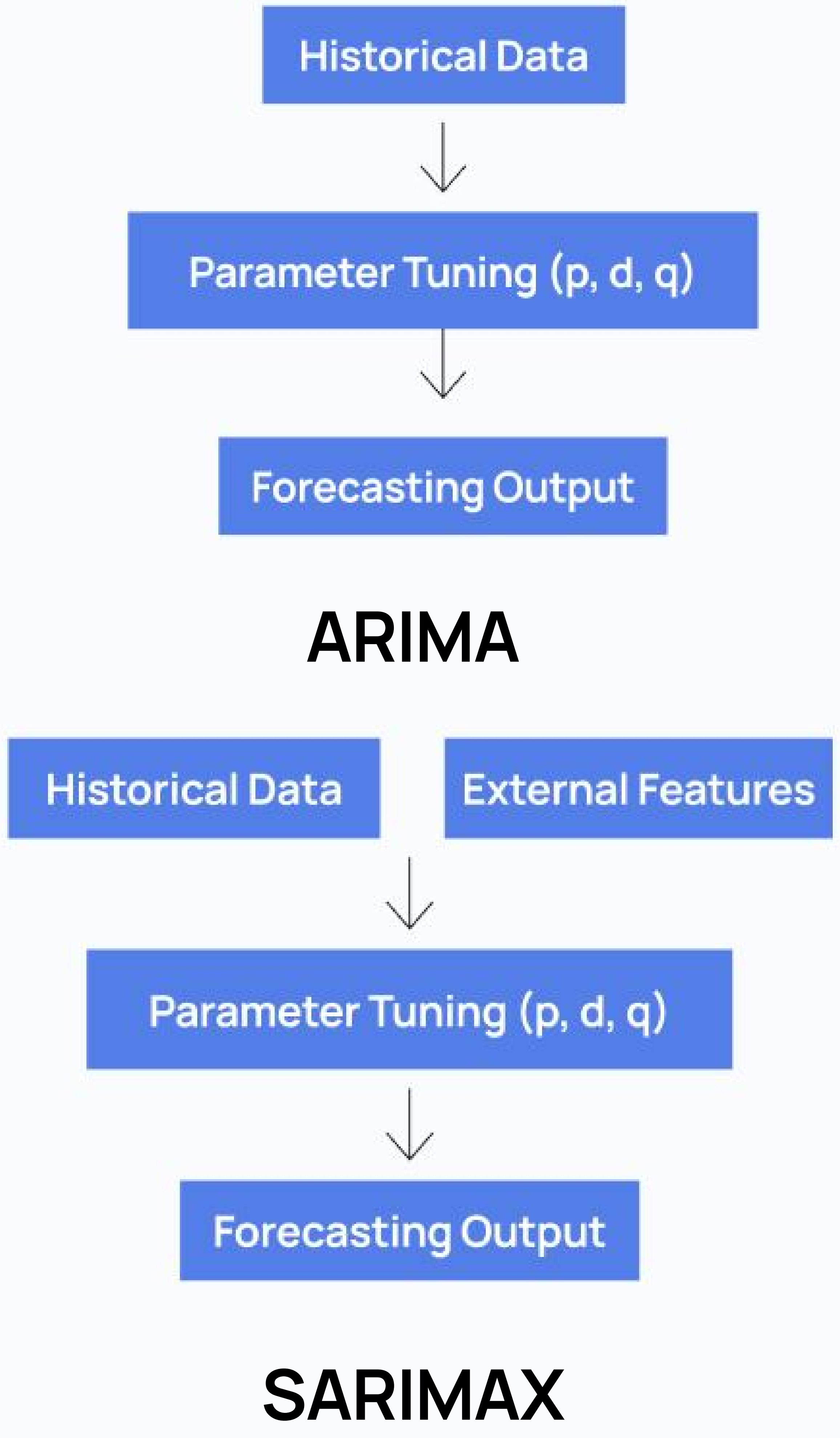
ARIMA & SARIMAX Benchmark Statistical Models

ARIMA (AutoRegressive Integrated Moving Average)

- Models univariate time-series trends using past values.
- Uses (p, d, q): autoregressive, differencing, and moving average terms.

SARIMAX (Seasonal ARIMA with Exogenous Variables)

- Extends ARIMA with external variables (e.g., S&P 500) and seasonal patterns.
- Key Features:
  - Exogenous Inputs: Includes economic indicators.
  - Seasonality: Adds terms for cyclical trends (P, D, Q, s).



Method: Patch TST

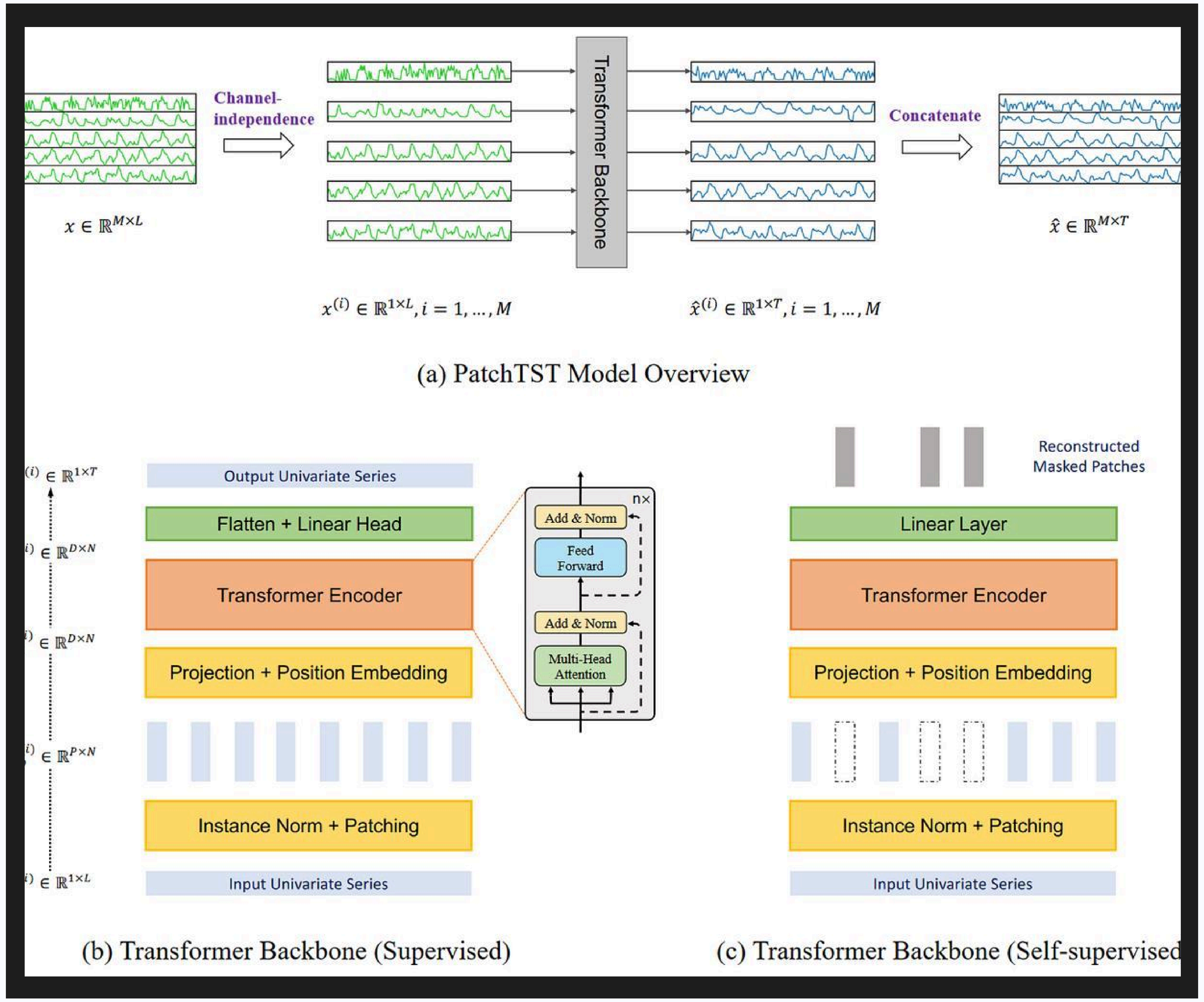
Purpose: Autoregressive transformer-based model for time-series forecasting.

Approach:

- Divides historical data into patches to capture long-term dependencies.
- Attention mechanism focuses on relevant time steps.

Technical Details:

- Input: Historical oil prices only.
- Key Hyperparameters: Patch Length, Attention Heads, Layers.
- Optimized using Optuna.



Method: Temporal Fusion Transformer (TFT)

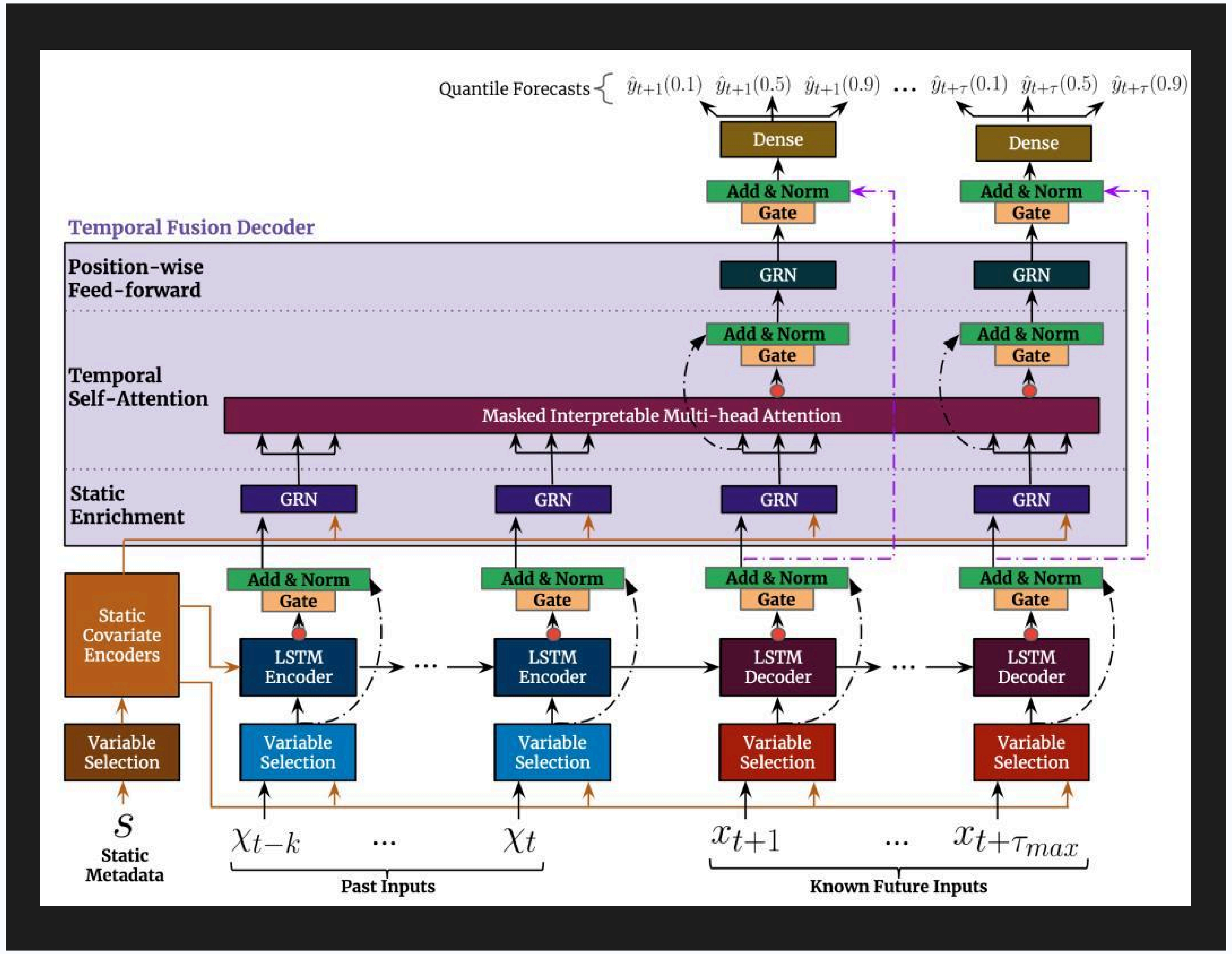
Purpose: Multivariate transformer model that incorporates external features.

Approach:

- Combines historical oil prices and exogenous variables (S&P 500, Sentiments, OVX, etc.).
- Variable selection + Attention + LSTM decoder for prediction.

Technical Details:

- Input: Lagged targets, economic indicators, and sentiment scores.
- Hyperparameters: Hidden Size, Attention Heads, Dropout.



Implementation

Optimization Method:

- Optuna for hyperparameter tuning.
- Forecast horizons fixed at 16, 48, 96 steps for short, medium, and long-term predictions.

PatchTST

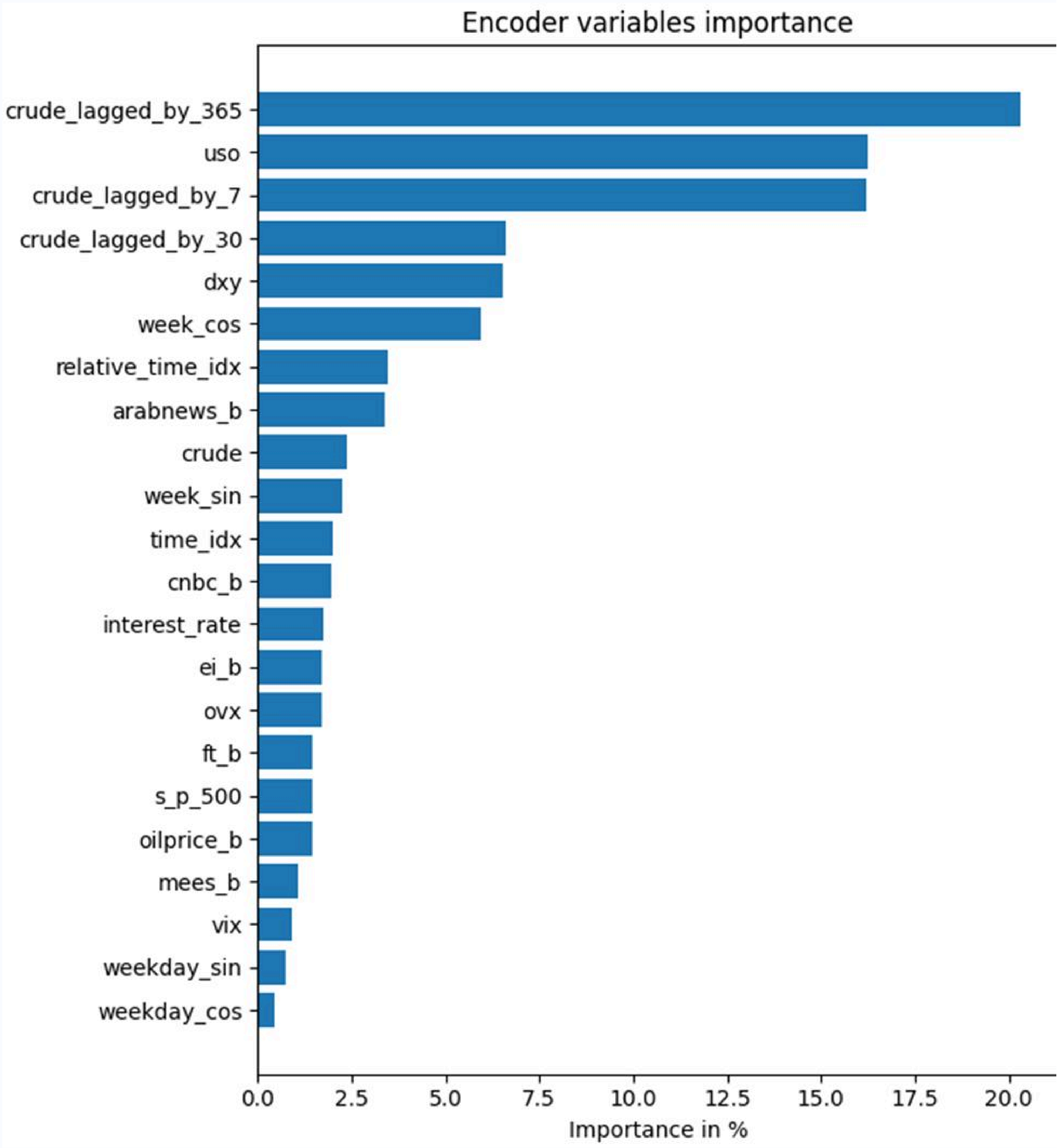
- Forecast History: 96
- Attention Heads (n heads): 1
- Attention Dropout: 0.1
- Embedding Dimension (d model): 120
- Patch Length: 96
- Dropout: 0.3
- Encoder Layers (n layers): 3
- Feedforward Network Dimension (dff): 152
- Batch Size: 24
- Epochs: 20
- Learning Rate: Cyclical (tuned between lr min and lr max).
- Stride: 2

TFT (Temporal Fusion Transformer)

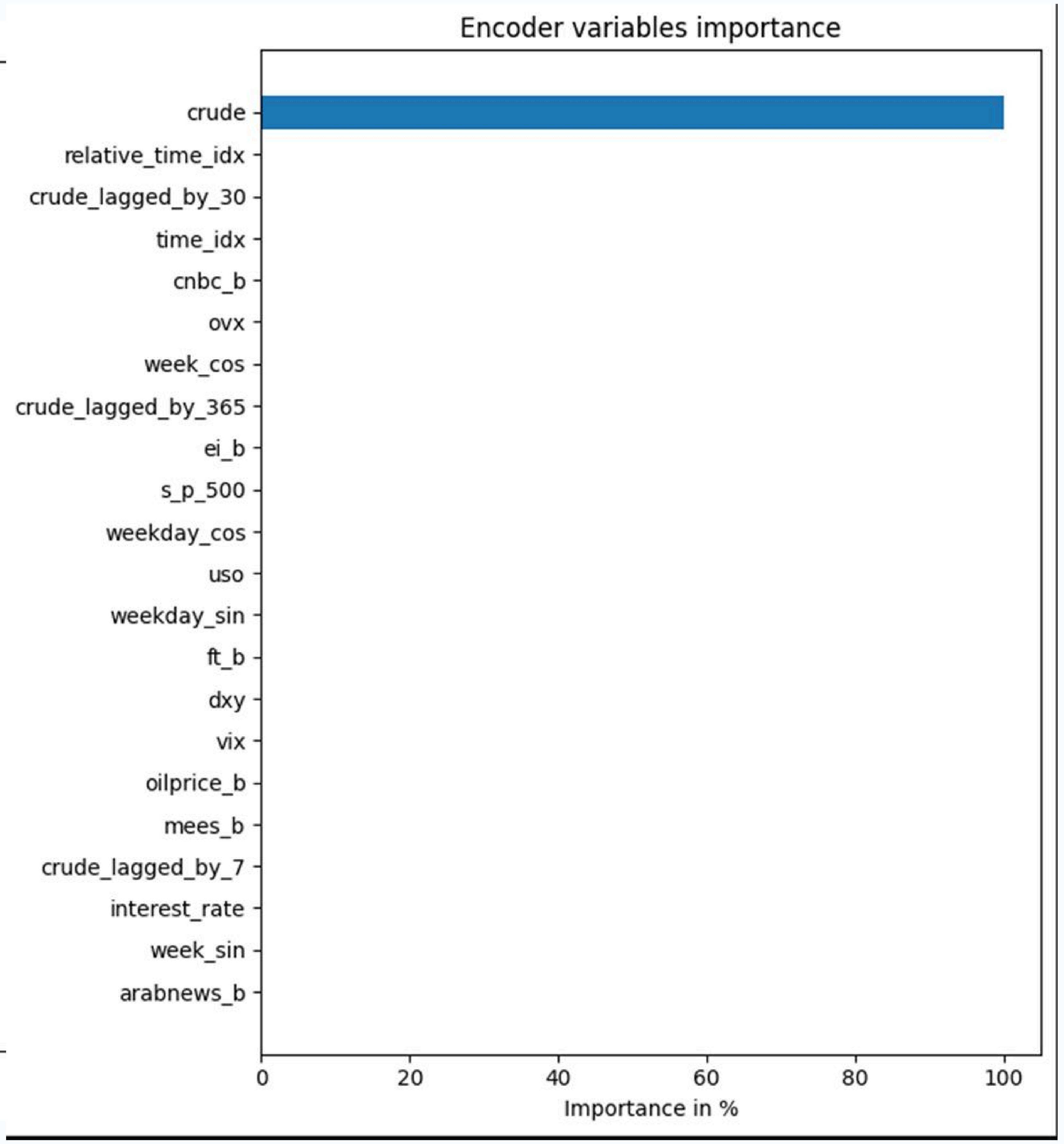
- Hidden Size: 12
- Attention Heads: 3
- LSTM Layers: 4
- Dropout: 0.1
- Batch Size: 96
- Epochs: 15
- Learning Rate: 0.002
- Gradient Clip Value: 1.0
- Loss Function: MSE
- Max Encoder Length: 16

Results & Experimental Evaluation: TFT Feature Importances

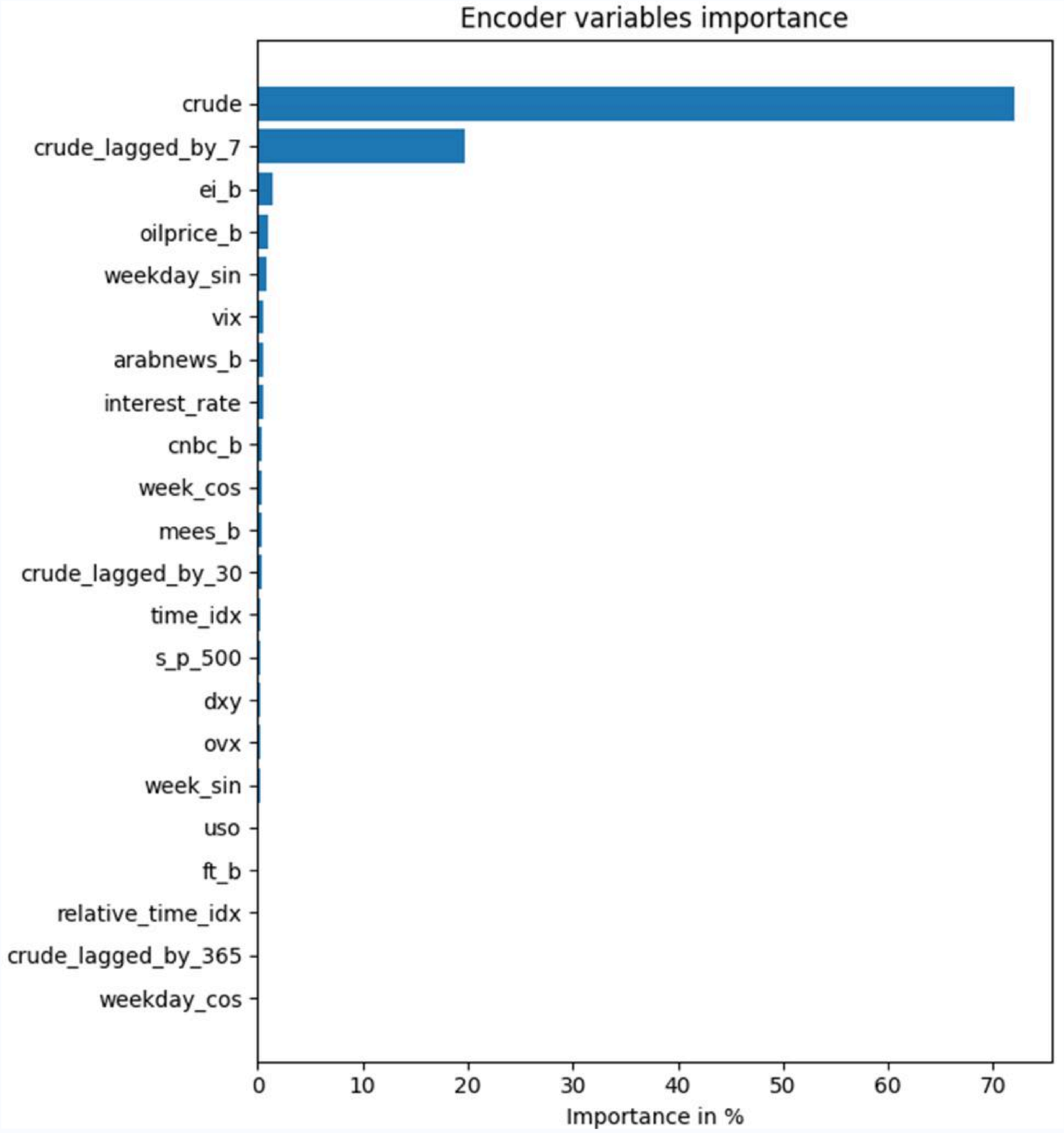
Short-Term model (16 future days)



Medium-Term model (48 future days)



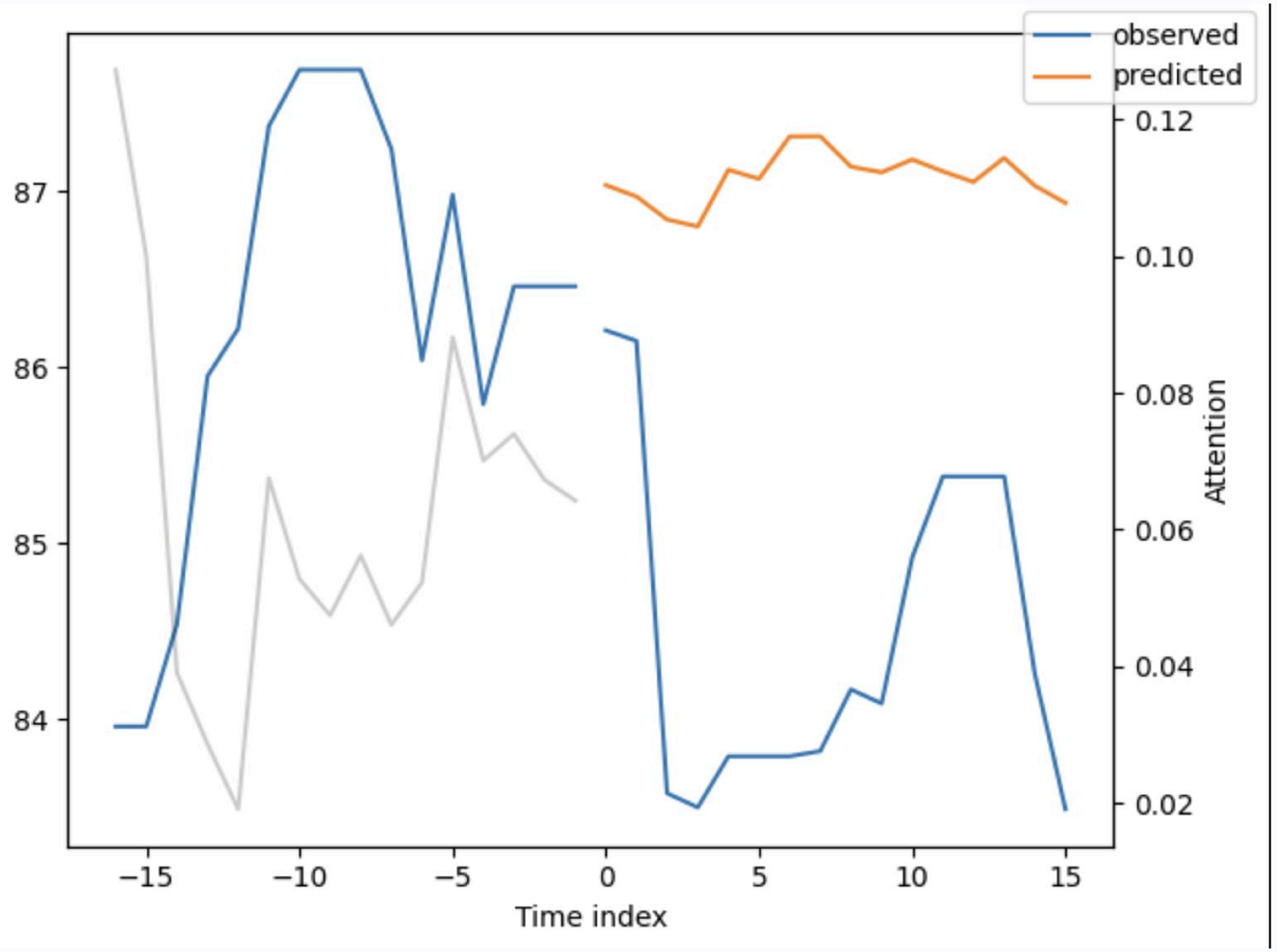
Long-Term model (96 future days)



Results & Experimental Evaluation

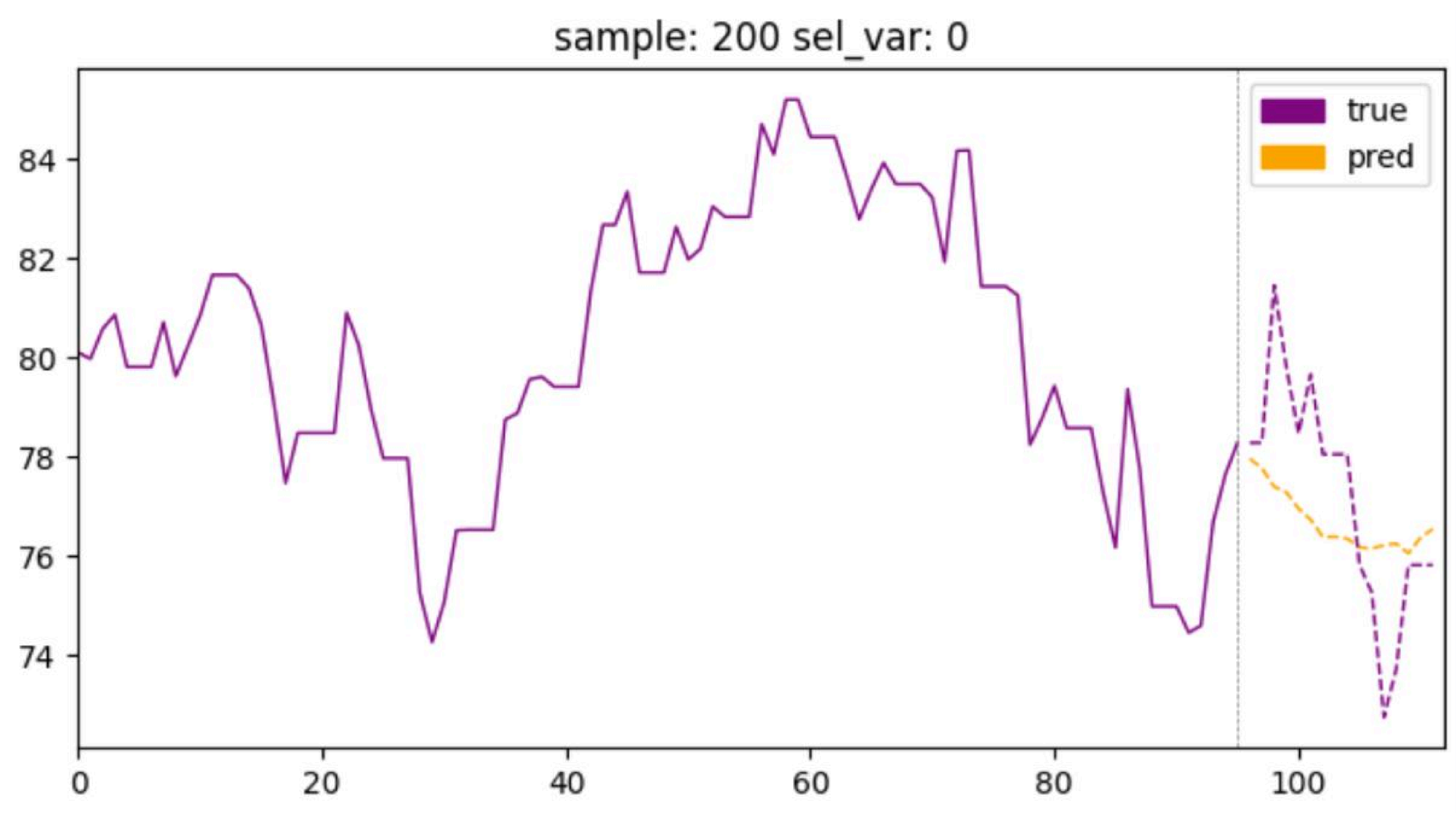
Model Performance Comparison

Horizon	Model	SMAPE (%)	MAE (USD)
16 Steps	ARIMA	8.99	6.90
	SARIMAX	6.82	5.30
	PatchTST	**3.15**	**2.47**
	TFT	3.56	2.80
48 Steps	ARIMA	9.72	7.48
	SARIMAX	6.722	5.26
	PatchTST	**4.86**	**3.84**
	TFT	7.36	6.96
96 Steps	ARIMA	10.45	8.07
	SARIMAX	**6.4**	**5.23**
	PatchTST	6.65	5.24
	TFT	10.37	8.55

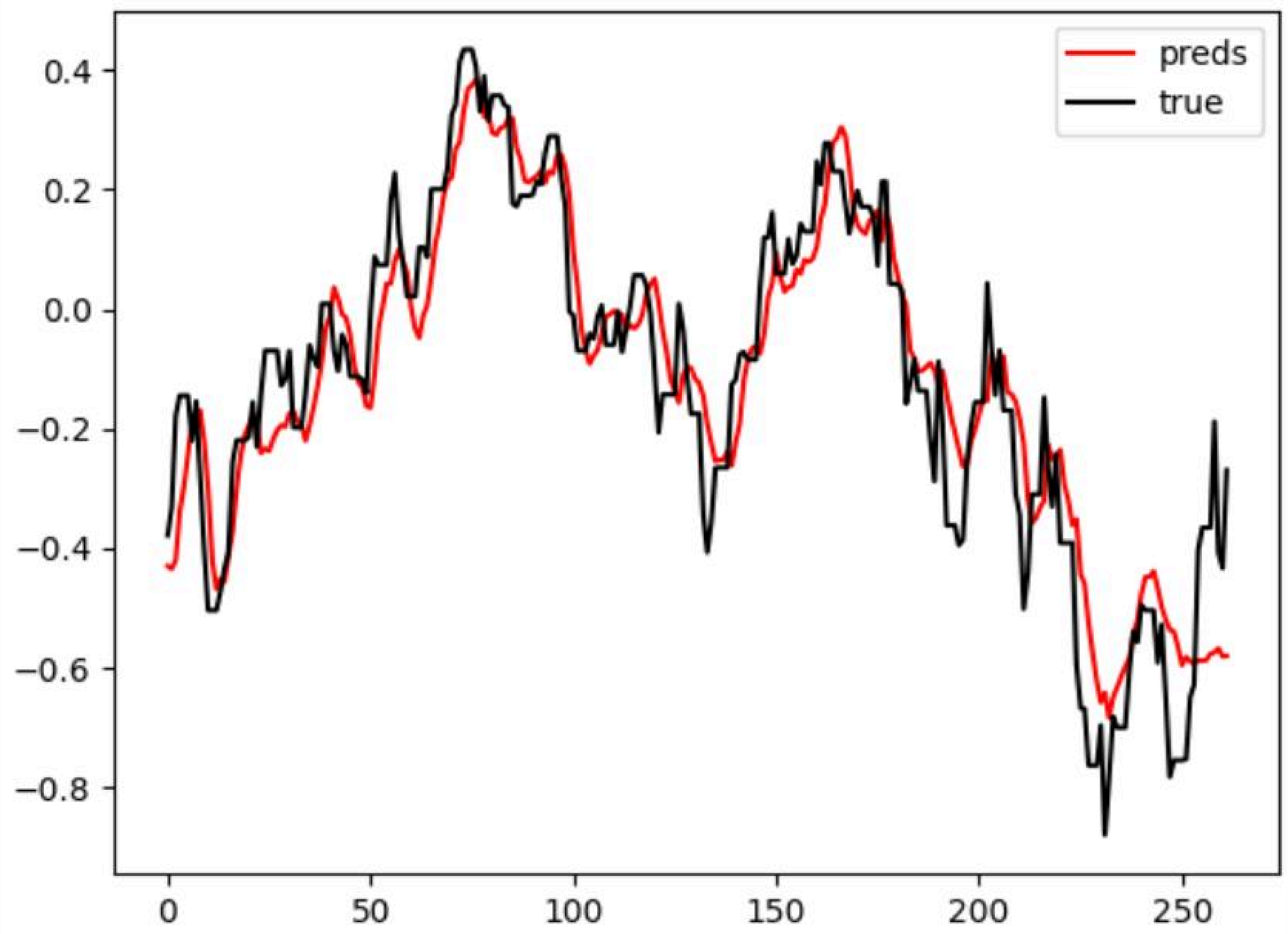


TFT Prediction Across a Sample Test Set Window

Results & Experimental Evaluation



PatchTST Prediction Across the Last Test Set Window



PatchTST First Prediction on Every WIndow

Conclusion

Oil is best predicted autoregressively with lower model complexity

- The multivariate transformer could not outperform autoregressive models
- PatchTST outperformed SARIMAX with a reduced feature set

Improvements:

1. Sentiment modeling
2. Feature set
3. Increase dataset size
4. Analyze factors that contribute to model performance

Thank You

Richie, Rohan, & Uday