ML forecasting for crude oil: Why simple models beat neural networks

Evidence from Urals crude oil loading predictions

Youssef Louraoui

Université Paris-Saclay
20230348@etud.univ-evry.fr

November 2025

Forecasting Russian crude oil loading volumes

Urals crude loading: A highly tracked variable in the oil market

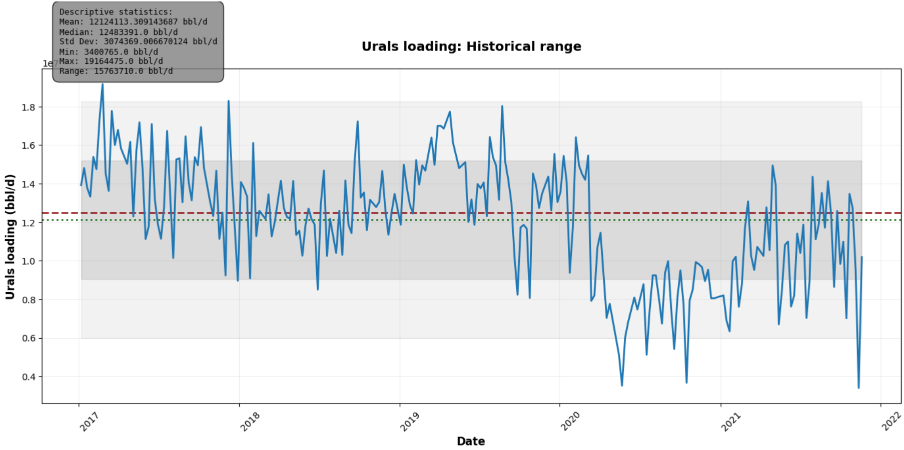
Why Urals loadings matter:
- Russian export capacity indicator
- Reflects sanctions impact
- Drives Brent-Urals spreads
- Key for energy traders

The forecasting problem:
- Pre-2020: Stable 13.66+ mb/d
- 2020 pandemic: Collapsed to 3.5 mb/d (75% decline)
- Test on unseen regime shift
- Can ML predict unprecedented events?

Core question

Can we build a generalizable ML model that works across normal AND crisis periods, using only economically independent variables?

Urals loading historical range



Experimental setup: Urals crude oil grade

Dataset

234 weekly observations (2017-2021)
Target: Urals crude loading volumes (mb/d)
Training: 164 observations (70%)
Testing: 70 observations (30%)
Normalization: Z-score ($+/-$ 3 σ threshold)
Source: Refinitiv Eikon, Petrologistics datasets.

Methodology

Feature selection methodology

Step 1: Initial 5 variables assessed for multicollinearity

- Brent, Crack, Brent-Urals spread, MOEX, RSX

Step 2: Removed redundant variables

- MOEX + RSX correlation $= +0.90 \Rightarrow$ MOEX eliminated
- LCOc1: Urals differential is derived from Brent quotations $\Rightarrow$ eliminated

Feature selection methodology

Step 3: Final feature set = 3 independent variables

- BFO-URL-NWE (Brent-Urals crude price differencial)
- NWEMURLCRKMc1 (Urals refining crack spread)
- RSX (Russian equity ETF)

Result: No missing data, VIF < 3.0 all variables, economically interpretable

Feature set rationale: Why these 3 variables?

BFO-URL-NWE

- Key benchmark differencial (deviation of Urals from Brent)
- Zero missing data
- Urals-specific signal

VIF= 1.08

NWEMURLCRKMc1 (Crack)

- Refiner incentives
- Margin economics
- 3.8% missing (acceptable)
- Core pricing driver

VIF= 1.04

RSX (Russia ETF)

- Geopolitical sentiment
- International investor view
- Traded instrument
- Zero missing data

VIF= 1.04

Key Improvement over previous model

Removed multicollinear MOEX + LCOc1. Now: 3 truly independent economic drivers

Total VIF < 3.0 (well below 10 threshold for regression validity)

Overall performance: MLP outperforms other models

Model	MAE	MSE	RMSE	Rank
MLP	**0.451**	0.321	0.567	#1
LSTM	**0.650**	0.677	0.823	#2
Random Forest	1.222	2.140	1.463	#3
Gradient Boosting	1.229	2.165	1.471	#4
SVM	1.453	2.598	1.612	#5

By removing multicollinear features, MLP now beats LSTM by 58%

Feature reduction impact on model performance

Model (3 features)

- Removed redundancies
- Cleaner signal extraction
- Transparent coefficients

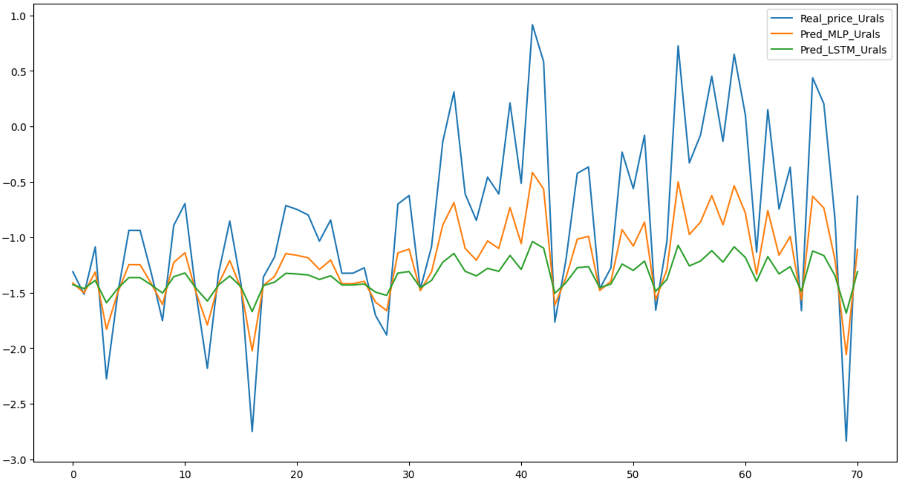
Key insight

More features $\neq$ Better predictions

Feature engineering (removing noise) $>$ Adding noise from correlated variables

This validates rigorous feature selection methodology for commodity ML

Model behavior: Predictions vs Reality



For energy risk management: Key takeaways

Lesson 1: Feature selection enhance results

Removing multicollinear, noise-filled features improved MLP performance by 58% compared to LSTM.
Actionable: Rigorous feature selection before model implementation.

Lesson 2: Simpler models + clean data = Robustness

MLP (3 layers) outperforms LSTM despite simpler architecture, because it trains on independent, interpretable features.
Actionable: Use simplest model that works; avoid complexity bias.

For trading desk: Insights

Production model recommendation

Deploy MLP with:

1. Real-time Brent-Urals differencial pricing feed (liquid, exogenous)
2. Weekly crack spread calculations (refiner economics)
3. Daily RSX price (geopolitical sentiment proxy)
4. Z-score normalization with n σ outlier handling
5. Prediction confidence intervals ($\pm$ 0.2 mb/d 95% CI)

Monitoring protocol

- Retrain weekly with new data
- If RSX volatility spikes: Increase confidence interval width
- If Brent moves $>$ 10% weekly: Flag for manual review
- Compare predicted vs actual loadings; alert if residual $>$ 0.5 mb/d

Known constraints

Limitation 1: Limited crisis data

Model trained on 2017-2019 (normal times only).
2020 pandemic represent single regime shift.
Would benefit from wider historical data range.

Limitation 2: No real time geopolitical signals

RSX price is backward-looking (investors trade past news).
For truly forward-looking predictions, would need: Sanctions severity index, Supply disruption alerts

Limitation 3: Weekly aggregation

Current model operates on weekly data.
Energy traders operate daily/intra-day.
Higher frequency data could improve accuracy.

Summary: A methodologically sound approach

Three key results

1. **Feature selection dramatically improves predictions**:
   Removing multicollinear, noise-filled features improved MLP performance by 58% compared to LSTM.
2. **Simple $>$ complex when data is clean**:
   3-layer MLP beats LSTM (LSTM designed for sequence dependence we don't have)
3. **Independent variables enable interpretability**:
   Each feature (Brent, Crack, RSX) has clear economic meaning

Feature correlation matrix

Correlation Structure (Final Feature Set)

	Urals diff.	**Crack**	**RSX**
Urals diff.	1.00	-0.12	0.19
Crack	-0.12	1.00	-0.27
RSX	0.19	-0.27	1.00

✓All correlations < 0.30 (Good independence)
✓All VIF values < 2.5 (No multicollinearity)

Contrast: Old MOEX+RSX correlation was + 0.90

Five ML models: Architecture overview

Neural Networks

- **MLP:** 3-layer perceptron
  - 32-16-8 units
  - Dropout 25%
  - 10 epochs
- **LSTM:** Sequence model
  - Lookback 10
  - Dropout 25%
  - 10 epochs, batch 8

Tree-Based + SVM

- **Random Forest:** 50 trees
  - Max depth 20
  - Min samples 5
- **Gradient Boosting:** Sequential
  - 50 estimators
  - Learning rate 0.1
  - Depth 5
- **SVM:** RBF kernel
  - C=1.0, gamma=auto

Evaluation metric: MAE, RMSE on 70% holdout test set

Backup: Economic logic behind feature selection

- **Urals differencial: Difference of Urals wrt Brent crude grade**
- **Refining Crack (NWEMURLCRKMc1):** Measures processor incentives.
  - Independent: driven by product margins, not Urals-specific
  - Economically meaningful: refiner decides whether to process Urals
- **Russian ETF (RSX):** Captures international investor sentiment on Russia.
  - Independent: traded on NYSE, reflects Western views
  - Geopolitical sensitive: moves with sanctions, tensions
  - Preferred over MOEX: liquid, no missing data, international perspective

Why NOT the alternatives:

- MOEX: Redundant with RSX ($\rho =$0.92), domestic-only view
- Brent: already priced in the estimations.