

TTF Front Month Gas Price Forecasting (2020–2024)

This project builds a machine learning model to predict monthly TTF gas prices using LNG trade flows, temperatures in key European markets, alternative benchmark prices (Henry Hub, JKM, Brent), and EU country-level gas imports. It aims to demonstrate the application of real-world LNG pricing dynamics for analytics and forecasting.

Objective

- Forecast TTF natural gas prices using historical market data and energy fundamentals.
 - Understand how supply, demand, weather, and global price benchmarks affect TTF pricing.
 - Develop a machine learning model (RandomForest) for forecasting monthly prices.
-

Quick Summary

- Data Period: January 2020 – December 2024 (monthly)
 - Model: RandomForest Regressor
 - Target Variable: TTF Price (USD/MMBtu)
 - Key Drivers: US LNG exports, HH/JKM prices, weather (temp), EU gas imports
-

Key Insights

- TTF is consistently higher than Henry Hub due to Europe's import dependence and supply risks.

- TTF and JKM diverged from HH post-2021, reflecting LNG arbitrage potential.
 - Gas imports and temperature trends correlate with price spikes, particularly in winter months.
 - Russia-Ukraine war in Feb 2022 led to sharp price volatility and structural break in trends.
-

Feature Engineering Strategy

- To avoid overfitting due to the small dataset size (~38 observations), we deliberately limited the number of engineered features while ensuring they were strategically meaningful.
- We included only one country's import volume (Germany), as Germany is a major gas consumer and a key destination for LNG and pipeline imports.
- For temperature-based features, we selected Spain due to initial correlation strength. However, we acknowledge that countries like France and the Netherlands are more influential LNG importers (based on 2024 data from IEEFA).
- Adding features for all countries (imports or temperatures) would have introduced multicollinearity and reduced model generalizability. Instead, we selected a small subset that captured regional demand/supply dynamics effectively. This reflects a pragmatic trade-off between model robustness and physical market relevance.

Model Highlights

- Applied lag features, rolling averages, and spread variables.
- Best performance with selected engineered features: $R^2 \sim 0.87$, MAE ~ 1.3 , RMSE ~ 1.7

- Useful for exploratory pricing applications or trading signal prototypes.
-

Strategic Applications

- Can be extended into gas portfolio optimization or LNG arbitrage modeling.
 - Helpful as an interview showcase for roles in LNG analytics, pricing, or trading.
 - Highlights understanding of EU energy pricing, supply/demand fundamentals, and quant modeling.
-

Assumptions & Notes

- Project focuses on monthly prices, not day-ahead or intraday spot forecasting.
 - Temperature data used as a proxy for demand in key EU countries.
 - Dataset constructed from multiple public sources; merged with consistent monthly frequency.
 - All price data assumed in USD/MMBtu.
-

Final Verdict

- Despite strong test-set performance ($R^2 \sim 0.87$), this project faced key limitations common in LNG macro forecasting:
- Very small dataset: Only ~38 monthly data points, which limits model generalization and increases risk of overfitting.

- Feature dimensionality: Even after selecting top features, adding lags and rolling metrics can inflate model complexity.
- Time series modeling constraints: TimeSeriesSplit was used to respect chronology, but data scarcity reduced cross-validation reliability.

Takeaway

This project demonstrates:

- A pragmatic trade-off between model complexity and physical LNG market structure.
- The importance of market-relevant feature engineering (e.g., gas flows, regional temperatures, price spreads).
- That test-set accuracy alone does not guarantee model robustness, especially in small, temporal datasets.

Notebook

View full notebook on nbviewer:

[Open Notebook](#)

Author

Mohit Kumar

LNG Analyst | Energy Market Research | Quantitative Modeling

Email: mohitkr.h@gmail.com