QF621 Sample Report - Group 7

Table of Contents

1. Executive Summary

This project explores the development of a machine learning-enhanced trading strategy for the Title Transfer Facility (TTF), Europe's primary natural gas trading hub. The European energy market is volatile and complex due to geopolitical shifts, weather variability, and supply dynamics. In this project, we aim to design a data-driven, interpretable trading framework that can generate alpha while managing risk.

By leveraging real-world data from the Bloomberg Terminal, we integrated fundamental factors (e.g., gas storage levels, LNG imports), macroeconomic indicators (e.g. EUR/USD, Brent crude, carbon prices), and technical signals into a predictive model. We . To forecast front-month TTF prices, we trained two machine learning models, LSTM and XGBoost. XGBoost ultimately outperformed LSTM across key metrics and was used to power our final strategy.

Our trading strategy focused on mean-reversion spread trades between TTF and related commodities (Brent and Carbon). Signals were generated based on deviations between our model's predicted prices and the implied TTF prices which were derived from rolling historical ratios. Backtests showed promising results: the TTF-Brent spread strategy returned 22%, while the TTF-Carbon strategy started strong but fell behind in the later part of the test period.

Overall, this project shows how machine learning and domain expertise can work hand in hand to build smarter, more adaptive trading strategies. There's real potential for this approach to be used in practice, and plenty of room for further development.

Most Important Findings:

The most important findings from this project highlight that XGBoost significantly outperformed LSTM in forecasting TTF front-month prices, achieving lower prediction errors (MAE: 2.04 vs. 4.98) and more stable trend tracking. When integrated into a mean-reversion spread trading strategy, XGBoost-powered signals enabled a 22.31% return in the TTF-Brent strategy and -5.72% in the initial TTF-Carbon strategy. However, with risk controls such as dynamic position sizing and a 5% trailing stop loss, the improved strategies delivered stronger performance, raising the Brent strategy return to 28.66% and converting the Carbon strategy to a 9.52% gain. A combined, diversified portfolio achieved an 18.90% return with a Sharpe ratio of 1.49 and reduced drawdowns, demonstrating the effectiveness of ML-enhanced trading in volatile energy markets.

2. Introduction (Business Problem)

The volatility of natural gas prices in Europe presents both a challenge and an opportunity for energy market participants. In recent years, geopolitical tensions, decarbonization policies, and unpredictable weather patterns have led to extreme fluctuations in gas prices, with the TTF benchmark hitting a record high of €345/MWh in March 2022. For traders and risk managers, this volatility highlights the need for advanced forecasting tools and systematic trading strategies.

The core objective of this project is to develop and validate a machine learning-based trading strategy for the TTF natural gas market. By combining technical, fundamental, and sentiment data into a unified predictive framework, we seek to generate actionable trading signals that optimize profitability while managing risk. Unlike traditional technical trading rules or purely statistical approaches, our methodology integrates domain knowledge of gas market mechanics with machine learning's ability to uncover nonlinear patterns and relationships.

In doing so, we address key questions such as:

- Can ML models reliably forecast short-term TTF price movements?
- Which features are most predictive of price direction?
- And can these forecasts be turned into profitable, risk-adjusted trading strategies?

3. Market & Data Overview

3.1 Market Context

The European energy crisis, triggered by geopolitical events and structural market changes, has highlighted the importance of having robust predictive models for energy commodities. The TTF market's complexity, with influences ranging from weather patterns and storage levels to geopolitical developments and regulatory changes, makes it an ideal candidate for machine learning applications, which excel at capturing multidimensional nonlinear relationships.

3.2 What is TTF and Why It Matters

The Title Transfer Facility (TTF) is a virtual trading hub for natural gas in the Netherlands and serves as the main benchmark for European natural gas prices. Established in 2003 by Gasunie, TTF has grown to become the most liquid gas market in Europe, with prices at other hubs (e.g., THE in Germany) typically trading at a spread to TTF.

As Europe's benchmark, TTF plays a crucial role in price discovery and market arbitrage. For example, gas flows can be redirected across borders when TTF prices diverge significantly from those in neighboring markets. With natural gas being critical to electricity generation, industrial use, and household heating, especially during winter, the TTF market reflects both short-term shocks and long-term structural shifts in the energy ecosystem.

3.3 Why Focus on Front-Month Prices

Our focus is on front-month (FM) contracts, as they are the most actively traded and relevant for short-term trading and hedging. Market participants base many of their decisions, such as storage injections, withdrawals, and risk hedges, on FM prices rather than on highly volatile within-day (WD) prices (Van Der Maat, 2015).

These FM prices are also strongly influenced by broader supply-demand imbalances and market expectations, making them ideal for machine learning-based forecasting.

For example, when WD prices trade at a discount to FM prices (e.g., €38/MWh vs. €40/MWh), traders typically inject gas into storage, creating a long exposure to the FM contract that requires hedging. Conversely, when WD prices trade at a premium to FM, withdrawal from storage becomes economical, creating a natural short position in the FM contract.

3.4 Data Sources and Features

We sourced our data primarily from the Bloomberg Terminal at Singapore Management University. Drawing from both domain knowledge and research into natural gas market fundamentals (Chen et al., 2023), we designed a dataset tailored for machine learning-based forecasting of TTF front-month prices. The dataset included:

1. **TTF Front Month Price:** Our target variable for prediction.
2. **Fundamentals:**
    a. Demand indicators (daily temperature, industrial demand, household heating needs)
    b. Supply metrics (Netherlands storage percentage, LNG imports, interconnector flows)
3. **Macroeconomic & Financial Variables:**
    a. EUR/USD exchange rate, Brent crude price, coal price, EUA carbon price
    b. NBP-TTF spread as a measure of market arbitrage
4. **Technical Features:**
    a. Moving averages, rolling price ratios
5. **Model Features Example:**
    a. Summer-Winter Spread, UK-Netherlands Spread, etc.

4. Data Processing & Model Training

To develop a robust predictive framework for TTF price forecasting, we implemented a carefully structured data processing pipeline followed by two types of model training: LSTM and XGBoost. Each step of the pipeline is not only a technical necessity but is also informed by the nature of the energy market, the characteristics of our data, and the modeling objectives.

4.1 Data Processing

Feature Selection and Cleaning
To construct a forecasting-ready dataset, we extracted a curated subset of explanatory variables reflecting key aspects of gas price dynamics. These variables encompass physical fundamentals, macroeconomic indicators, and spread-based signals commonly used by energy traders (*Financial Energy Spreads | EBF 301: Global Finance for the Earth, Energy, and Materials Industries*, n.d.). Table below provides an overview of the features selected for modeling.

Category	Features
Supply & Demand	NL_Storage_%Full, NL_LNG_Import_tonne, UK_Interconnector_MCM_d
Macroeconomic	EURUSD, Brent_USD_per_barrel, Coal_USD_per_tonne, EUA_Euros_per_tonne
Spread Signals	Summer_Winter_Spread, UK_NL_Spread
Climate Proxy	Daily_Temperature

Normalization

All numerical features were normalized to a [0, 1] scale using MinMaxScaler. This method preserves the shape of the original distribution while ensuring that all input variables lie within the same numerical range.

Normalization is particularly critical for deep learning models such as LSTM, which are sensitive to the scale of input data. Without normalization, input features of varying magnitudes may dominate gradient updates or lead to instability during training.

Temporal Sequence Construction

To prepare the data for LSTM, we restructured it into time series sequences using a **rolling window technique** with a 30-day lag. Instead of feeding single-day feature snapshots into the model, we created overlapping input sequences of the past 30 days of features, and used them to predict the TTF price on the next day , the 31st day. This method allows the model to learn temporal dependencies, trend shifts, and lagged effects , all of which are crucial in volatile markets like TTF.

Example:

- Day 31 → Input: Days 1–30 → Target: Day 31 TTF price
- Day 32 → Input: Days 2–31 → Target: Day 32 TTF price
  ... and so on, until the end of the series.

After transformation:

- X_all.shape = (1271, 30, 13) → 1,271 samples, each a 30-day sequence with 13 features
- y_all.shape = (1271,) → One TTF price target per sequence

4.2 Model Selection & Architecture

To evaluate performance across different learning paradigms, we implemented both a recurrent neural network (LSTM) and a tree-based ensemble model (XGBoost). The former is capable of modeling sequential relationships, while the latter excels at handling structured, tabular data with engineered lags.

4.2.1 Long Short-Term Memory (LSTM)

LSTM is a variant of recurrent neural networks (RNNs) equipped with memory cells and gating mechanisms that enable the model to learn long-term dependencies across sequential data. Such architecture is particularly advantageous for financial time series, where price movements are shaped by delayed reactions to supply shocks, macroeconomic indicators, and seasonal demand patterns.

Model Architecture

Our LSTM model includes:

- **Two stacked LSTM layers**: 128 units and 64 units respectively
- **Dropout layers** (0.2) after each LSTM layer to prevent overfitting
- A **dense output layer** for final regression prediction

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 128)	72,704
dropout (Dropout)	(None, 30, 128)	0
lstm_1 (LSTM)	(None, 64)	49,408
dropout_1 (Dropout)	(None, 64)	0
dense (Dense)	(None, 1)	65

The model was trained using the Adam optimizer with a learning rate of 0.001 and Mean Squared Error (MSE) as the loss function. Training was conducted for 50 epochs with a batch size of 4, and a 10% validation split was applied.

These hyperparameters were selected after empirical tuning using trial runs across multiple configurations. Lower learning rates and dropout regularization were especially effective in improving convergence stability and minimizing overfitting, given the relatively limited dataset.

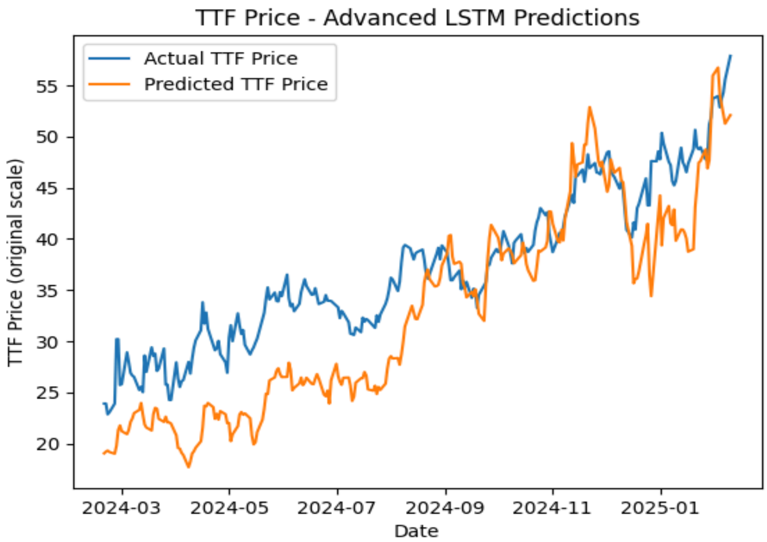
Evaluation Metrics

After inverse-scaling the predictions to the original price scale, the LSTM model yielded the following results on the test set:

- MSE: 34.68
- RMSE: 5.89
- MAE: 4.98

These metrics suggest that, on average, the model's predicted TTF prices deviate from the true observed values by approximately €4.98 per MWh, which is a non-trivial error in volatile energy markets but within an acceptable range for initial modeling efforts.

Forecast Visualization and Analysis



As shown in the above figure, the LSTM model captures the general directional trend of TTF prices across the test period. The predicted values (orange line) follow the actual prices (blue line) reasonably well during mid-2024. However, in the earlier part of the year, the model underpredicts the price level consistently, and in later months, volatility is underestimated, with the model failing to capture several sharp drops and spikes.

This behavior reflects a known limitation of sequence models trained on limited data: while they can learn smoothed long-term dependencies, they often struggle to react to short-term structural breaks or external shocks, such as geopolitical disruptions or policy changes.

Feature Importance via Permutation Analysis

To further interpret the model, we conducted a **permutation importance analysis** on the test set. This method involves randomly shuffling each input feature one at a time and measuring the impact on model MSE. A greater increase in error indicates higher feature importance.

```
Permutation importance (feature -> MSE difference):
NL_Storage_%Full                        39.2281
EUA_Euros_per_tonne                      6.6691
EURUSD                                   6.4904
Summer_Winter_Spread                     5.1579
UK_Interconnector_MCM_d                  3.5796
Brent_USD_per_barrel                     2.9804
NL_LNG_Import_tonne                      1.7691
UK_NL_Spread                             1.5201
Coal_USD_per_tonne                      -0.3239
NL_Industrial_Demand_GWh_d              -0.6163
Daily_Temperature                       -0.7567
```

Key Findings:

- **NL_Storage_%Full** emerged as the most critical feature, with a **+39.2 increase in MSE** when permuted.
- **EUA_Euros_per_tonne**, **EURUSD**, and **Summer_Winter_Spread** also contributed significantly to the model's forecasting accuracy.

- Even features with negative MSE changes (e.g., Daily_Temperature) showed values close to zero, suggesting they were either redundant or marginally helpful.

These results are consistent with domain knowledge, as **storage levels**, **carbon market signals**, and **macro FX trends** are all well-known drivers of natural gas prices.

The LSTM model demonstrated a reasonable capacity to model trend and seasonality within the TTF time series. However, its limitations in capturing volatility and reacting to exogenous market shifts suggest potential value in exploring non-sequential models. Therefore, we proceeded to train a second model, XGBoost, to evaluate whether a flattened, tree-based approach could better generalize under these constraints.

4.2.2 XGBoost Regression

XGBoost (Extreme Gradient Boosting) is a decision-tree-based ensemble model that has proven highly effective for structured, tabular datasets. Unlike LSTM, which models temporal dependencies implicitly, XGBoost requires explicit feature engineering for time series tasks. In our implementation, we flattened each 30-day rolling sequence of features into a single 390-dimensional vector (30 days × 13 features), enabling the model to learn lagged interactions directly.

Model Architecture and Hyperparameter Tuning

The XGBoost model was configured with the following hyperparameters:

- n_estimators = 200
- max_depth = 5
- learning_rate = 0.05
- subsample = 0.8
- colsample_bytree = 0.8

These parameters were selected through iterative experimentation and validation. A moderate depth and learning rate were used to prevent overfitting while enabling meaningful feature interactions. The subsample and colsample_bytree parameters introduced randomness to enhance model generalization.

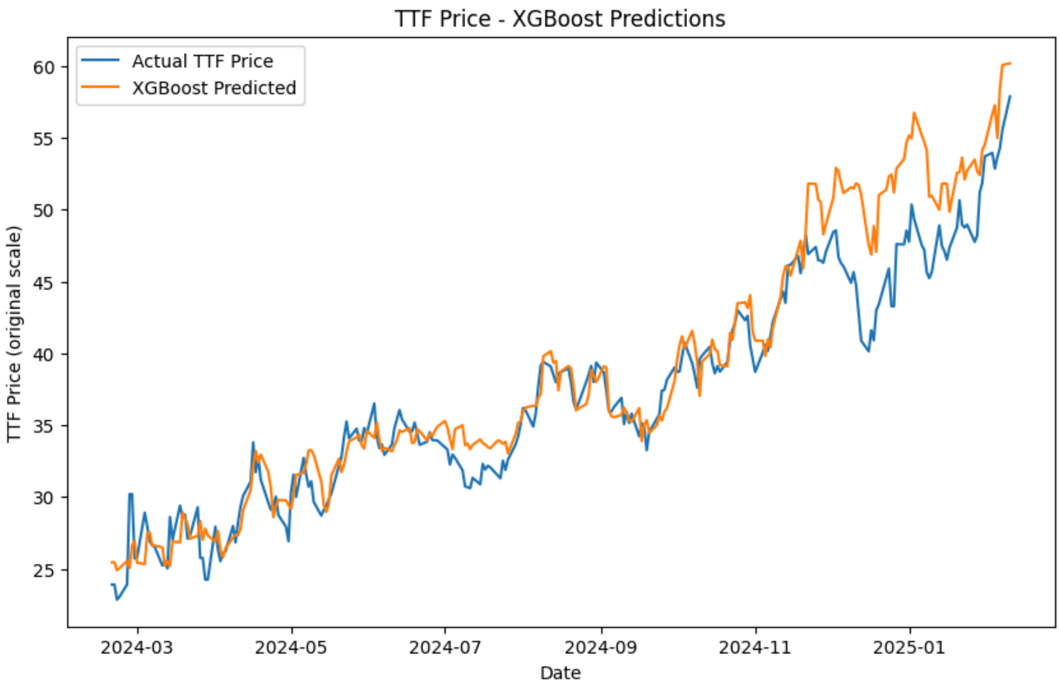
Evaluation Metrics

After rescaling predictions to the original price scale, the XGBoost model demonstrated superior performance across all standard regression metrics:

- MSE: 8.44
- RMSE: 2.91
- MAE: 2.04

These metrics suggest that, on average, the model's predicted TTF prices deviate from the true observed values by approximately €2.04 per MWh.

Forecast Visualization and Analysis

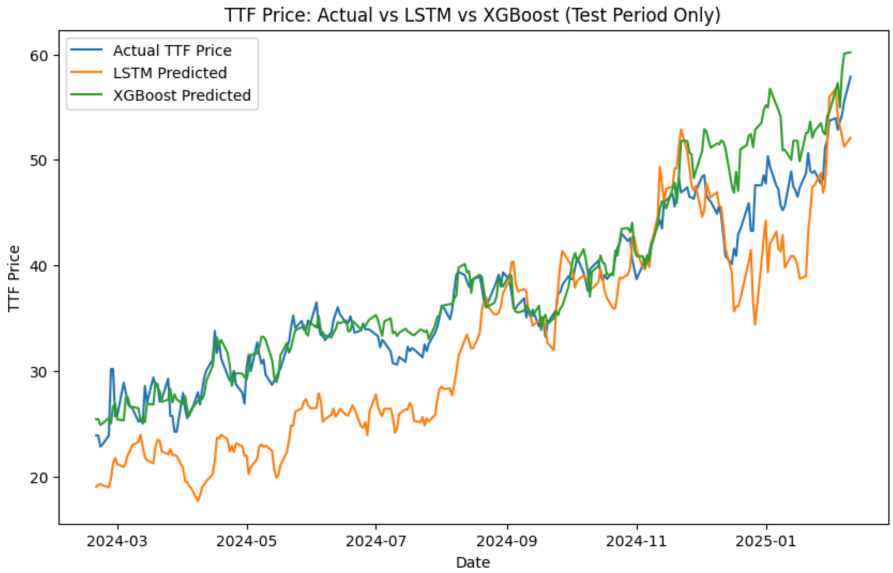


The XGBoost model tracked the actual TTF price trajectory more closely than the LSTM model, especially during the earlier months of the test period. It successfully captured both trend and relative magnitude, while remaining robust to high-frequency noise. However, like LSTM, it exhibited prediction drift near the end of the test period, likely due to out-of-distribution data.

4.3 Comparative Analysis and Final Model Selection

To highlight the performance gap between both models, we summarize their results below:

Metric	LSTM	XGBoost
MSE	34.68	8.44
RMSE	5.89	2.91
MAE	4.98	2.04



XGBoost's predictions aligned more closely with observed TTF prices in early and mid-testing periods. Both models diverged significantly by early 2025, which may reflect out-of-distribution shifts, such as market regime changes or external shocks not present in the training window. Based on this comparative evaluation, we selected the XGBoost model for downstream trading strategy tasks. Its lower error rates, superior trend tracking, and enhanced interpretability made it more reliable for signal generation in a live trading environment. Moreover, its ability to produce consistent predictions without the need for sequence-based modeling simplifies deployment and reduces computational overhead.

5. Strategy Development

5.1 Correlation & Cointegration Analysis

Pair	Correlation	Cointegration Test Statistic	P-value
TTF-Brent	0.6251	-3.7145	0.0176
TTF-Carbon	0.5326	-2.9870	0.1133

The critical values for the cointegration test were [-3.90488438, -3.34083411, -3.04771407] at the 1%, 5%, and 10% significance levels respectively.

The TTF-Brent pair showed both strong correlation (0.6251) and statistically significant cointegration ($p < 0.05$), with the test statistic (-3.7145) exceeding the 5% critical value. This indicates a robust long-term equilibrium relationship between TTF and Brent prices, making it an ideal candidate for spread trading.

The TTF-Carbon pair exhibited moderate correlation (0.5326) but failed to meet the traditional threshold for cointegration ($p = 0.1133$). The test statistic (-2.9870) fell short of even the 10% critical value, suggesting a weaker long-term relationship. Despite this, we included both pairs in our strategy development for comparison purposes, as even non-cointegrated pairs can sometimes exhibit profitable trading opportunities over shorter horizons.

5.2 Spread Trading Framework

5.2.1 Historical Relationship Quantification

For each pair (TTF-Brent and TTF-Carbon), we calculated rolling 30-day price ratios. The ratio on any given day represents the TTF price divided by the respective commodity price, while the rolling ratio is the average of these daily ratios over the past 30 days.

The rationale behind this approach is that while absolute prices of energy commodities may fluctuate significantly, their relative value relationships tend to exhibit mean-reversion properties. These ratios effectively capture the structural price relationships without requiring explicit cointegration model parameters.

5.2.2 Implied Price Mechanism

The core insight of our strategy lies in determining an "implied fair value" for TTF based on the current price of the paired commodity and the historical relationship. This is calculated by multiplying the current commodity price by the rolling ratio.

This calculation provides a model-free baseline estimate of where TTF should trade relative to the paired commodity if their historical relationship holds. Deviations from this implied price potentially represent exploitable trading opportunities.

5.2.3 Signal Generation with ML Enhancement

Our innovation comes from comparing the XGBoost model's predicted TTF price (which incorporates numerous fundamental factors) against the implied price:

1. **Long TTF / Short Commodity Signal** generated when the XGBoost predicted TTF price exceeds the implied TTF price by more than a threshold (TTF undervalued relative to the commodity)
2. **Short TTF / Long Commodity Signal** generated when the XGBoost predicted TTF price is below the implied TTF price by more than a threshold (TTF overvalued relative to the commodity)
3. **Neutral Signal** (close positions) when the absolute difference between XGBoost predicted TTF and implied TTF is less than the threshold

The threshold parameter (set to 1% of the implied price) filters out noise and reduces excessive trading during minor price fluctuations.

5.2.4 Dynamic Position Sizing

To optimize risk-adjusted returns, we implemented a sophisticated position sizing model that scales exposure based on:

1. **Signal Strength:** Larger deviations between predicted and implied prices warrant larger positions
2. **Recent Spread Volatility:** Lower position sizes when correlation between assets begins to weaken
3. **Price Ratio:** Maintains dollar-neutral exposure between the paired assets

The position size for TTF is determined by multiplying a base position size (Adjustable) by the signal strength (the proportional difference between predicted and implied TTF prices) and dividing by the recent spread volatility.

As for how the spread volatility is calculated, we take the annualised difference between the daily price percentage changes over the last 20 days. This results in a parameter that measures how much one asset is trending away from the other. In an ideal scenario where one asset responds quickly in correlation with the other in price changes this percentage would be close to zero allowing us to take a larger position.

For the commodity position, we take the negative of the TTF position multiplied by the price ratio to maintain dollar-neutral exposure. This approach concentrates capital in the highest-conviction opportunities while reducing exposure during uncertain market conditions.

6. Trailing Stop Loss and Hedging

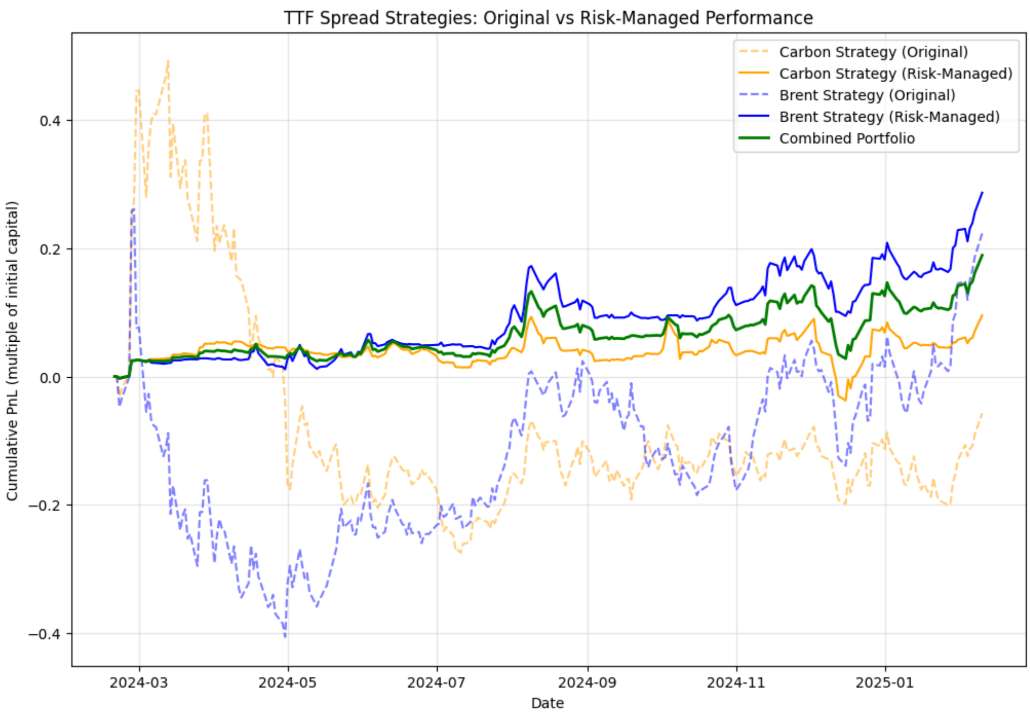
Through testing different stop loss sizes (Baviera & Baldi, 2017), we determined a good trailing stop loss to be 5% matching our risk appetite and striking a balance between risk and reward. Whenever a position is opened, there will be a stop loss of 5% on our total position. This means that whenever we hit a cumulative drawdown of 5% on our positions trading in opposite directions, we will exit the trade.

As a hedging measure we also calculated a strategy combining both our Carbon and Brent Strategy investing half in each which gave us a better return than if we had just bet all on our carbon strategy. This shows that diversification can sometimes give us a better chance at getting a higher return, lowering risk as well when we are not sure of future performances.

7. Evaluation & Results

Original Strategy	Carbon Strategy	Brent Strategy
Final PnL	-5.72%	22.31%
Max Drawdown	-76.66%	-66.73%
Sharpe Ratio	0.17	0.60
Win Rate	49.41%	54.51%

Improved Strategy	Carbon Strategy	Brent Strategy	Combined Portfolio
Total Return	9.52%	28.66%	18.90%
Sharpe Ratio	0.74	1.92	1.49
Max Drawdown	-11.93%	-8.70%	-10.00%
Win Rate	49.02%	53.73%	52.94%
Number of Trades	127	128	128
Avg Trade Duration	6.4 days	7.4 days	6.9 days



The improved strategy clearly performed better than the original. While the original Carbon Strategy ended with a loss of 5.72%, the improved version turned that into a positive return of 9.52%. The Brent Strategy also saw a large jump from 22.31% to 28.66%. What's more important is that risk was reduced. Max drawdown dropped sharply for both strategies, with the Carbon Strategy improving from a painful -76.66% to just -11.93%, showing much better downside control.

Combining the Carbon and Brent strategies gave a balanced portfolio that performed well overall. The combined version returned 18.90%, with a strong Sharpe ratio of 1.49 and a solid win rate of nearly 53%. This shows how diversification helped smooth out performance and reduce risk. It also helped keep trade results more consistent without giving up too much in returns.

8. Conclusion

In closing, this project successfully demonstrates how machine learning, when combined with solid domain expertise, can drive a profitable and risk-adjusted trading strategy in Europe's natural gas market. The XGBoost model produced robust TTF price forecasts that exceeded LSTM performance. Leveraging these forecasts within a spread-trading framework not only enhanced return potential but also offered a systematic way to manage drawdowns and volatility.

Importantly, this project addressed the core questions posed at the outset. First, we showed that machine learning models, especially XGBoost can indeed forecast short-term TTF price movements with a high degree of accuracy, achieving a mean absolute error of just €2.04/MWh. Second, our permutation analysis revealed that storage levels, carbon prices, and FX rates were among the most predictive features of price direction, aligning well with industry intuition. Finally, we translated these forecasts into a live trading framework that not only generated positive alpha but also demonstrated resilience through dynamic position sizing, stop-loss mechanisms, and portfolio diversification.

The final results reinforce the value of diversification across Brent and Carbon pairs, as well as the importance of dynamic position sizing and well-calibrated stop-loss mechanisms. Together, these elements generated higher Sharpe Ratios, reduced drawdowns, and ultimately delivered attractive risk-adjusted returns. Although the strategy's real-world implementation would face challenges such as sudden geopolitical shifts, liquidity constraints, and execution costs, the demonstrated framework can be refined to adapt to evolving market conditions. With further feature engineering and retraining on expanded datasets, the approach holds promise for continued outperformance and deeper insights into energy market dynamics.

9. References

Baviera, R., & Baldi, T. S. (2017, June 21). *Stop-loss and Leverage in optimal Statistical Arbitrage with an application to Energy market*. arXiv.org. https://arxiv.org/abs/1706.07021

Bloomberg L.P. (2024). *TTF front-month prices, storage levels, LNG imports, EUR/USD, Brent crude, and carbon data*. Bloomberg Terminal. Accessed at Singapore Management University.

Chen, Y., Hartley, P. R., & Lan, Y. (2023). Temperature, storage, and natural gas futures prices. *Journal of Futures Markets*, *43*(4), 549–575. https://doi.org/10.1002/fut.22402

Financial Energy Spreads | EBF 301: Global Finance for the Earth, Energy, and Materials Industries. (n.d.). https://www.e-education.psu.edu/ebf301/node/549

Factors affecting natural gas prices - U.S. Energy Information Administration (EIA). (n.d.). https://www.eia.gov/energyexplained/natural-gas/factors-affecting-natural-gas-prices.php

Van Der Maat, J. H. J., s1878581. (2015). *Impact of natural gas storage on natural gas prices* https://repositorio.uchile.cl/bitstream/handle/2250/142637/Van%20Der%20Maat%20Jan-Pieter.pdf

* Data sources were taken from Bloomberg Terminal in Singapore Management University