Haute école de gestion
Genève

Algorithmic Trading in Commodity Markets: Assessing Performance through Machine Learning and Technical Analysis

Bachelor Project submitted for the degree of
Bachelor of Science HES in International Business Management

by

David ALVAREZ

Bachelor Project Mentor:
Luis DEL POZO, PhD

Geneva, 31st of May 2024
Haute école de gestion de Genève (HEG-GE)
International Business Management

Hes·so GENÈVE
Haute Ecole Spécialisée
de Suisse occidentale

Disclaimer

This report is submitted as part of the final examination requirements of the Haute école de gestion de Genève (HEG), for the Bachelor of Science HES-SO in International Business Management. The use of any conclusions or recommendations made in or based upon this report, with no prejudice to their value, engages the responsibility neither of the author, nor the author's mentor, nor the jury members nor the HEG or any of its employees.

Declaration of authorship

By submitting this manuscript, the author declares that it constitutes their own original work. The author acknowledges full responsibility for the work's design; the acquisition, analysis, and interpretation of the data; as well as the formulation of conclusions within the work. Furthermore, the author asserts sole responsibility for the drafting of the manuscript and its thorough review for significant intellectual content. The author is accountable for all aspects of the work, including matters related to its accuracy and integrity.

Any external assistance, including the use of Artificial Intelligence (AI)-assisted technologies, has been systematically identified in accordance with the HEG Referencing guide and are declared in the reporting summary appended at the conclusion of this manuscript. The author further declares that there is no plagiarism in the text and images within the work.

Acknowledgements

I would like to express my gratitude to Luis del Pozo, PhD, for his continuous guidance, mentorship, and unwavering belief in my abilities. His invaluable advice, thoughtful insights, and support were essential.

I am also thankful to Jan Erik Meidell, PhD, my professor in the field of machine learning and statistics, whose knowledge and passion for the subject have inspired me to explore new horizons in machine learning.

Furthermore, I would like to thank Julie Noller and Richard Watts, whose expertise and dedication have been invaluable assets throughout this research journey.

Executive Summary

This research evaluates the effectiveness of machine learning (ML) and technical analysis strategies in commodity trading. Six strategies were backtested across various commodity futures markets: three based on technical indicators (simple moving average crossover, dual momentum crossover, and momentum strategy in volume) and three using machine learning algorithms (Logistic Regression, Decision Tree, and Random Forest). The performance of these strategies was evaluated using win rate, profit factor and Sharpe ratio.

The analysis revealed that ML, particularly the Random Forest model, often outperformed technical analysis strategies and their counterparts in terms of risk-adjusted returns and profitability. This success can be attributed to ML's ability to capture complex, non-linear relationships within market data. However, the effectiveness of trading strategies varied significantly across different commodity markets and market phases. While machine learning strategies demonstrated resilience in certain periods, particularly during phases of steady volatility, they encountered notable challenges amidst periods of heightened volatility or significant market events, with the exception of Random Forest, which tended to maintain its performance.

Firms interested in using machine learning models for commodity trading could benefit from several key recommendations. First, they should prioritize a robust data strategy by ensuring clean, relevant, and diverse market-specific input variables combining a mix of technical indicators, sentiment analysis, fundamental and market-depth data to capture different market patterns. These choices need to be driven by feature selection methods and analysis. Feature engineering is also vital, involving data transformation, creating new indicators, and aggregating information across various timeframes. For model development, leveraging hybrid models can offer an edge, and ensuring adaptability by developing models that can adjust to different markets with regular recalibration. Incorporating integrated risk management principles within the models, such as position sizing based on volatility is also critical.

Contents

Algorithmic Trading in Commodity Markets: Assessing Performance through Machine Learning and Technical Analysis i
Disclaimer ii
Declaration of authorship ii
Acknowledgements iii
Executive Summary iv
Contents v
List of Tables viii
List of Figures viii
1. Introduction 1
1.1 Commodities pricing 1
1.2 Financial markets and trading 2
1.3 Commodity contracts 2
1.4 Algorithmic trading 3
2. Literature review 5
2.1 Technical and Fundamental analysis 5
2.2 Efficient markets 5
2.3 Machine learning in commodity trading 6
3. Methods 9
3.1 Data description 10
3.1.1 Commodity prices data 10
3.1.2 Economic and financial data 12
3.2 Technical Indicators 12
3.2.1 Simple Moving Average 13
3.2.2 Bollinger Bands 13
3.2.3 Relative Strength Index (RSI) 14
3.2.4 Money Flow Index (MFI) 15
3.2.5 Rate of Change (ROC) 16
3.2.6 Momentum Strategy in Volume (MSV) 16
3.3 Performance metrics 17
3.3.1 Win Rate 17
3.3.2 Profit factor 17
3.3.3 Sharpe Ratio 18
3.3.4 Benchmark Model 19
4. Technical Analysis Strategies 20
4.1 Simple Moving Average (MA) Crossover 20
4.1.1 Results 21
4.2 Dual Momentum (MFI – RSI) Crossover 22
4.2.1 Results 23
4.3 Momentum Strategy in Volume (MSV) 24
4.3.1 Results 25

5. Machine Learning Models **27**
5.1 Structure of the algorithms **27**
5.1.1 Implementation workflow *27*
5.1.2 Input variables *30*
5.1.3 Strategy cross-validation *31*
5.2 Model Evaluation and Interpretation **33**
5.2.1 Classification Report *33*
5.2.2 Feature importance *35*
5.3 Logistic Regression **42**
5.3.1 Results *42*
5.4 Decision Tree **43**
5.4.1 Results *44*
5.5 Random Forest **46**
5.5.1 Results *46*
6. Analysis of trading performance **48**
6.1 Overall performance **48**
6.2 Evolution of performance **50**
6.2.1 Crude Oil *50*
6.2.2 Natural gas *52*
6.2.3 Copper *53*
6.2.4 Silver *55*
6.2.5 Corn *56*
6.2.6 Sugar *58*
6.3 Robustness Analysis– Random Forest **60**
6.3.1 Comparative analysis *60*
6.3.2 Sensitivity Analysis *61*
6.3.3 Performance cross-validation *62*
6.3.4 Portfolio Comparison *64*
7. Experts' opinion **65**
7.1 Data and market requirements **65**
7.2 Main factors affecting machine learning profitability **66**
7.3 Advantages of machine learning **66**
7.4 Disadvantages of machine learning **66**
7.5 Market phases **67**
8. Limitations **68**
9. Recommendations **69**
10. Conclusion **71**
References **72**
Use of Artificial Intelligence-assisted tools **76**
Appendix 1 – Interview of respondent A **77**
Appendix 2 – Interview of respondent B **80**
Appendix 3 – Interview of respondent C **85**
Appendix 4 – Summary of parameters **89**
Appendix 5 – Feature Importance (Logistic Regression) **90**

Appendix 6 – Feature Importance (Decision Tree)..90
Appendix 7 – Feature Importance (Random Forest)..92

List of Tables

Table 1 Basic statistics of daily returns [%]..11

Table 2 Trading performance of MA Crossover 22

Table 3 Trading performance of MFI-RSI Crossover 24

Table 4 Trading performance of MSV...26

Table 5 Description of input variables...31

Table 6 Classification Report of Machine Learning Models 34

Table 7 Performance of Logistic Regression 43

Table 8 Trading performance of Decision Tree...45

Table 9 Performance of Random Forest...47

Table 10 Summary of trading performance..48

Table 11 Comparative analysis of trading performance – Random Forest............60

Table 12 Sensitivity analysis of Random Forest 61

Table 13 Trading performance per year for cross-validation – Random Forest.....63

Table 14 Portfolio Analysis – All strategies...64

List of Figures

Figure 1 Methodology of strategies...9

Figure 2 Boxplot of commodity prices...10

Figure 3 Histogram of log returns 11

Figure 4 Example of MA Crossover Strategy – Crude Oil 21

Figure 5 Example of MFI-RSI Strategy – Natural Gas 23

Figure 6 Example of MSV Strategy – Sugar 25

Figure 7 Implementation of a machine learning algorithm....................................29

Figure 8 Cross validation on a rolling basis 32

Figure 9 Features Importance – Crude Oil..36

Figure 10 Features Importance – Natural Gas...37

Figure 11 Features Importance – Copper 38

Figure 12 Features Importance – Silver 39
Figure 13 Features Importance – Corn 40
Figure 14 Features Importance – Sugar 41
Figure 15 Example of Logistic Regression Strategy – Natural Gas 42
Figure 16 Example of Decision Tree Strategy – Natural Gas 44
Figure 17 Example of Random Forest Strategy – Corn 46
Figure 18 Distribution of Sharpe Ratios across commodities 49
Figure 19 Prices, log returns and volatility of Crude oil 50
Figure 20 Crude oil quarterly profit factor 51
Figure 21 Natural gas quarterly profit factor 52
Figure 22 Copper quarterly profit factor 53
Figure 23 Silver quarterly profit factor 55
Figure 24 Corn quarterly profit factor 57
Figure 25 Sugar quarterly profit factor 58
Figure 26 Cumulative returns of Random Forest strategy 60
Figure 27 Evolution of cumulative returns when removing features 62
Figure 28 Cross-validation of Random Forest in other periods 63

1. Introduction

This research investigates the effectiveness of two distinct approaches to predicting commodity markets: machine learning algorithms and technical analysis. We will backtest six distinct strategies: three based on technical indicators and three leveraging machine learning algorithms across various type of commodity futures markets such as crude oil, natural gas, silver, copper, corn, and sugar. Several performance metrics will be used to compare the algorithms such as win rate, profit factor, and Sharpe ratio. The performance of these strategies will be compared and analyzed, with a focus on how their performance evolves over time. Once we have identified the best-performing strategy, we will conduct several tests to ensure its effectiveness, robustness and how its performance compares to “traditional” investments such as the S&P500 and Swiss Market Index buy&hold strategies.

In addition to our quantitative analysis, we will complement our findings with the perspective of industry experts. These specialists will offer insights into the practical applications and real-world implications of the examined techniques within the commodity trading industry.

1.1 Commodities pricing

Commodities, which include agricultural products, metals, and energy resources, play a crucial role in the global economy (Baker, Filbeck, Harris 2018). Their pricing is affected by a range of factors such as supply and demand dynamics, geopolitical events, and macroeconomic trends (Borensztein, Reinhart 1994; Commodity price cycles: Causes and consequences 2022). In particular, commodity prices can be a leading indicator of future inflation (How Commodity Pricing May Correlate to Inflation)

Currencies, especially the US dollar, can impact commodity prices as its depreciation/appreciation can lead to an increase/decrease of purchasing power of market participants trading in a foreign currency (Hannah Baldwin 2022). Furthermore, consumer confidence data has proven to be a good predictor of some commodities, especially oil prices (Su et al. 2023).

The complexity of these elements presents an interesting ground for the development of diverse trading strategies.

1.2 Financial markets and trading

Financial trading can be defined as the process of buying and selling financial instruments such as stocks, bonds, commodities, currencies, and derivatives in various markets. Financial markets are populated by a diverse array of participants, each with their own objectives, strategies, and risk profiles. These participants may include:

- Retail Traders: Individual investors who trade for their personal accounts, ranging from casual investors to active day traders.
- Institutional Investors: Large financial institutions such as banks, hedge funds, pension funds, and mutual funds that trade on behalf of their clients or shareholders.
- Market Makers: Entities that provide liquidity by quoting both buy and sell prices for financial instruments, thereby facilitating trading and maintaining orderly markets.
- Speculators: Traders who seek to profit from price movements by taking directional bets on the market, often using leverage to amplify returns.
- Producers and suppliers: They include farmers, miners, energy companies, and agricultural cooperatives whose goal is usually to hedge against a price increase/decrease.

Market participants have always been interested in accurately predicting financial time series. Many studies have explored different methods and models. ARIMA is an example of a model based on “classical” statistics that is often used to predict financial prices (e.g. Namin, 2018). Later, many researchers focused their efforts on techniques based on AI/machine learning which will be detailed later in this paper.

1.3 Commodity contracts

Commodities are traded in specialized markets known as commodity exchanges, where standardized contracts are bought and sold. (What Is a Commodities Exchange? How It Works and Types)

The three primary types of contracts traded in commodity markets are:

- Spots
- Forwards
- Futures

Spot contracts involve the immediate exchange of a commodity for cash. This type of transaction is straightforward and is typically used for immediate delivery of the commodity. (Forward Rate vs. Spot Rate: What’s the Difference?)

Forward contracts, on the other hand, involve an agreement between two parties to buy or sell a commodity at a specified price on a future date. These contracts are customized and are not traded on exchanges but rather over the counter (OTC), which means they are not as liquid and standardized as futures contracts. (Forward Contracts vs. Futures Contracts: What’s the Difference?)

Futures contracts are standardized agreements to buy or sell a specified quantity of a commodity at a predetermined price on a specified future date. (Forward Contracts vs. Futures Contracts: What’s the Difference?).

These contracts are traded on organized exchanges, such as the Chicago Mercantile Exchange (CME) or the Intercontinental Exchange (ICE), and are highly liquid, making them attractive to traders and investors.

Each type of contract serves different purposes for market participants. Spot contracts are ideal for those looking for immediate delivery of a commodity, while forward contracts allow for customization but may lack liquidity. Futures contracts, with their standardized nature and high liquidity, are an easy method to speculate or hedge for markets participants, such as producers, consumers, speculators, and investors. (Forward Contracts vs. Futures Contracts: What’s the Difference?) (Forward Rate vs. Spot Rate: What’s the Difference?)

For these reasons, futures contracts will be used in this thesis.

1.4 Algorithmic trading

Algorithmic trading, also known as algo trading or automated trading, is a method of executing trading orders using computer algorithms or programs. Instead of relying on human decision-making, algorithmic trading relies on predefined instructions to analyze market data, identify trading opportunities, and execute trades automatically. (Basics of Algorithmic Trading: Concepts and Examples)

Algorithmic trading algorithms analyze vast amounts of market data, including price movements, trading volumes, and other relevant metrics. This analysis helps identify patterns, trends, and potential trading opportunities.

Based on the analysis of market data, algorithmic trading strategies are formulated to determine when to buy or sell financial instruments. These strategies can range from simple to complex and may incorporate various technical indicators, statistical models, or machine learning algorithms. Once a trading opportunity is identified, algorithmic trading systems automatically generate and execute trading orders according to predefined rules and parameters.

These systems typically include risk management mechanisms to control and mitigate trading risks. These mechanisms may involve setting position limits, implementing stop-loss orders, or dynamically adjusting trading parameters based on market conditions. (Bluestock 2023)

It has become increasingly prevalent in financial markets due to its potential to improve trading efficiency, reduce transaction costs, capitalize on market inefficiencies and because trading decisions are not taken by humans anymore, human error, subjectivity and emotions are removed leading to more objective trades. (Pros and Cons of Automated Trading Systems)

2. Literature review

2.1 Technical and Fundamental analysis

Technical analysis is a method that predicts price directions by examining past market data, primarily price and volume. Advocates of this approach, such as Murphy (1999), argue that price movements are not entirely random and often follow identifiable patterns. Technical analysts use various indicators such as Moving Averages, the Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) to generate trading signals, as detailed by Edwards, Magee, and Bassetti (2018), Murphy (1999) and Brock, Lakonishok and Lebaron (1992). This method operates under the assumption that historical price patterns and trends can repeat themselves, thus providing predictive value.

We were not able to find any study that can justify the use of technical analysis as a benchmark for other strategies. However, Yen,Hsu used in their research several strategies such as MSV that are similar to the strategies presented in our paper.

On the other hand, fundamental analysis focuses on assessing economic, financial, geopolitical, and other factors to determine the market's or a commodity's intrinsic value. This approach, supported by research from Ball, Kothari, and Watts (1993) and Fama (1970), operates under the premise that markets may not always efficiently incorporate new or complex information into prices, thereby creating opportunities for investors to capitalize on mispricings. Fundamental analysts argue that, over time, the market will correct these discrepancies, aligning prices with the intrinsic values of assets.

Researchers have assessed the performance of technical analysis strategies in a few papers. Yen, Hsu tested 8061 trading strategies across several commodity futures markets and found that most strategies underperform the buy-and-hold strategy. Their best strategy in crude oil yielded a Sortino ratio of 0.096 whereas the other metrics were not used in our thesis. In addition, Jevtic and Délèze used a moving-average crossover strategy as a benchmark for other strategies based on machine learning and found a Sharpe ratio of 0.69 and a profit factor of 2.83 for crude oil.

2.2 Efficient markets

The efficient market hypothesis (EMH) stands as one of the pillars of modern financial theory. Proposed by Eugene Fama in 1970, the efficient market hypothesis theorizes that financial markets are efficient in reflecting all available information, thereby rendering it impossible for investors to consistently outperform the market through active trading strategies.

The concept of market efficiency comes in three forms: weak, semi-strong, and strong. In the weak form, prices reflect all past market data, implying that technical analysis cannot consistently predict future price movements. In the semi-strong form, prices reflect all publicly available information, meaning that neither fundamental nor technical analysis can consistently outperform the market. Finally, in the strong form, prices reflect all public and private information, making it impossible for anyone, including insiders, to consistently earn abnormal returns.

Technical analysis often contrasts with the Efficient Market Hypothesis (EMH), particularly its semi-strong and strong forms but it may align more closely with the weak form of the EMH. (Bodie, Treussard, Willen 2007; Latif et al. 2011). These contrasting approaches highlight the ongoing debate about market efficiency. The EMH, in its various forms, contends that markets are efficient to different degrees in reflecting information in prices. The degree to which a market adheres to these forms of EMH often influences the effectiveness of technical and fundamental analysis. In less efficient markets, where information is not quickly or fully reflected in prices, fundamental analysis might have an edge. Conversely, in markets closer to semi-strong or strong efficiency, the value of both technical and fundamental analysis could be limited (Harrington 2003).

While the EMH has been widely debated and its validity challenged, it remains a cornerstone of financial theory. However, with the advent of machine learning and its transformative potential in finance, the EMH is being revisited and re-evaluated.

This application of machine learning challenges the traditional assumptions of the EMH, particularly its semi-strong and strong forms. If machine learning models can consistently outperform the market by leveraging patterns hidden in complex data, it could imply that markets are not fully efficient in reflecting all available information. (Singh 2024)

2.3 Machine learning in commodity trading

In recent times, the advancement of computational tools and algorithms has significantly transformed the methods of formulating and evaluating trading strategies. This evolution has made it more accessible for financial analysts and traders to analyze market data and develop sophisticated trading models (e.g., Feng, Du 2022).

Machine learning in trading is a distinct approach that can incorporate elements of both technical and fundamental analysis, but it doesn't strictly fall under either category. It involves using algorithms capable of learning from data and making informed predictions.

They are applied to vast amounts of financial data in an attempt to uncover hidden patterns and predict future market movements. These algorithms can process and analyze complex data sets far beyond human capabilities, potentially identifying inefficiencies that human traders might miss. (Shen, Jiang, Zhang)

The key advantage of a machine learning-based strategy is its adaptability to changing market conditions and its ability to process large and varied datasets. However, challenges such as overfitting and the need for extensive data for training are important concerns. (Machine Learning in Finance - Overview, Applications)

Moreover, machine learning can be used to enhance traditional trading strategies. By combining machine learning with technical and fundamental analysis, traders can potentially develop more robust and accurate models for predicting market trends. For example, machine learning algorithms can be used to optimize trading signals based on historical data or to analyze news sentiment and assess its impact on stock prices.

However, the use of machine learning in finance also raises new questions and challenges. The opacity of some machine learning models, often referred to as "black boxes," makes it difficult to understand the reasoning behind their predictions. (Mane 2019).

Many researchers have extensively explored the use of machine learning in stock trading. For example, Phuoc et al. used LSTM to predict stock prices and found that “the price forecast from the LSTM model tends to be very similar to the variation trend of the actual price on the data of the test set.”. On the other hand, Rundo et al. (2019) used LSTM to predict stock prices and obtained returns p.a. of 0.82 after transactions costs.

In contrast, the exploration of machine learning techniques in commodity trading has not been as extensive, despite its rising popularity. Jevtic and Délèze compared several machine learning models such as Random Forest, SVM and K-Nearest Neighbor in crude oil markets which yielded profit factors significantly above the buy-and-hold strategy (5.77 for the long-short strategy for Random Forest) and Sharpe ratio of 1.15. Finally, Manokhin compared several strategies based on ARIMA, GARCH and SVM in crude oil markets and he demonstrated that “due to non-linear and complex nature of oil price, traditional modelling techniques indeed performed poorly in comparison with more advanced machine learning techniques such as SVM regression". The best model (SVM) yielded returns p.a. of 19.8% and a Sharpe Ratio of 0.58

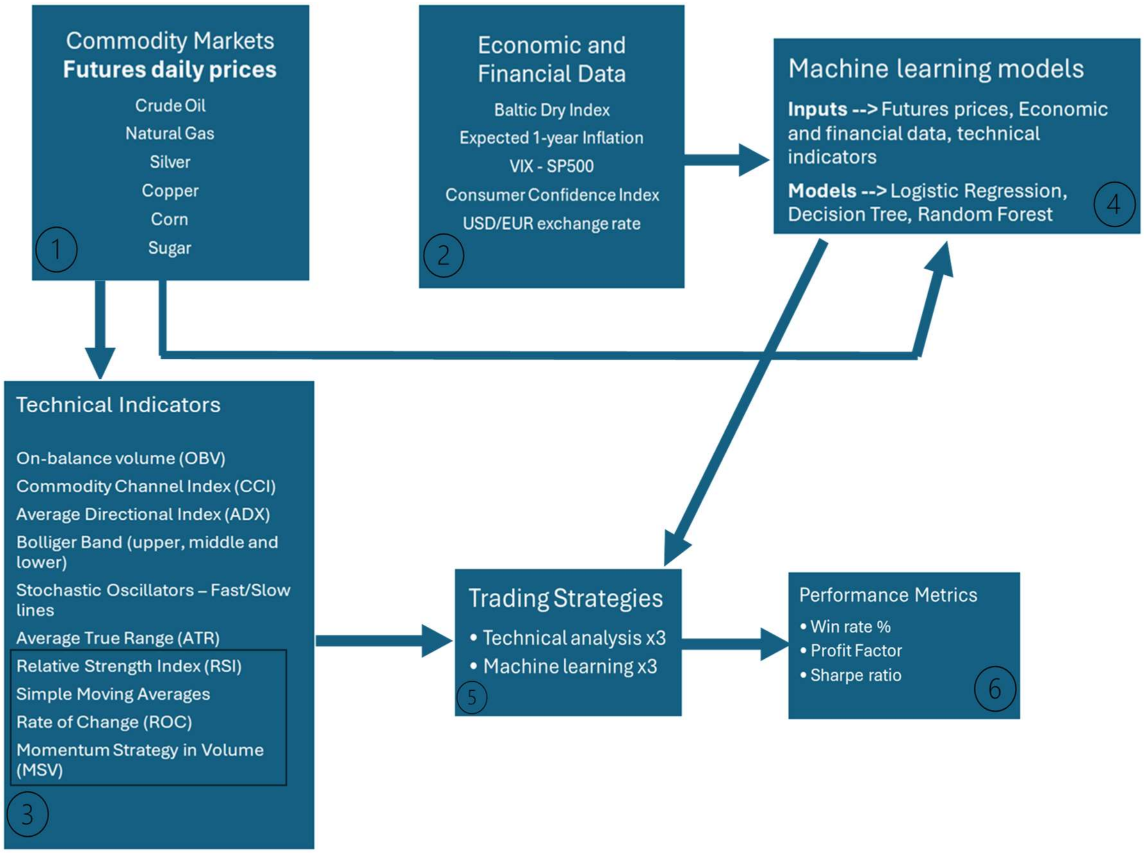
Overall, machine learning strategies are demonstrating a significant advantage in commodity and financial markets, often generating positive performance due to their superior predictive capabilities, and adaptability.

3. Methods

In the methods section, we detail the methodology that was used in this paper. First, we will explain where we obtained historic prices from and the period they cover. In addition, we will show some basic statistics and statistical charts.

Lastly, we will describe the data that was used for our machine learning strategies (including financial and economic data) and the technical indicators that were used for both technical analysis strategies and machine learning models. For the technical indicators shown, this includes how they are calculated and why they are important for our study. Finally, we will present the performance metrics that were used throughout this study.

Figure 1 Methodology of strategies



3.1 Data description

3.1.1 Commodity prices data

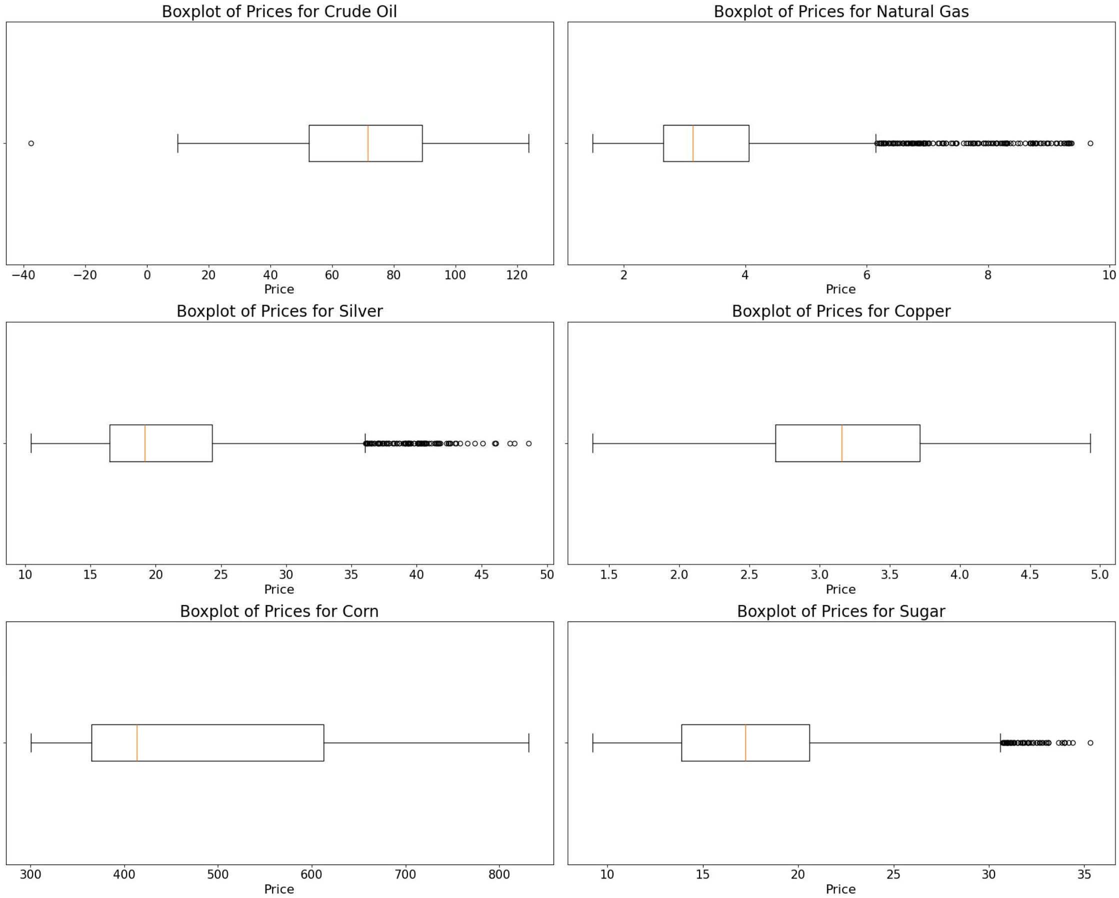
The financial data consists of daily futures prices, which have been obtained using the Python library *yfinance*, a tool that enables to fetch historical market data from Yahoo Finance. We have gathered this data for various commodities: crude oil, natural gas, silver, copper, corn, and sugar from 01-01-2009 to 12-03-2024.

The data was structured in a Pandas DataFrame, in OHLCV format, which stands for Open, High, Low, Close, and Volume.

3.1.1.1 Basis statistics

Figure 2 shows a boxplot of each commodity studied in this thesis. We note that a higher number of outliers are present in natural gas, silver, and sugar markets while high variability is present in corn.

Figure 2 Boxplot of commodity prices



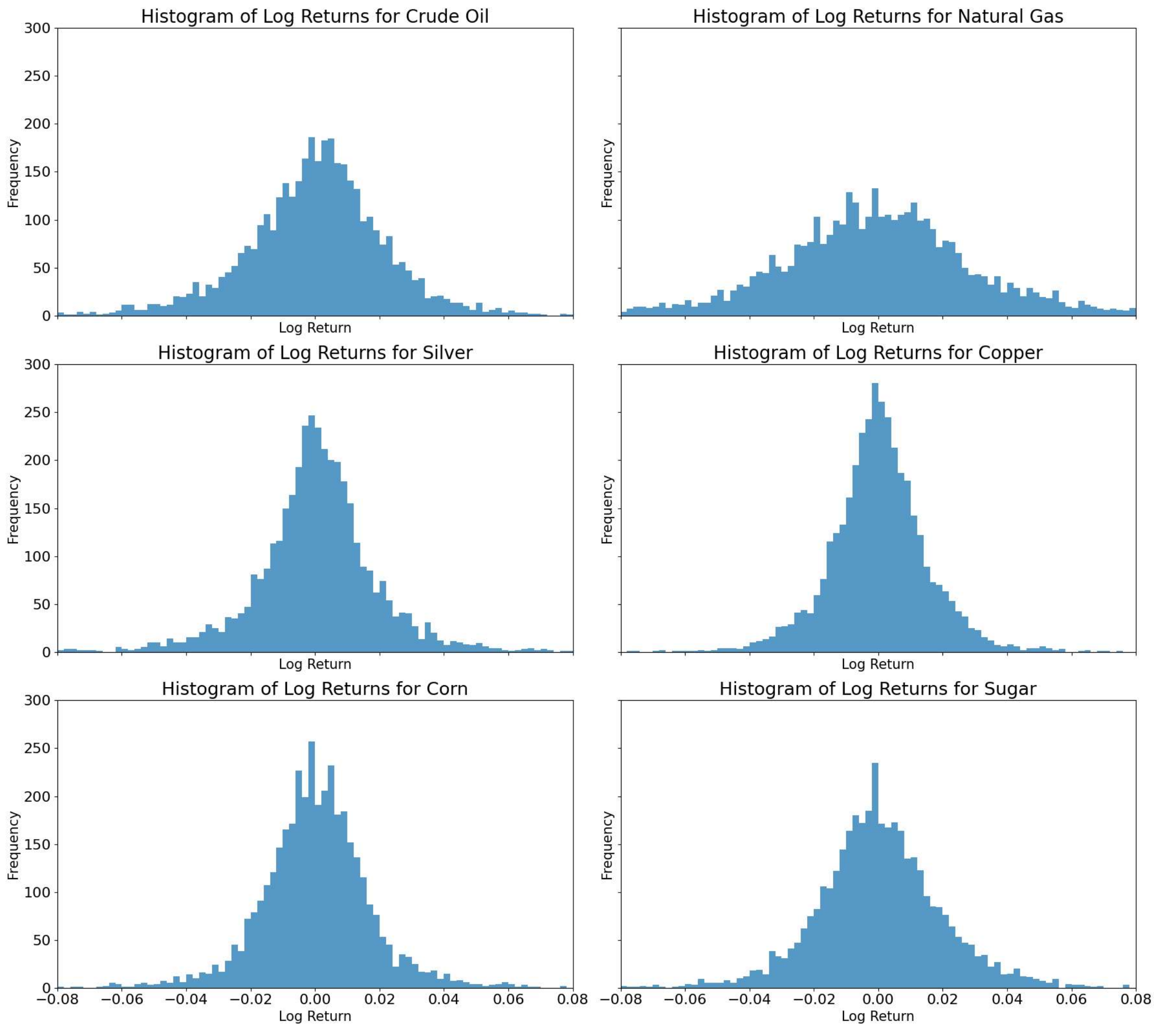
In table 1, we note that volatility is greater in energy commodities (crude oil and natural gas) while copper has the lowest volatility. The mean daily returns are relatively stable across the different markets. From figure 3, we can also see that natural gas and crude oil exhibit "fatter" tails than the other commodities.

Table 1 Basic statistics of daily returns [%]

Commodity	**Min**	**Max**	**Std. dev.**	**Mean**
Crude oil	**-307***	**37.7**	**5.99**	**0.05**
Natural gas	**-26.0**	**46.5**	**3.62**	0.03
Copper	-7.46	6.63	1.54	0.04
Silver	-13.0	13.0	1.96	0.04
Corn	-23.6	13.6	**1.82**	**0.02**
Sugar	-11.6	11.4	2.00	0.04

*This happened when prices plummeted from ~ 18$/bb to ~ -37$/bb in 2020.

Figure 3 Histogram of log returns



3.1.2 Economic and financial data

We used the *Pandas DataReader* library in Python to fetch economic data such as the 1-year expected inflation (USA), and the consumer confidence (USA) index from the Federal Reserve Economic Data (FRED) database. They are both published **on a monthly basis.**

For our machine learning model, we have also gathered financial data **(daily frequency)** which includes the USD to Euro spot exchange rate, showing the dollar's value in Euros at a given moment, and the VIX, or "fear index," measuring implied volatility in the S&P 500 stock market. Freight rates (Baltic Dry Index) were fetched from their platform and are also published on a **daily basis.**

This data was used as input variables for the machine learning strategies backtested in this research paper and the explanation on why we chose those variables will be presented in section 5.1.2.

3.2 Technical Indicators

In this section, we will present and discuss the technical indicators that have been employed in our technical analysis strategies and some in our machine learning models. We will elaborate on each technical indicator used, explaining its mathematical formulation, the logic behind its use, and how it can be interpreted in the context of this analysis. They will later be used to build our strategies; therefore, we will explain why we specifically choose them in their respective sections of this thesis.

Technical indicators are mathematical calculations based on past market data such as price, volume, and interest. Common types include **trend** indicators, which show market direction; **momentum** indicators, which indicate the speed of price changes; **volatility** indicators, which reveal market instability; and **volume** indicators, which show trading activity levels.

We chose indicators that fall into the following categories: Simple moving average **(trend)**, Bollinger bands (**volatility**), Money Flow Index **(momentum),** Rate of change **(momentum/volume),** and Momentum strategy in volume **(momentum/volume).**

3.2.1 Simple Moving Average

A simple moving average is an average of a dataset's points over a specified period, recalculated continuously as new data becomes available. It smooths out short-term fluctuations and highlights longer-term trends in data.

$$MA_t(N) = \frac{\left(\sum_{i=1}^{N-1} p_{t-1}\right)}{N}$$

The formula for calculating the simple moving average (MA) at time t for N periods, denoted as $MA_t(N)$ is given by the sum of the closing prices for the N−1 days leading up to day t, divided by N.

3.2.2 Bollinger Bands

Bollinger Bands are a technical analysis tool widely used in finance to provide a visual representation of price volatility. They consist of three lines plotted on a price chart: the middle band, which is typically a simple moving average (MA), and two outer bands, which are derived from the standard deviation of price data. The standard deviation is a measure of volatility, indicating how much the price of a security deviates from its average. The upper band is calculated by adding a specified number of standard deviations to the middle band, while the lower band is calculated by subtracting the same number of standard deviations from the middle band. The most common default setting for the number of standard deviations is two.

Upper Bollinger Band:

$$BB_{upper,t}(N, K) = MA_t(N) + K * \sigma_t(N)$$

Lower Bollinger Band:

$$BB_{lower,t}(N, K) = MA_t(N) - K * \sigma_t(N)$$

Where:

$MA_t(N) = \frac{(\sum_{i=1}^{N-1} p_{t-1})}{N}$ is the simple moving average at time t calculated over the last N periods.

$\sigma_t\,(N) = \sqrt{\frac{\sum_{i=1}^{N}(p_{t-i} - M\ \ {}_t(N))^2}{N}}$ is the standard deviation at time t calculated over the last N periods.

K is the number of standard deviations.

3.2.3 Relative Strength Index (RSI)

The Relative Strength Index (RSI) is a momentum oscillator that measures the speed and change of price movements. It oscillates between 0 and 100 and is typically used to identify overbought or oversold conditions in a market. An RSI above 60-70 (BO_{RSI}) indicates that an asset might be overbought while and RSI below 20-30 (SO_{RSI}) suggests the asset might be oversold.

Let n be the specified period (e.g., 14 days), U represent the average gain of the up periods among the selected period, and D represents the average loss of the down periods.

Average Gain (AG) and Average Loss (AL) over n periods is calculated as:

$$AL = \frac{\sum_{i=1}^{n} Li}{n}$$

$$AG = \frac{\sum_{i=1}^{n} G_i}{n}$$

Where $\sum_{i=1}^{n} G_i$ and $\sum_{i=1}^{n} L_i$ represent respectively the sum of gains/losses over the n periods.

The RSI is finally calculated as follows

$$RSI = 100 - (\frac{100}{1 + \left(\frac{AG}{AL}\right)})$$

3.2.4 Money Flow Index (MFI)

The Money Flow Index (MFI) is a technical indicator that integrates price and volume data to gauge the buying and selling pressure of a security. It is often compared to the Relative Strength Index (RSI), with the primary difference being the inclusion of volume in the calculation of MFI, which makes it a volume-weighted version of the RSI.

To define the MFI mathematically, we start with the calculation of the Typical Price (TP). The formula for the Typical Price is given by:

$$TP_t = \frac{p_t^H + p_t^L + p_t^C}{3}$$

The trading volume for the day (Vol_t) is then used to calculate the Money Flow (MF) for each period. The Money Flow is adjusted based on whether the Typical Price has increased or decreased from the previous period.

$$MF_t^+ = TP_t * Vol_t \quad if \quad TP_t > TP_{t-1}$$

$$MF_t^- = TP_t * Vol_t \quad if \quad TP_t < TP_{t-1}$$

The Money Ratio (MR) is then calculated by dividing the sum of Positive Money Flow by the sum of Negative Money Flow:

$$MR_t(N) = \frac{\sum_{i=0}^{N-1} MF_t^+}{\sum_{i=0}^{N-1} MF_t^-}$$

Finally, the MFI is calculated using the Money Ratio to generate a value between 0 and 100. The formula is:

$$MFI_t(N) = 100 - (\frac{100}{1 + MR_t(N)})$$

The MFI is interpreted similarly to the RSI, with values over 60-80 (BO_{MFI}) typically considered overbought and values under 20-30 (SO_{MFI}) considered oversold.

3.2.5 Rate of Change (ROC)

The Rate of Change (ROC) is a momentum-based technical indicator that measures the percentage change in price between the current price and the price a certain number of periods ago. The ROC is used to identify the speed at which a security's price is changing, essentially capturing the momentum or the rate at which the price is rising or falling. The rate of change can also be applied to volume, which will be the case in our trading strategies.

This adaptation helps in identifying significant changes in trading volume over time. The formula for the volume ROC is similar to the price ROC but uses volume data:

Mathematically, the ROC is expressed as:

$$ROC_{volume} = \left(\frac{V_t - V_{t-e}}{V_{t-e}}\right) * 100$$

Where:

V_t is the current volume at time t

V_{t-e} is the volume e periods ago.

e is the specified number of periods for calculating the rate of change in volume.

3.2.6 Momentum Strategy in Volume (MSV)

The MSV indicator enhances the ROC by applying a simple moving average (MA) to the ROC values, smoothing out short-term fluctuations and emphasizing longer-term trends in volume movement. The MSV is calculated using the following formula:

$$MSV_t(N) = \frac{1}{N}\sum_{i=0}^{N-1} ROC_{t-i}(e)$$

Where:

$MSV_t(N)$ is the MSV at time t

N represents the number of periods over which the MA is computed.

$ROC_{t-i}(e)$ stands for the rate of change of volume at time $t-i$, utilizing the same lookback period e.

3.3 Performance metrics

We decided to focus on a select set of performance metrics—Sharpe ratio, profit factor, and win rate to first compare the strategies because they provide essential insights into risk-adjusted returns, profitability, and consistency, which are crucial for evaluating trading strategies. This focused approach allows for a meaningful comparison across various commodities and contract specifications while avoiding unnecessary complexity. After identifying the best performing strategy, we will ensure its robustness by testing it thoroughly and by using other metrics as well.

3.3.1 Win Rate

Win rate is the percentage of profitable trades out of the total number of trades executed using a specific strategy. It's a simple but crucial metric for evaluating the effectiveness of a trading strategy, although it doesn't tell the whole story. A high win rate doesn't guarantee profitability if the losses from losing trades outweigh the gains from winning trades. Therefore, the win rate should be considered alongside other factors and ideally with a benchmark model, which is what we will do.

$$Win\ rate\ [\%] = \frac{Number\ of\ winning\ trades}{Total\ number\ of\ trades} * 100$$

3.3.2 Profit factor

Profit factor is a performance metric used to evaluate the effectiveness of a trading strategy. This metric helps traders understand the profitability of their trading strategy as it measures the ability of the strategy to generate profit over loss, irrespective of the amount of money used in each trade and the initial capital invested. By not considering the investment size, the profit factor purely reflects the strategy's operational efficiency. A profit factor greater than one indicates a profitable strategy, as it means that gross profits exceed gross losses. Conversely, a profit factor less than 1 suggests that the strategy is losing money, with gross losses surpassing gross profits.

$$\text{Profit Factor} = \frac{Total\ Gross\ Profit\ Returns}{Total\ Gross\ Loss\ Returns}$$

Where total gross profit returns are the sum of all profitable trades (that resulted in a positive return), and total gross loss is the sum of all losing trades.

In this thesis, we will consider a profit factor **above or equal to 1 to be favorable.** Consequently, we will denote the markets and strategies that meet or exceed this threshold with a checkmark.

3.3.3 Sharpe Ratio

Risk-adjusted returns are a crucial financial metric that provide insight into the performance of an investment by considering not just the returns it generates but also the risk involved in achieving those returns. (Benhamou 2021; Sharpe 1994) Calculating risk-adjusted returns allows investors to compare the performance of various investments on a level playing field, considering both the returns and the risks (Sharpe Ratio: Definition, Formula, and Examples).

This metric offers a measure to assess the performance of an investment relative to its risk.

A higher Sharpe Ratio indicates a more desirable risk-adjusted return, suggesting that the portfolio is offering more return per unit of risk. (Sharpe ratio 2024)

The Sharpe Ratio is calculated by subtracting the risk-free rate from the portfolio's returns and then dividing this result by the portfolio's standard deviation of returns. The formula is expressed as:

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

Where:

R_f is the expected portfolio return

R_f is the risk-free rate, **2.33% here**, representing the average yield of the 10-year US Treasury Note during our strategy's timeframe.

σ_p is the standard deviation of the portfolio's excess returns.

In this thesis, we will consider a Sharpe ratio **above or equal to 0.3 to be favorable.** Consequently, we will denote the markets and strategies that meet or exceed this threshold with a checkmark.

3.3.4 Benchmark Model

We established a basic model to serve as a benchmark for evaluating more complex trading strategies. Trading signals are generated as follows:

Long Signal:

- When the daily returns are positive for five consecutive days.

Short Signal:

- When the daily returns are negative for five consecutive days.

4. Technical Analysis Strategies

This section presents the empirical findings from the backtesting of technical analysis strategies.

The analysis covers a total of three different strategies based on technical indicators that were covered in detail in the methodology section. The following results were obtained in the out-of-sample dataset for better comparison with machine learning strategies. All parameters used for the strategies can be found in appendix 4.

4.1 Simple Moving Average (MA) Crossover

The strategy, known as the MA Crossover, involves using two Simple Moving Averages (MA) with different periods to generate buy or sell signals based on their crossover points.

Two MAs are calculated for the commodity's price data over two different periods (short-term and long-term). The shorter period MA reacts quicker to price changes, while the longer period MA moves slower, providing a trend indication.

Trade signals are generated when the short-term MA crosses above or below the long-term MA adjusted by a certain buffer percentage (b%). This buffer is used to filter out noise and false signals.

We decided to use this strategy as it probably is the most known technical indicator from which this strategy is derived. MA are often scrutinized by analysts and traders to uncover patterns and reversals in price (Moving Average (MA): Purpose, Uses, Formula, and Examples), one in particular known as the Death Cross (which is when the 50-day MA crossed below/above the 200-day MA) is often analyzed. (Trading the Death Cross [ChartSchool])

Long Signal:

- When the moving average $MA_t(N1)$ calculated over the last N1 periods crosses above the moving average $MA_t(N2)$ calculated over the last N2 periods, and the value of $MA_t(N1)$ exceeds $MA_t(N2)$ by b% of $MA_t(N2)$

Short Signal:

- When the moving average $MA_t(N1)$ calculated over the last N1 periods crosses below the moving average $MA_t(N2)$ calculated over the last N2 periods, and the value of $MA_t(N1)$ is less than $MA_t(N2)$ by b% of $MA_t(N2)$.

In figure 4, we show an example of the strategy showing in blue the short-term MA and in orange the long-term MA. Red arrows indicate a short position, and green arrows indicate a long position (none here).

Figure 4 Example of MA Crossover Strategy – Crude Oil

4.1.1 Results

The MA Crossover strategy demonstrates varied win rates across different commodity markets. In all the markets, the win rate is above the benchmark model although it is only slightly higher in commodities such as copper. In addition, win rates are lower than 50% for all commodities except for sugar indicating a limited proportion of profitable trades. Across different commodities, the MA Crossover strategy exhibits higher profit factors than the benchmark model. In addition, crude oil, silver, and sugar are all above 1. There are, however, significant variations across commodities. The Sharpe ratios also vary a lot. While some markets, such as sugar, silver and crude oil, show positive Sharpe ratios, others, such as copper, exhibit negative Sharpe ratios. Overall, the MA Crossover strategy demonstrates limited performance in terms of win rate, profit factor and Sharpe ratios. Using our qualitative table, we see that only the sugar strategy ticks all the boxes.

Table 2 Trading performance of MA Crossover

Commodity	Strategy	Win rate [%]	Delta (MA-Benchmark) [%]	Profit factor		Sharpe ratio	
Crude oil	MA	47.8	5.4	**1.9**	✓	**0.2**	
	Benchmark	42.5		0.8		-0.6	
Natural gas	MA	49.5	8.7	0.9		0.0	
	Benchmark	40.8		0.8		-0.5	
Silver	MA	48.9	4.0	**1.1**	✓	**0.1**	
	Benchmark	44.9		1.0		0.2	
Copper	MA	**41.9**	1.3	**0.3**		**-0.9**	
	Benchmark	40.6		0.8		-0.5	
Corn	MA	47.2	**14.3**	1.0		-0.2	
	Benchmark	32.9		0.7		-0.5	
Sugar	MA	**50.6**	**12.4**	**1.5**	✓	**0.3**	✓
	Benchmark	38.2		0.8		-0.6	

4.2 Dual Momentum (MFI – RSI) Crossover

This trading strategy combines two technical indicators, the Money Flow Index (MFI) and the Relative Strength Index (RSI).

We have chosen this strategy first because both indicators are widely recognized and used by traders for their effectiveness in assessing market momentum and identifying potential price reversals (Murphy 1999). In addition, by integrating these two indicators, the strategy aims to capitalize on short-term price movements driven by shifts in market sentiment and momentum, particularly in commodity markets known for their volatility.

The strategy initiates a long/short position under the following conditions:

Long signal:

- The MFI crosses above a specified threshold, indicating a shift from an oversold condition to a more neutral or bullish sentiment. This crossover is considered a bullish signal, suggesting increasing buying pressure. Simultaneously, the RSI crosses above its own specified threshold, further confirming the bullish momentum by indicating that the price is moving out of the oversold territory. The entry is triggered when both these conditions align.

Short signal:

- The MFI crosses below a certain threshold, signaling that the asset might be overbought. This condition implies that the buying pressure is possibly overextended, and a price pullback could be imminent. In conjunction with the MFI, the RSI also crosses below its overbought threshold. The entry is triggered when both these conditions align.

Figure 5 Example of MFI-RSI Strategy – Natural Gas

4.2.1 Results

The MFI-RSI Crossover strategy demonstrates higher win rates compared to the benchmark model across different commodity markets. For instance, in the crude oil market, the MFI-RSI Crossover strategy achieves a win rate of 70.97%, while the benchmark model has a lower win rate of 42.47%. The delta is also high across most commodities.

The strategy also exhibits higher profit factors than the benchmark. This indicates that the MFI-RSI Crossover strategy generates more profit relative to losses, suggesting its potential effectiveness in generating profits. There are, however, important variations in performance and some markets have very low profit factors (silver). Regarding Sharpe ratios, while some markets, such as copper and sugar, show positive Sharpe ratios others exhibit negative Sharpe ratios. Overall, the MFI-RSI Crossover strategy demonstrates limited performance, and no strategy meets all the requirements.

Table 3 Trading performance of MFI-RSI Crossover

Commodity	Strategy	Win rate [%]	Delta (MFI-RSI-Benchmark) [%]	Profit factor		Sharpe ratio	
Crude oil	MFI-RSI	**70.97**	**28.5**	**4.2**	✓	0.0	
	Benchmark	42.47		0.8		-0.6	
Natural gas	MFI-RSI	60.87	20.1	0.7		-0.5	
	Benchmark	40.81		0.8		-0.5	
Silver	MFI-RSI	44.23	-0.7	**0.6**		**-0.5**	
	Benchmark	44.89		1.0		-0.2	
Copper	MFI-RSI	**67.27**	26.6	**1.2**	✓	**0.1**	
	Benchmark	40.64		0.8		-0.5	
Corn	MFI-RSI	65.00	**32.1**	0.9		-0.4	
	Benchmark	32.86		0.7		-0.5	
Sugar	MFI-RSI	**77.05**	**38.9**	**1.5**	✓	**0.2**	
	Benchmark	38.16		0.8		-0.6	

4.3 Momentum Strategy in Volume (MSV)

This strategy employs a “dual-layered” moving average system, where trades are initiated based on the interaction between a short-term and a long-term moving average of the volume's rate of change. The underlying logic mirrors traditional momentum strategies, but with a focus on volume rather than price. Incorporating volume data into the strategy provides a different insight into the market's momentum, offering a distinct perspective than conventional price-based indicators.

The MSV indicator is used to generate trading signals based on the relationship between the short-term and long-term MSVs. If a holding period is active (post-entry), the strategy waits until this period expires before evaluating new entry conditions. The days to hold position before exiting is defined with the parameter g.

Long signal:

- This signal is triggered when the short-term $MSV_t(N1)$ exceeds the long-term $MSV_t(N2)$ by a predefined percentage threshold (b%) over the last lookback period defined as e. This scenario suggests an uptick in trading volume that could precede an upward price movement, signaling an opportunity to initiate a long position.

Short signal:

- Conversely, a signal for a short position is generated when the short-term $MSV_t(N1)$ falls significantly below the long-term $MSV_t(N2)$, beyond the threshold (b%). It indicates a drop in trading volume that might indicate a downward price trend, signaling an opportunity to enter a short trade.

Figure 6 Example of MSV Strategy - Sugar

4.3.1 Results

Across all commodities listed, the MSV strategy consistently demonstrates higher win rates compared to the benchmark model.

In specific markets such as silver and corn, the MSV strategy stands out with notably higher win rates compared to the benchmark model. In corn, for example, the MSV strategy's win rate of 52.0% surpasses the benchmark model's 32.9%.

Regarding profit factor, the strategy consistently demonstrates comparable or higher profit factors compared to the benchmark, but they remain below 1 for most commodities.

In addition, the Sharpe ratios of the MSV strategy have high variability. In markets such as crude oil and natural gas, the MSV strategy shows positive but relatively low Sharpe ratios.

Overall, the strategy demonstrates relatively bad performance.

Table 4 Trading performance of MSV

Commodity	Strategy	Win rate [%]	Delta (MSV-Benchmark) [%]	Profit factor		Sharpe ratio	
Crude oil	MSV	**50.0**	7.5	0.9		-0.1	
	Benchmark	42.5		0.8		-0.6	
Natural gas	MSV	50.0	9.2	**1.0**	✓	**0.1**	
	Benchmark	40.8		0.8		-0.5	
Silver	MSV	54.9	**10.0**	**1.5**	✓	**0.8**	✓
	Benchmark	44.9		1.0		-0.2	
Copper	MSV	49.0	8.4	0.8		**-0.4**	
	Benchmark	40.6		0.8		-0.5	
Corn	MSV	**52.0**	**19.1**	0.9		-0.1	
	Benchmark	32.9		0.7		-0.5	
Sugar	MSV	45.1	6.9	0.9		-0.2	
	Benchmark	38.2		0.8		-0.6	

5. Machine Learning Models

In this chapter, three machine learning strategies will be presented and backtested across six commodities.

For each model, a signal vector (-1 or 1) was generated corresponding respectively to a decrease or increase in prices over a specified period (in this case, the next trading day). The models therefore aim to forecast whether the market will increase or decrease over this specified period. These signals are used as target variables.

$$Signal_t = \begin{cases} 1\ if\ return_{t+1} > 0 \\ -1\ if\ return_{t+1} \leq 0 \end{cases}$$

In the following section, we will detail the implementation workflow from gathering data to performance assessment. Next, we will detail the input variables that were used and the motivation behind their selection. Finally, we will explain the cross-validation strategy we have employed to ensure the robustness of our results.

5.1 Structure of the algorithms

5.1.1 Implementation workflow

The method and structure used to build our models is explained step by step in figure 7. This workflow is segmented into five distinct stages, ensuring systematic progression from raw data to the backtesting of our trading strategies.

The initial stage is data collection, a phase where key decisions are made regarding the software for data processing, the data source, and the inputs needed and that will lead to a profitable strategy. It is therefore key to consider a range of options and ensure there is no overlap between them. A correlation analysis can be conducted to ensure we avoid this. Next, Python's libraries are used to collect economic and financial data: *Pandas Datareader* and *yfinance*. In addition, the technical indicators were calculated using TA-lib, a library specialized in technical analysis.

After data collection, the data was cleaned using Python's panda's library. This step is vital because datasets often contain incomplete or missing (NA) values which need data cleaning. For this specific case, the data was completed by filling in missing datapoints with the most recent known values (known as forward-fill method) especially for the economic data that is released every month.

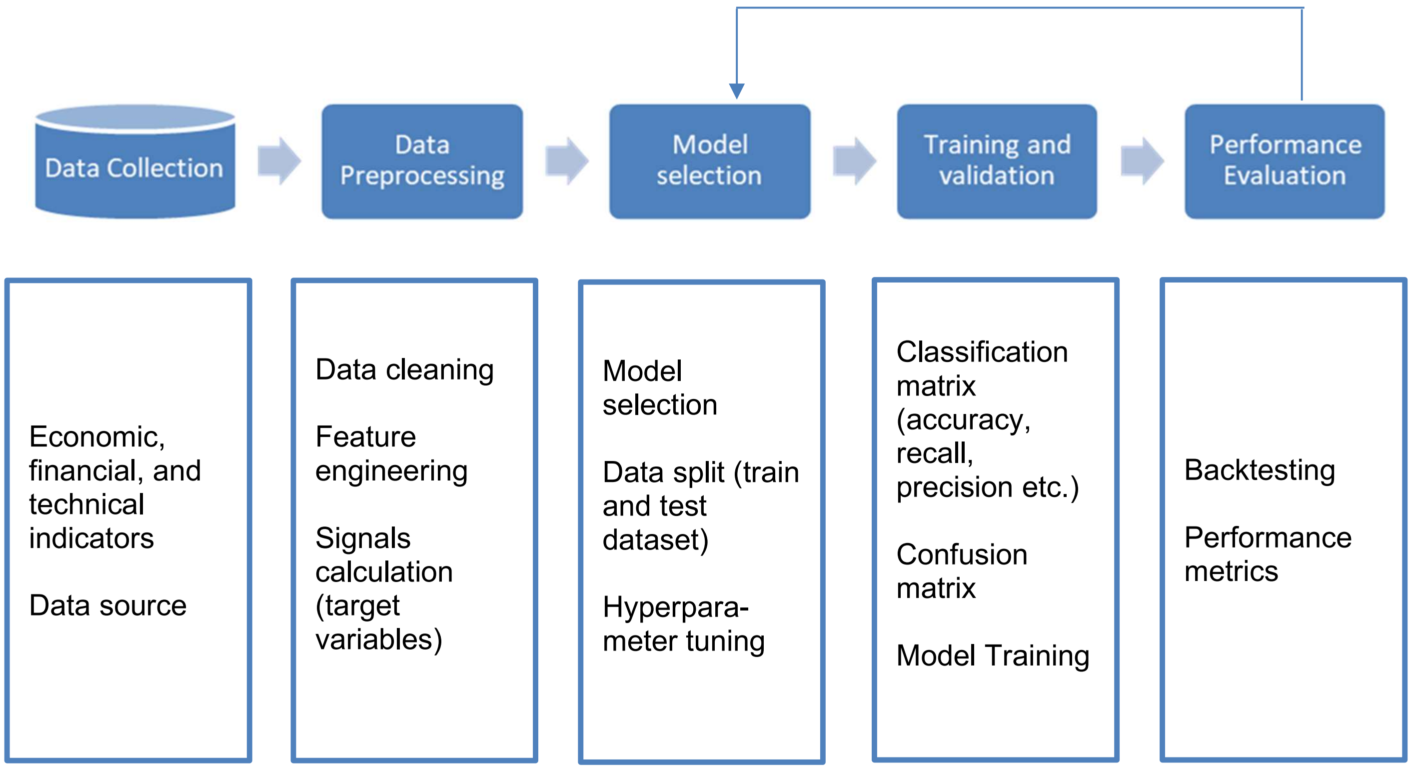
Next, trading signals are created for the daily futures prices. This means calculating each period as a gain (1) or a loss (-1) based on the returns of the next trading day. This served as target variables. At the end, the input variables were normalized with the function StandardScaler of the Sklearn library. This function transforms the data such that each feature has a mean of 0 and a standard deviation of 1. This was used because some models perform better when the data is normalized beforehand.

The third phase centers on selecting and configuring the ML model. This involves delving into the specifics of available algorithms, understanding their functionality, parameters, and usage examples, which are typically documented on relevant programming language websites. The dataset is split into training and test sets, facilitating hyperparameter tuning to optimize the model based on the given data. For hyperparameter tuning we used RandomizedSearchCV which is a function that searches for the hyperparameters that lead to the highest accuracy for each model.

In the fourth step, the models were trained on historical data. This step also means evaluating the model's predictions against the actual outcomes (is the model predicting an increase or decrease of returns the next trading day or not and is it accurate). Consistent evaluation methods are applied regardless of the model being used. In this case, classification reports and confusion matrices were used to ensure that the models were performing as expected.

Finally, we backtested our strategies on the out-of-sample dataset to ensure the models could work well in real-life situations by testing them with past data. This part was done with a library named *backtesting*. Performance metrics were then manually calculated based on the models' trades.

Figure 7 Implementation of a machine learning algorithm



5.1.2 Input variables

Several input variables were selected and analyzed to assess their predictive power. The variables are categorized into three main types: technical indicators, which are derived from market data, financial indicators which are derived from financial data, and economic indicators, which represent broader economic trends.

The core idea is to identify key features, with the expectation that they will reveal underlying patterns within the data and will allow us to predict commodity returns. These patterns are important in forecasting whether the market will experience an upward or downward trend the following day.

Referencing table 5, one can observe a comprehensive list of the input variables that were used across the models.

We decided to use these inputs for several reasons. First, inflation, and consumer confidence directly impact the supply and demand for commodities and therefore prices (Ajmera, Jook and Crilley, 2012)

The currency data ($) was chosen because commodities are traded globally, often denominated in US dollars. Therefore, fluctuations in currency exchange rates can significantly impact commodity prices. (Q. Farooq Akram, 2009)

The VIX volatility index was chosen as several studies pointed out that the returns of some commodities are influenced by the implied volatility of the stock market. (Hoon Kang, Maitra, Dash et al, 2019).

Finally, we decided to use the Baltic Dry Index as it is a direct measure of the cost of freight for dry commodities.

Table 5 Description of input variables

Type	Input variable
Technical	On-balance volume (OBV)
	Commodity Channel Index (CCI)
	Delta MA15-30 (DELTA_MA15-30)
	Delta MA50-200 (DELTA_MA50-200)
	Average Directional Index (ADX)
	Relative Strength Index (RSI)
	Bolliger Band (upper)
	Bolliger Band (middle)
	Bolliger Bands (lower)
	Stochastic Oscillator – Fast line (STOCH_K)
	Stochastic Oscillator – Slow line (STOCH_D)
	Average True Range (ATR)
Economic	1-Year Expected Inflation - USA
	Consumer Confidence Index - USA
	Baltic Dry Index
Financial	U.S. Dollars to Euro Spot Exchange Rate
	Volatility - VIX

5.1.3 Strategy cross-validation

In our study, two type of datasets have been used to train and test our models:

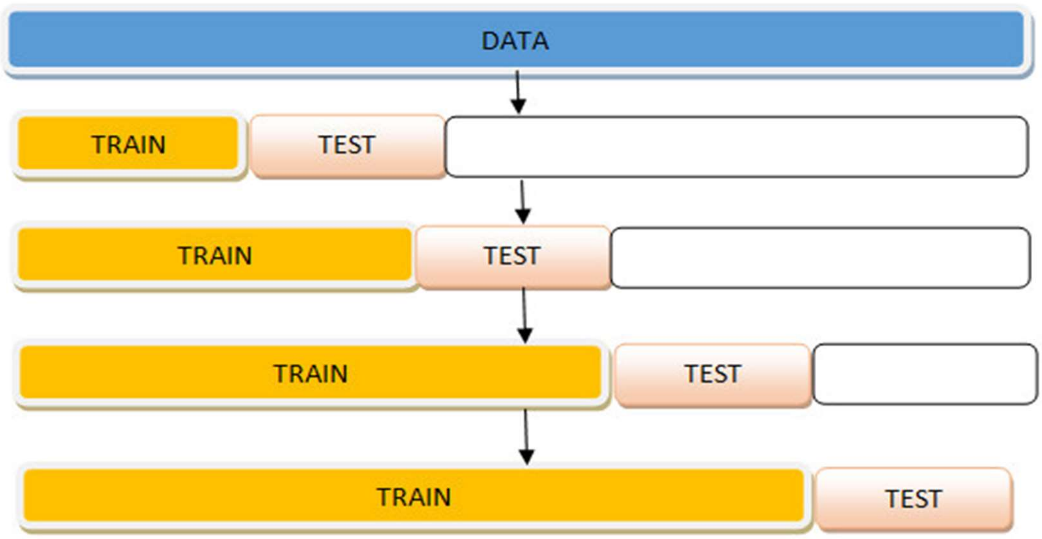
- In-sample dataset: This subset of the data is used to train the machine learning models. It contains historical financial data up to a certain point in time, and the models use this data to learn the patterns, trends, and relationships between different market variables. **This period lasts from 01-01-2009 to 02-08-2019 (70%).**

- Out-of-sample dataset: This is the data that the model has never seen before, set aside specifically for evaluating its performance. The out-of-sample dataset is used to evaluate how well the trained model can generalize its learned patterns to new, unseen data for machine learning. It is a crucial test of the model's effectiveness in live trading. Technical strategies will also be compared on this dataset as it easier to compare the six strategies on the same period. **This period lasts from 03-08-2019 to 12-03-2024 (30%).**

In the context of this research, we have used **rolling cross-validation**. The initiation training phase involves training the trading models on a subset of historical data, representing 70% of the total dataset period.

After the initiation training phase, the model is then tested on a separate subset of data to evaluate its performance. This testing phase spans a fixed period of time, 100 days in our case. The initiation training and testing phases are part of a rolling process where the training dataset gradually expands to include more recent historical data, and the testing period remains fixed at 100 days. By following this approach, we ensure that the models are trained on a sufficient amount of historical data to capture relevant market patterns while also providing a standardized testing period for evaluating their performance.

Figure 8 Cross validation on a rolling basis (Erfanian, Zhou, Razzaq et al. 2022)



5.2 Model Evaluation and Interpretation

5.2.1 Classification Report

To evaluate the predictive performance of our machine learning models, we have generated a comprehensive classification report, as depicted in table 6. This report provides essential insights into the quality of predictions across various classes within our dataset. The metrics assessed in the classification report include precision, recall, F1 score, and accuracy.

Precision measures the proportion of positive predictions that are actually correct. It is especially important in scenarios where the cost of false positives is high. It is defined as:

$$2 * (Recall\ *\ Precision) \,/\, (Recall\ +\ Precision)$$

Recall, on the other hand, delves into the model's ability to correctly identify all relevant instances within the dataset. It signifies the percentage of actual positive cases that were correctly classified as positive by the model. It is defined as:

$$\frac{True\ Positives}{True\ Positives + False\ Positives}$$

The F1 score offers a balanced assessment of the model's performance, with one indicating the best possible performance and zero representing the worst. It is defined as:

$$\frac{True\ Positives}{True\ positives{+}False\ Negatives}$$

Accuracy measures the overall correctness of predictions made by the models, representing the proportion of correctly classified instances out of the total instances in the dataset. It is defined as:

$$\frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

For Logistic Regression, the precision scores vary across commodities but generally hover around the 0.5 mark. This indicates that the model tends to make a moderate number of correct positive predictions relative to the total number of positive predictions. Similarly, the recall scores are moderate, indicating that the model identifies a fair proportion of true positives out of all actual positives. The F1-scores are also in the moderate range, suggesting a balance between precision and recall. The accuracy scores are generally consistent with the F1-scores, indicating that the model's overall performance in correctly classifying instances across all commodities is fair.

For Decision Tree, the precision scores are also around 0.5, indicating a similar performance to Logistic Regression in terms of correctly identifying positive instances. The recall scores are also moderate, suggesting that it identifies a reasonable proportion of true positives. The F1-scores are consistent with precision and recall, indicating a balanced performance. Finally, the accuracy scores are also around the 0.5 mark, similar to Logistic Regression, indicating a fair overall performance across all commodities.

For Random Forest, precision scores are slightly higher compared to Logistic Regression and Decision Tree, particularly for some commodities. Recall scores are also in the same range as the other models, indicating a similar ability to identify true positives. On the other hand, F1-scores are relatively consistent, suggesting a balanced performance. Finally, accuracy scores are generally similar to the other models, indicating a fair overall performance. Across all models, there is a similarity in performance metrics, with precision, recall, F1-score, and accuracy hovering around moderate levels. While there are slight variations in performance across different commodities, the overall patterns remain consistent. Random Forest slightly stands out compared to the other models in crude oil corn and sugar.

Table 6 Classification Report of Machine Learning Models

Crude Oil	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.49	0.51	0.48	0.51
Random Forest	0.50	0.50	0.50	0.50
Decision Tree	0.49	0.50	0.49	0.50

Natural gas	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.51	0.51	0.51	0.51
Random Forest	0.49	0.49	0.49	0.49
Decision Tree	0.50	0.51	0.50	0.51

Copper	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.51	0.51	0.51	0.51
Random Forest	0.53	0.52	0.52	0.52
Decision Tree	0.53	0.53	0.53	0.53

Silver	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.52	0.52	0.52	0.53
Random Forest	0.56	0.52	0.52	0.52
Decision Tree	0.50	0.50	0.50	0.50

Corn	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.51	0.51	0.50	0.51
Random Forest	0.53	0.53	0.53	0.53
Decision Tree	0.53	0.52	0.52	0.52

Sugar	Precision	Recall	F-1 score	Accuracy
Logistic Regression	0.50	0.50	0.50	0.50
Random Forest	0.53	0.53	0.52	0.53
Decision Tree	0.49	0.50	0.48	0.49

5.2.2 Feature importance

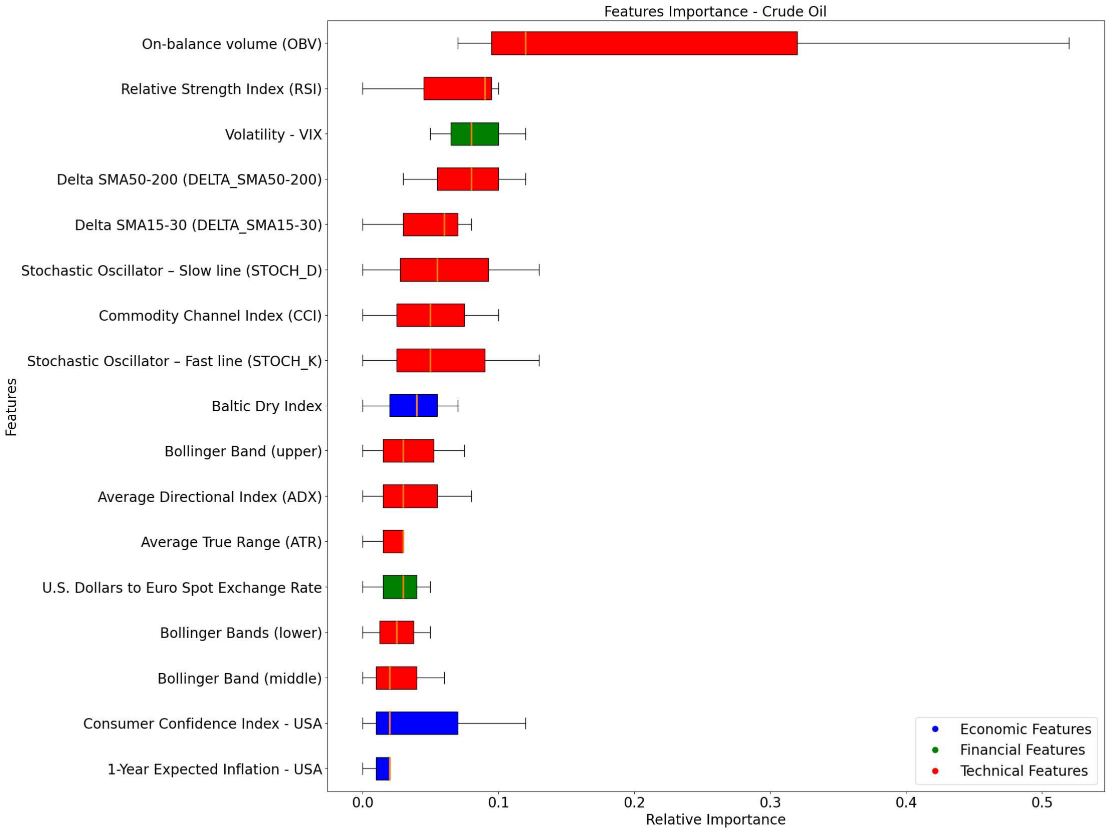
Feature importance analysis refers to methods that are used to calculate the importance of all the input variables used in a machine learning algorithm. It is a fundamental concept that helps us understand the influence individual features have on the predictions of a model. A boxplot will display the feature importance for our three models, arranged in descending order based on the median value for each feature.

We opted to use permutation for analyzing feature importance.

5.2.2.1 Crude oil

The results indicate varying degrees of effectiveness across different features and trading strategies. For instance, features like OBV and RSI show the best performance while volatility is the first financial input we see in the chart. This may indicate that knowing the volume, if the commodity is overbought or oversold and the current volatility of the stock market may be great predictors of futures returns. This may be explained by the high volatility inherently present in the energy markets but to confirm this we would have to check if the market's volatility is correlated with VIX. On the other hand, economic indicators showed mixed results. For instance, the consumer confidence index and 1-year expected inflation were last. To enhance the choice of input variables, we believe it would be beneficial to remove these "generalist" economic indicators to replace them with some that are directly related to the commodity, in this case oil production, exports, etc. This would certainly provide variables with greater importance, but further analysis should be conducted.

Figure 9 Features Importance – Crude Oil

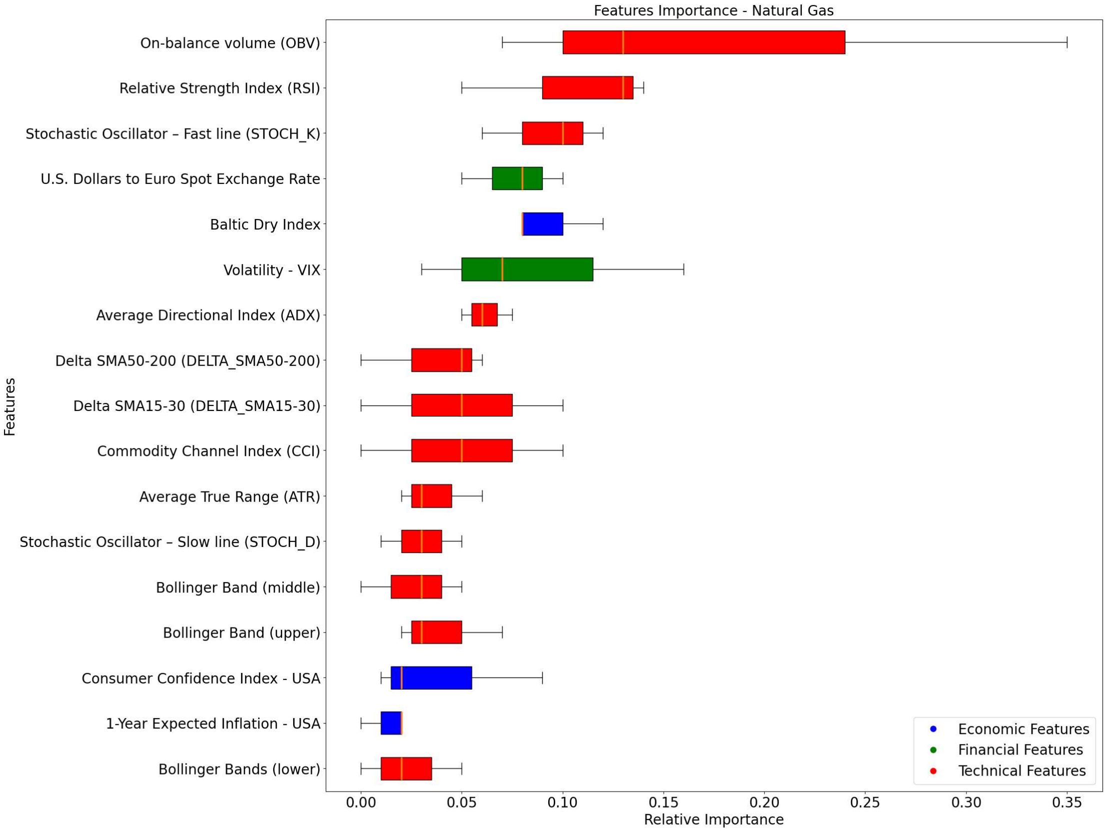


5.2.2.2 Natural gas

Given the unique dynamics of the natural gas market, certain features naturally hold more predictive power than others. For instance, we believe the strong influence of technical indicators such as OBV, STOCK_K, and RSI makes sense, as natural gas is a highly traded commodity with prices often driven by short-term supply and demand fluctuations. These indicators capture market sentiment and momentum, crucial for understanding these types of price movements.

However, the moderate importance of economic factors such as the U.S. Dollars to Euro Spot Exchange Rate and Consumer Confidence Index (USA) warrants a closer look. While these factors can indirectly impact natural gas demand (through industrial activity or consumer spending), their influence might be less direct and pronounced compared to crude oil, which is more tightly linked to global economic activity. Similarly, while the Baltic Dry Index (BDI) is a useful indicator of overall economic health, its connection to natural gas prices is significantly less robust than for dry commodities which may indicate that noise is present for the next inputs. The VIX is again relevant here as well. To enhance the choice of input variables, we think it could be beneficial to remove generalist and lagging economic indicators and add inputs such as weather data, production, storage levels etc. and focus on better granularity.

Figure 10 Features Importance – Natural Gas



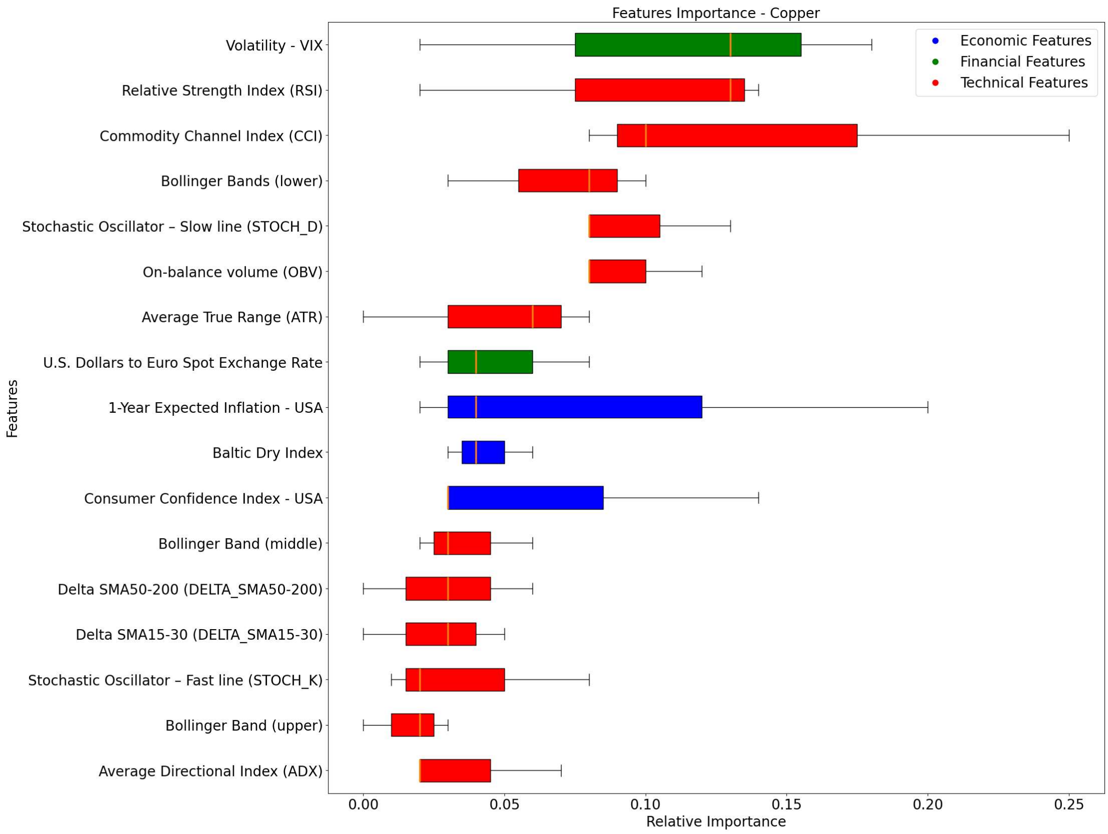
5.2.2.3 Copper

For copper, technical, economic indicators and financial indicators play significant roles. Among technical indicators, the CCI and RSI are highly influential, emphasizing the importance of momentum and trend strength. The lower Bollinger Bands and the ATR also show considerable importance, reflecting their effectiveness in capturing price volatility and range. The VIX is a top financial indicator again.

Economic indicators, particularly the 1-year expected inflation in the USA, exhibit substantial importance, suggesting that inflation expectations significantly affect copper prices. Other important economic indicators include the Consumer Confidence Index and the Baltic Dry Index. This is interesting as copper is an important industrial input for the economy.

In summary, while economic indicators are crucial, especially inflation expectations, technical indicators dominate the prediction model. The VIX plays a significant role as well. To enhance the choice of input variables, we believe it is beneficial to conduct a thorough analysis of technical indicators to keep those with greatest importance, add different ones and/or with different granularity and remove those that do not work well for this specific market. We also believe that market-specific inputs could work well. For example, production, industrial demand, manufacturing activity, etc.

Figure 11 Features Importance – Copper



5.2.2.4 Silver

Given the unique dynamics of the silver market, the feature importance ranking aligns with expectations to a certain extent. The dominance of technical indicators such as ADX, RSI, and the delta SMA50-200 is understandable, as silver is a heavily traded commodity with prices often driven by short-term speculative activity. These indicators capture momentum, and trends, which are crucial for understanding silver price movements.

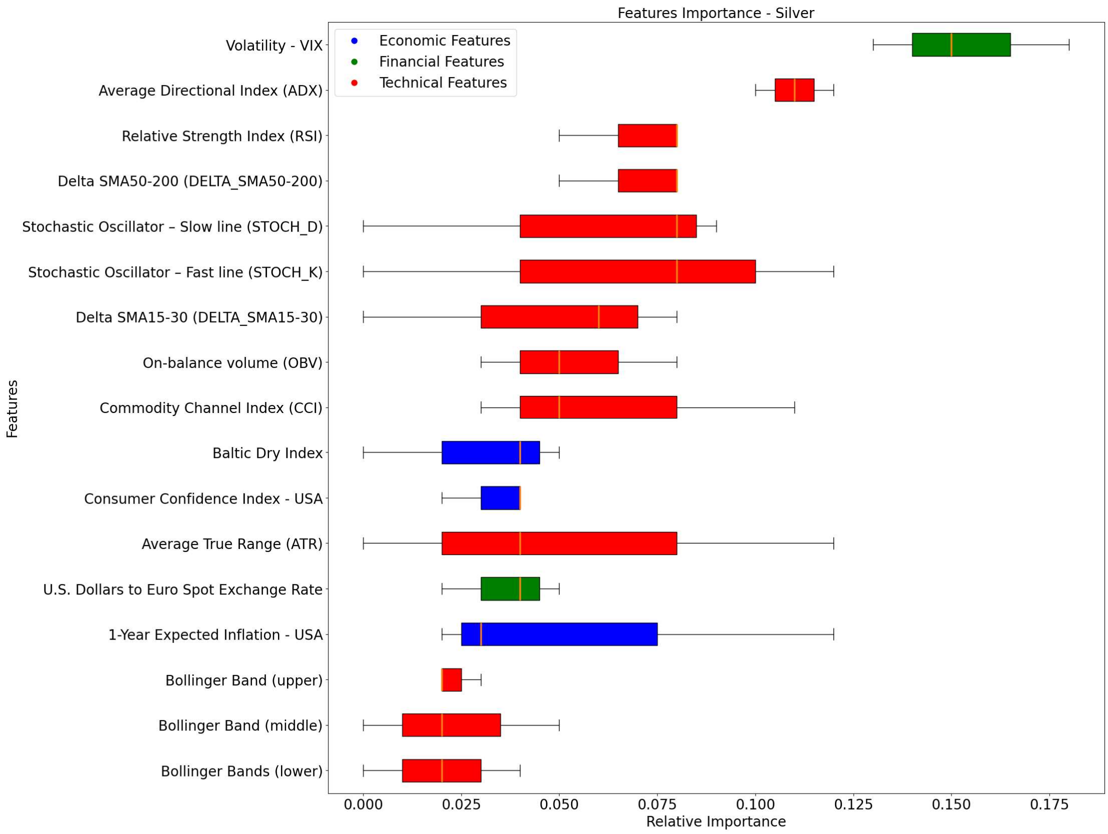
However, the moderate importance of economic factors such as the U.S. Dollars to Euro Spot Exchange Rate and inflation expectations is somewhat surprising. As a traditional safe-haven asset, silver is often sought during times of economic uncertainty or high inflation. One might expect these factors to have a more significant impact on its price. It's possible that these factors are already reflected in the technical indicators, or that the specific time period used for training the model might have influenced their importance.

The relatively low importance of the Consumer Confidence Index (USA) is also noteworthy. Unlike commodities directly tied to consumer spending, silver is primarily driven by investment demand and macroeconomic factors.

The Baltic Dry Index (BDI) has limited importance. While it indicates global shipping activity, its relevance to silver prices might be limited compared to commodities more heavily reliant on maritime transportation.

To enhance the choice of input variables, we believe it would be beneficial to add market-specific inputs such as production and inventory data, mines output, etc.

Figure 12 Features Importance – Silver



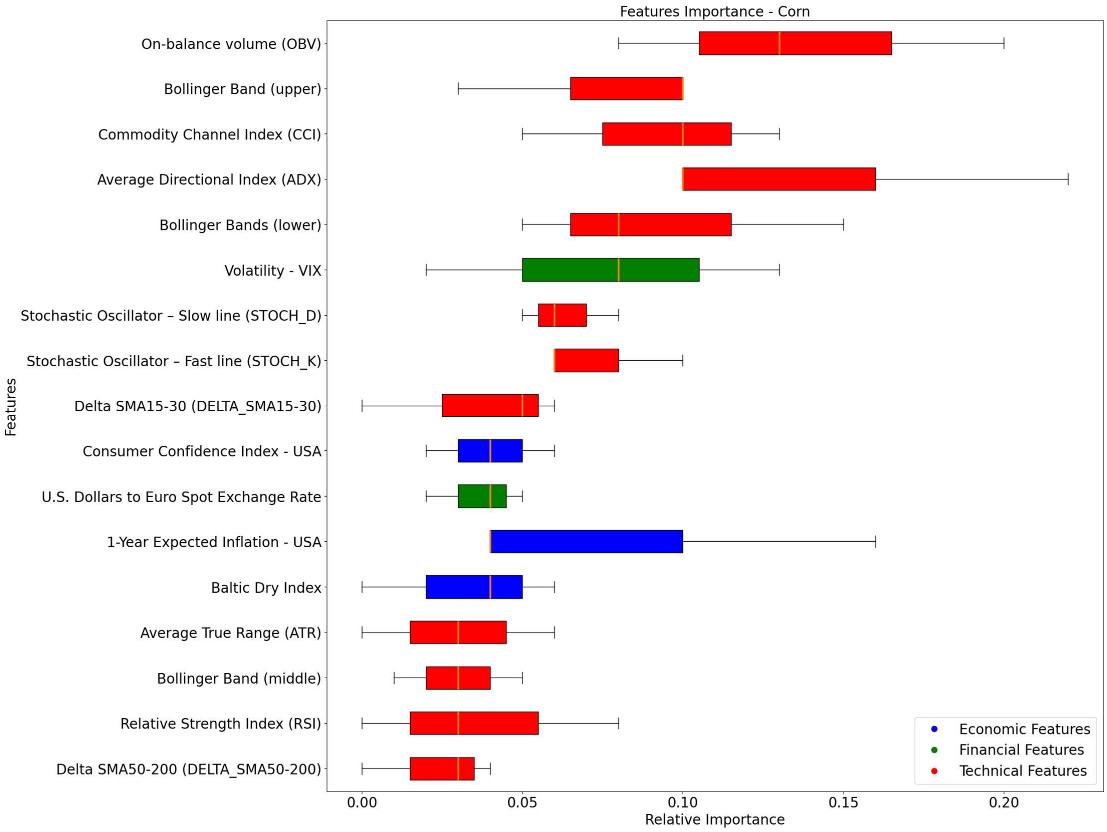
5.2.2.5 Corn

Given that corn is a globally traded agricultural commodity heavily influenced by both market dynamics and environmental factors, the feature importance ranking aligns somewhat with expectations. The prominence of technical indicators such as OBV, Bollinger Bands, and CCI could be explained by the fact that they capture short-term price fluctuations, volatility, and momentum driven by trader sentiment and speculative activity. In addition, the relatively moderate importance of economic factors such as inflation expectations and exchange rates, while still relevant for agricultural commodities, might suggest that corn prices are more sensitive to other inputs which are not directly captured in this model.

The relatively low importance of the Consumer Confidence Index (USA) could be attributed to the fact that corn is primarily used for animal feed and ethanol production, making its demand less sensitive to immediate changes in consumer sentiment compared to other food commodities.

To enhance the choice of input variables, we believe it would be beneficial to add market-specific inputs such as projected yield, crop conditions, harvest progress, weather events etc. and remove inputs such as the Baltic Dry Index, expected inflation and Consumer Confidence data as they carry little importance here.

Figure 13 Features Importance – Corn

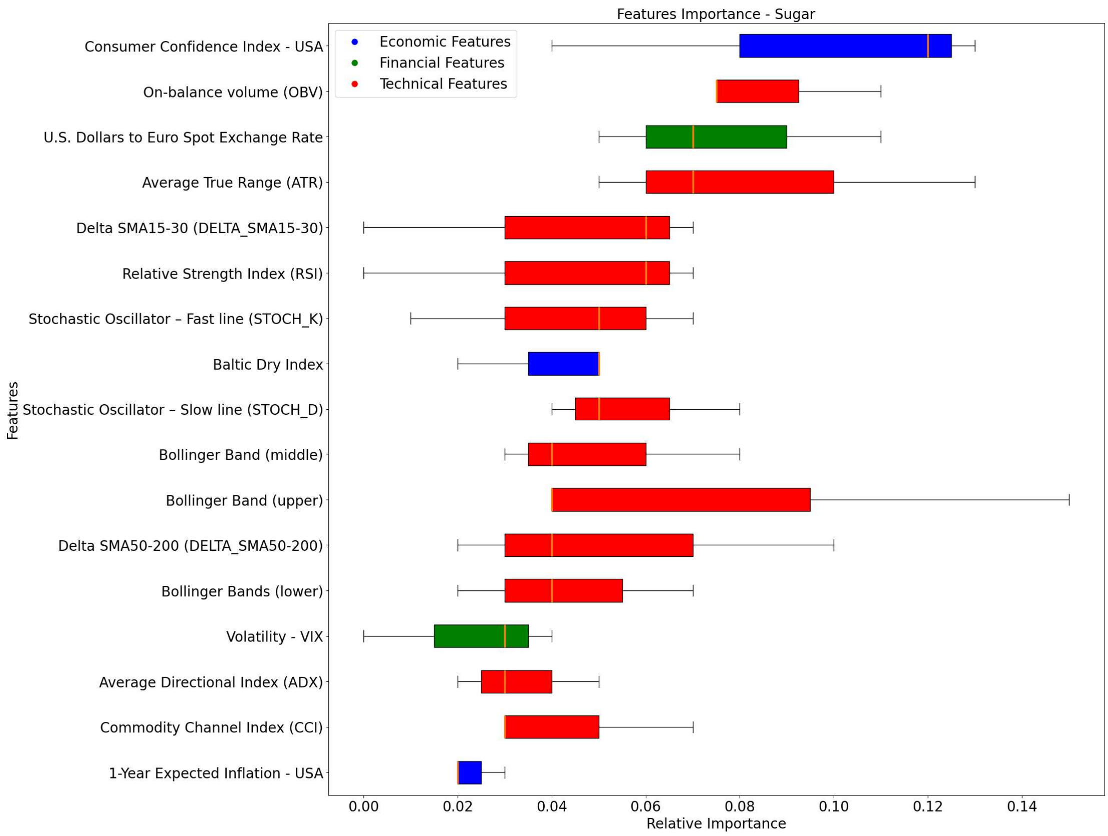


5.2.2.6 Sugar

The Consumer Confidence Index (USA) emerges as the most influential factor, surprisingly overshadowing traditional technical indicators. This suggests that consumer sentiment and expectations play a pivotal role, likely due to their indirect impact on sugar demand through processed food and beverage consumption. While technical indicators such as OBV and ATR retain some importance in capturing market dynamics and volatility, their influence is secondary to consumer confidence. This highlights the need for a comprehensive approach to sugar price prediction, incorporating both technical analysis and broader economic indicators to accurately forecast market trends. The USD exchange rate also ranks at the top.

To enhance the choice of input variables, we believe it would be beneficial to add market-specific inputs such as export demand, sugar production, biofuel production, harvested areas, projected yield, etc. Additional analysis of the inputs, especially of the technical indicators that perform the best, should also be conducted.

Figure 14 Features Importance – Sugar



5.3 Logistic Regression

Logistic Regression is a statistical method that models the probability of a binary dependent variable based on one or more independent variables. In our context, the dependent variable is the binary signal indicating buy or sell. The independent variables include a range of inputs. The logistic model operationalizes these forecasts by employing a logistic function. This curve transforms any input of real-valued numbers into a new value within a bounded range of 0 to 1. This feature is particularly useful in our context as it translates the linear combination of input variables into a probability measure. Thus, each outcome — be it a buy or sell signal — is associated with a specific probability value, indicating the confidence level of the prediction based on the current input data.

The advantage of Logistic Regression lies in its ability to handle variables that predict an outcome that can only be one of two possible states. It's particularly well-suited for our context where the decision is binary: short or long.

Figure 15 Example of Logistic Regression Strategy – Natural Gas

5.3.1 Results

The LR strategy generally demonstrates varied win rates across different commodity markets compared to the benchmark. In some markets, such as sugar, the win rates of the LR strategy are higher than the benchmark model. However, in other markets they remain relatively close to the benchmark with a delta of only 2.6 for crude oil.

Across various commodities, the LR strategy exhibits varied profit factors compared to the benchmark, but they only reached values above 1 in silver and copper markets. In the other markets, the model performed badly and was even below the performance of the benchmark in some cases.

The Sharpe ratios of the LR strategy consistently outperform the benchmark across different commodity markets but it demonstrates a significantly better performance only for copper. Overall, the model performed badly, and the strategy only meets all the requirements for copper.

Table 7 Performance of Logistic Regression

Strategy	**Commodity**	**Win rate [%]**	**Delta (LR-Benchmark) [%]**	**Profit factor**		**Sharpe ratio**	
Crude oil	LR	45.1	**2.6**	0.7		-0.1	
	Benchmark	42.5		0.8		-0.6	
Natural gas	LR	47.6	6.8	0.6		**-0.5**	
	Benchmark	40.8		0.8		-0.5	
Silver	LR	49.2	4.4	**1.1**	✓	0.0	
	Benchmark	44.8		1.0		-0.2	
Copper	LR	44.9	6.4	**1.7**	✓	**0.3**	✓
	Benchmark	38.5		0.7		-0.7	
Corn	LR	**37.7**	4.8	0.9		-0.3	
	Benchmark	32.9		0.7		-0.5	
Sugar	LR	**54.8**	**16.6**	0.6		-0.2	
	Benchmark	38.2		0.8		-0.6	

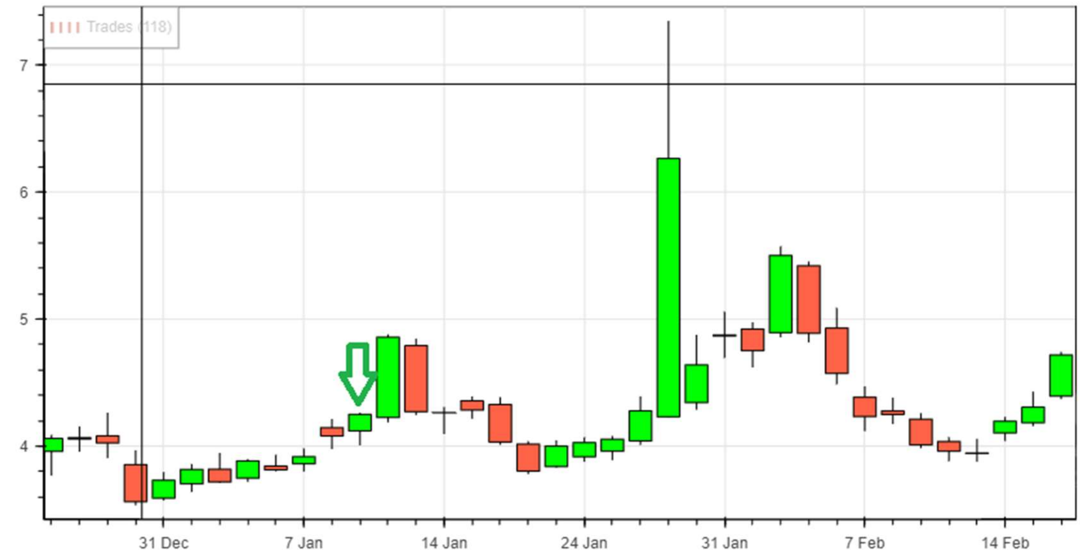
5.4 Decision Tree

The Decision Tree algorithm functions by splitting the dataset into smaller and more specific subsets. At each step, it selects the input variable that provides the greatest information gain, separating the data into more homogeneous groups. Each point where the data is split, called a node, represents a crucial decision based on the chosen feature.

As the algorithm advances, it continues to divide the data until it reaches the terminal nodes, also known as leaves. These leaves make explicit predictions about market direction, classifying future movements as either bullish (1) or bearish (-1) for the next day. This process creates a tree-like structure where each branch leads to a specific prediction.

This method not only gives traders actionable insights but also ensures transparency. Stakeholders can easily understand why a particular decision was made, as the tree structure clearly shows the factors considered at each decision point. This interpretability enhances trust and facilitates informed decision-making.

Figure 16 Example of Decision Tree Strategy – Natural Gas



5.4.1 Results

The DT model achieves higher win rates across all commodities compared to the benchmark model but by a slight margin for crude oil.

Across most commodities, the DT model also demonstrates higher profit factors compared to the benchmark model. The profit factor is also often above 1 (silver, copper, and corn) which indicates profitability for three out of six commodities. This metric is, however, only slightly above one for most of them.

The DT model yields positive Sharpe ratios in copper and corn markets, but they remain relatively low.

Overall, the analysis suggests that the Decision Tree model consistently outperforms the benchmark across all commodities in terms of win rate, profit factor, and Sharpe ratio. However, its performance remains limited.

Table 8 Trading performance of Decision Tree

Commodity	Strategy	Win rate [%]	Delta (DT-Benchmark) [%]	Profit factor		Sharpe ratio	
Crude oil	DT	**43.6**	**1.1**	0.9		-0.1	
	Benchmark	42.5		0.8		-0.6	
Natural gas	DT	52.9	**12.1**	**1.0**		-0.2	
	Benchmark	40.8		0.8		-0.5	
Silver	DT	**52.0**	7.1	**1.0**	✓	0.0	
	Benchmark	44.9		1.0		0.2	
Copper	DT	45.2	6.7	**1.1**	✓	**0.1**	
	Benchmark	38.5		0.7		-0.7	
Corn	DT	49.4	**16.5**	**1.0**	✓	**0.2**	
	Benchmark	32.9		0.7		-0.5	
Sugar	DT	47.9	9.7	**0.8**		**-0.3**	
	Benchmark	38.2		0.8		-0.6	

5.5 Random Forest

The Random Forest algorithm builds multiple Decision Trees using different samples of the dataset, generated through a process called bootstrapping. At each node within these trees, it randomly selects a subset of the available input features to determine the best split. This randomness helps to ensure that the trees are less correlated with each other, which reduces the risk of overfitting and improves the model's ability to generalize to new data.

In our research, each Decision Tree in the Random Forest examines the relationships between various inputs functioning similarly to a single Decision Tree. However, by combining the predictions from all the trees in the ensemble, the Random Forest can capture more complex and subtle patterns, better reflecting the complexities of market dynamics.

Figure 17 Example of Random Forest Strategy - Corn

5.5.1 Results

The RF model consistently achieved higher win rates across all commodities compared to the benchmark model.

In addition, the results demonstrate that RF consistently outperforms the benchmark model and exhibits a profit factor above 1 for every commodity except natural gas. This exception might be explained by the importance of weather data and other specific features that might have had some predictive power, but which were not included in the model.

The Sharpe ratios are also in line with the other metrics as they range from -0.02 to 0.40. 0.40 being the highest score obtained in our analysis across all models.

Overall, the performance of RF is robust.

Table 8 Performance of Random Forest

Commodity	**Strategy**	**Win rate [%]**	**Delta (RF-Benchmark) [%]**	**Profit factor**		**Sharpe ratio**	
Crude oil	RF	49.2	6.7	1.0	✓	**0.4**	✓
	Benchmark	42.5		0.8		-0.6	
Natural gas	RF	**45.8**	5.0	0.8		**0.0**	
	Benchmark	40.8		0.8		-0.5	
Silver	RF	**51.3**	6.4	1.2	✓	0.2	
	Benchmark	44.9		1.0		0.2	
Copper	RF	49.8	**11.3**	**1.3**	✓	0.2	
	Benchmark	38.5		0.7		-0.7	
Corn	RF	**51.9**	**19.0**	**1.5**	✓	**0.3**	✓
	Benchmark	32.9		0.7		-0.5	
Sugar	RF	47.1	8.9	1.2	✓	0.2	
	Benchmark	38.2		0.8		-0.6	

6. Analysis of trading performance

6.1 Overall performance

Table 11 Summary of trading performance

Market	Strategy	Win rate [%]	Delta (X-Benchmark) [%]	Profit factor		Sharpe ratio	
Crude oil	MA	47.8	**5.4**	**1.9**	✓	**0.2**	
	MFI-RSI	**71.0**	**28.5**	**4.2**	✓	0.0	
	MSV	50.0	7.5	0.9		-0.1	
	LR	45.1	2.6	0.7		-0.1	
	DT	**43.6**	1.1	0.9		-0.1	
	RF	49.2	6.7	1.1	✓	**0.4**	✓
	Buy&hold	-	-	0.8		-0.4	
	Benchmark	42.5	-	0.8		-0.6	
Natural gas	MA	49.5	8.7	0.9		-0.0	
	MFI-RSI	**60.9**	20.1	0.7		**-0.5**	
	MSV	50.0	**9.2**	**1.0**	✓	0.1	
	LR	47.6	6.8	0.6		**-0.5**	
	DT	52.9	12.1	1.0	✓	0.2	
	RF	**45.8**	5.0	0.8		-0.0	
	Buy&hold	-	-	1.1	✓	0.3	✓
	Benchmark	40.8	-	0.8		-0.5	
Silver	MA	48.9	4.0	1.1	✓	0.1	
	MFI-RSI	**44.2**	**-0.7**	**0.6**		**-0.5**	
	MSV	**54.9**	**10.0**	**1.5**	✓	**0.8**	✓
	LR	49.2	4.4	1.1	✓	0.0	
	DT	52.0	7.1	1.0	✓	0.0	
	RF	51.3	**6.4**	**1.2**	✓	**0.2**	
	Buy&hold	-	-	1.1		0.4	✓
	Benchmark	44.9	-	1.0		0.2	
Copper	MA	41.9	1.3	0.3		**-0.9**	
	MFI-RSI	**67.3**	26.6	1.2	✓	0.1	
	MSV	49.0	8.4	0.8		**-0.4**	
	LR	44.9	**6.4**	**1.7**	✓	**0.3**	✓
	DT	45.2	6.7	1.1	✓	0.1	
	RF	49.8	**11.3**	**1.3**	✓	0.2	
	Buy&hold	-	-	1.1		0.5	✓
	Benchmark	38.5	-	0.7		-0.7	
Corn	MA	47.2	14.3	1.0	✓	-0.2	
	MFI-RSI	**65.0**	32.1	0.9		**-0.4**	
	MSV	52.0	19.1	0.9		-0.1	
	LR	**38.0**	4.8	0.9		-0.3	
	DT	49.4	16.5	1.0	✓	0.2	
	RF	51.9	**19.0**	**1.5**	✓	**0.3**	✓
	Buy&hold	-		1.1	✓	0.3	✓
	Benchmark	32.9		0.7		-0.5	
Sugar	MA	50.6	**12.4**	**1.5**	✓	**0.3**	✓
	MFI-RSI	**77.1**	**38.9**	**1.5**	✓	0.2	
	MSV	45.1	6.9	0.9		-0.2	
	LR	54.8	**16.6**	**0.6**		-0.2	
	DT	47.9	9.7	0.8		**-0.3**	
	RF	47.1	8.9	1.2	✓	0.2	
	Buy&hold	-	-	1.1	✓	0.6	✓
	Benchmark	38.2	-	0.8		-0.6	

The analysis we conducted and that is shown in table 11 indicates that Random Forest (RF) outperformed technical analysis strategies examined for most of the commodities. In addition, within the machine learning group, Random Forest provided the best results while Logistic Regression and Decision Trees lagged. There are however exceptions, in certain commodities we see the outperformance of technical analysis, but Random Forest (ML) provided the most consistent results with profit factors often above 1 across most commodities.

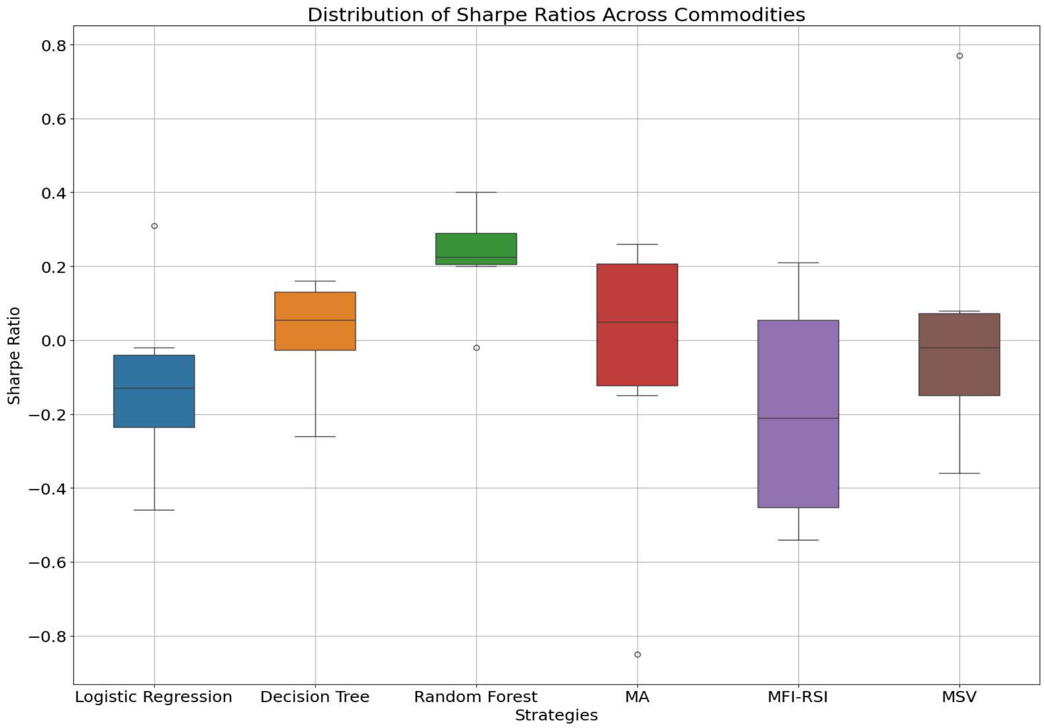
The ML strategies based on Random Forest which is a model that is able to capture more complex, non-linear relationships led to a better performance overall.

Among technical analysis strategies, the moving average crossover strategy showed robust performance in some markets, but RF strategies were able to further optimize and enhance the trading signals, leading to better risk-adjusted returns and profit factors.

The analysis also reveals significant variations in the performance of both the ML and technical analysis strategies across different commodity markets.

Some markets, such as silver or sugar, showed higher risk-adjusted returns, while others, such as energy commodities, exhibited lower performance overall. This suggests the effectiveness of the trading strategies is heavily dependent on the specific market environment and characteristics.

Figure 18 Distribution of Sharpe Ratios across commodities



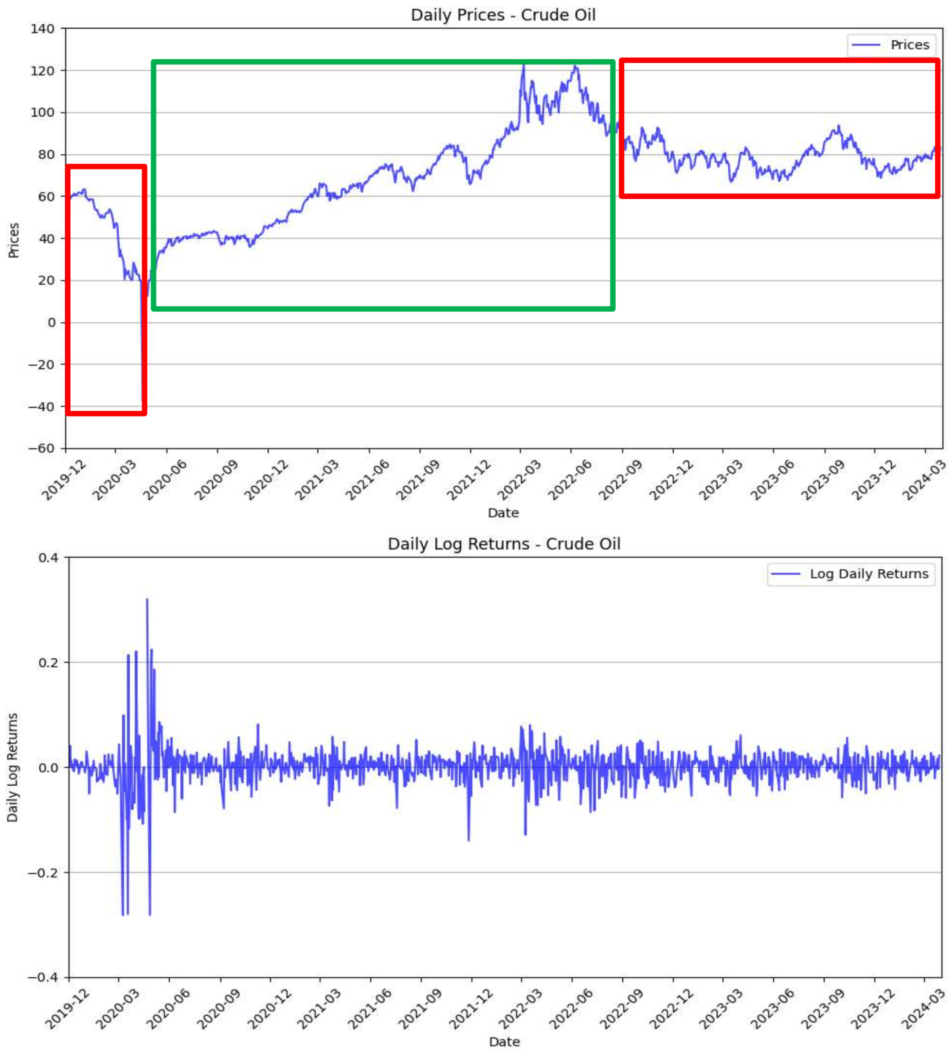
6.2 Evolution of performance

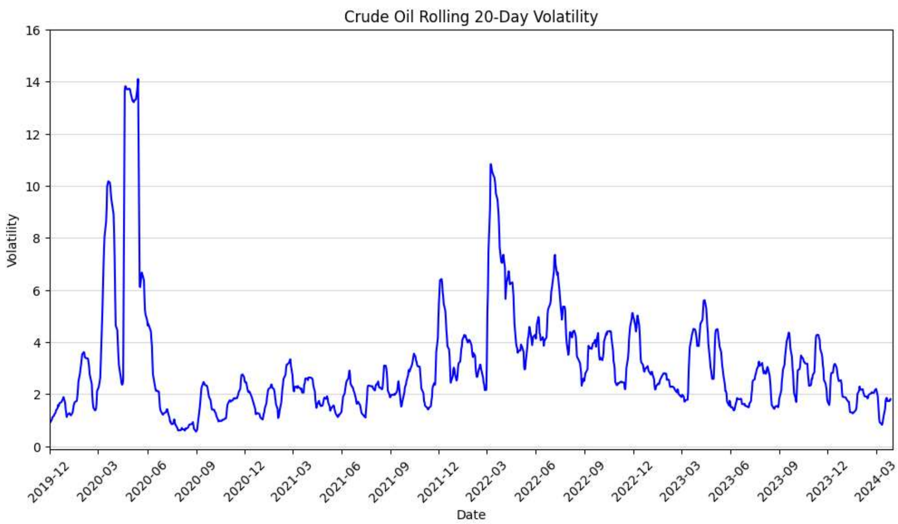
The evolution of performance is shown in the following figures. The goal is to analyse and interpret how different periods and market phases impact the performance of our strategies. Bull markets are indicated in **green** and bear markets in **red.**

6.2.1 Crude Oil

For this first commodity, we will show the prices, daily returns, and volatility over the timeframe that we are studying. These plots help us understand the impact of volatility and market phases on our models' performance. To maintain clarity, we will present these charts exclusively for crude oil. We note that there is a significant increase in volatility in March-April 2020 and again between February and April 2022. We see the market in a bull phase from after the drop in April 2020 to August 2022 after reaching ~120$/bb. Prices then decreased gradually (bear market).

Figure 19 Prices, log returns and volatility of Crude Oil





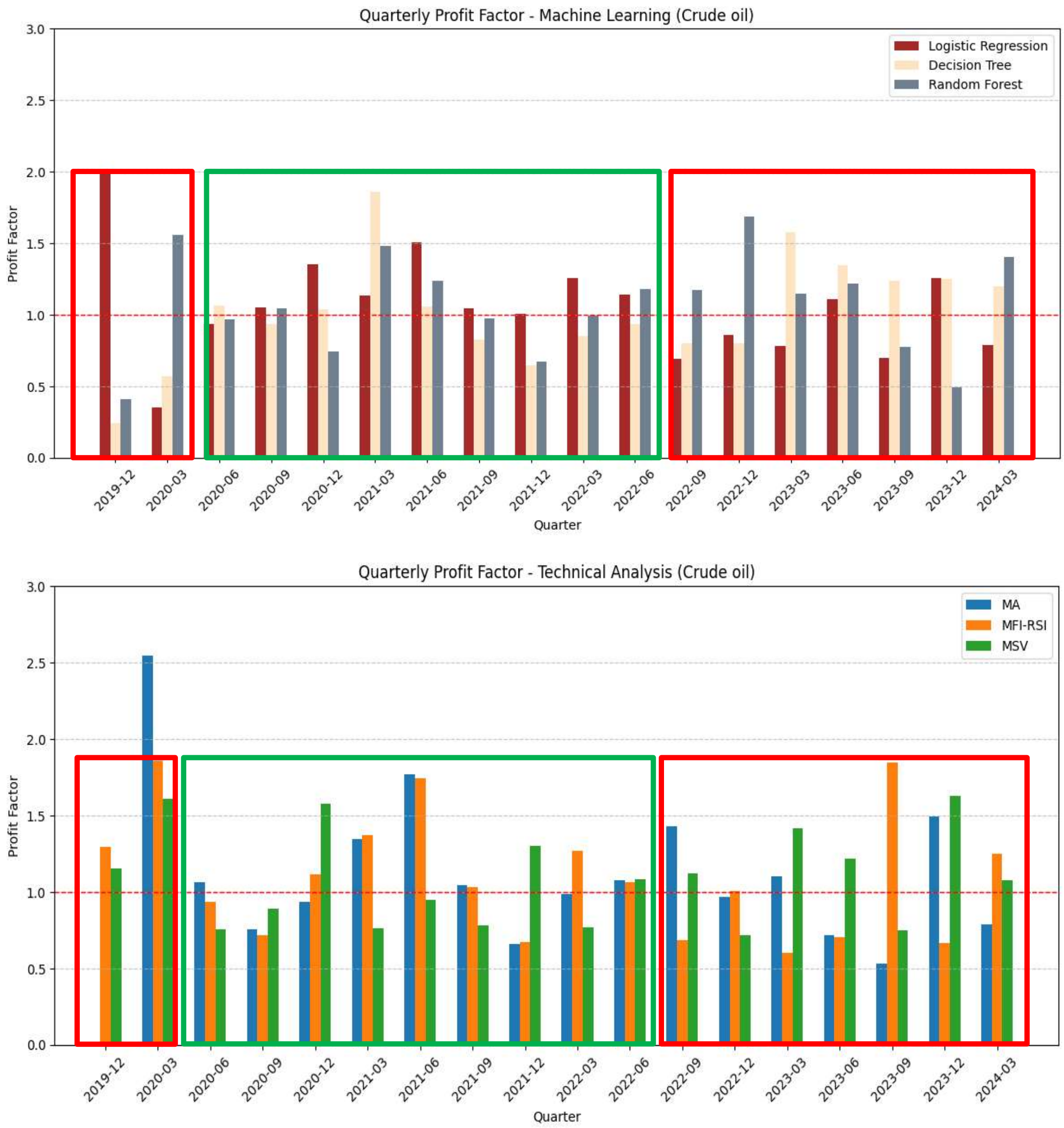
The quarterly profit factors for machine learning strategies exhibit significant variability in response to major market events. During the initial shock of the COVID-19 pandemic in 1Q20, all ML strategies experienced a drop in profit factors, reflecting the extreme market volatility that characterized this period although RF was able to deliver profit factor above 1.5. Indeed, among the ML strategies, Random Forest demonstrated a notable ability to adapt and pick up better during periods of high volatility, as evidenced by its performance in the beginning of 2020 and throughout 2021. This robustness was also apparent in 1Q-2Q22 amid the onset of the Russia-Ukraine conflict, where Random Forest achieved moderate improvements in profit factors. By 3Q-4Q22, Random Forest reached even higher profit factors, reflecting its effectiveness in leveraging the continued volatility and high oil prices driven by geopolitical tensions, outperforming Logistic Regression and Decision Tree in several key quarters.

For crude oil, ML strategies except Random Forest seem to have some difficulties generating consistent profits in periods of high volatility.

Overall, Logistic Regression and Decision Tree seems to get negatively affected by “big events” that drive up volatility while Random Forest seems to maintain its performance.

Technical indicator-based strategies displayed distinct patterns in response to market events with relatively high variability.

Figure 20 Crude oil quarterly profit factor



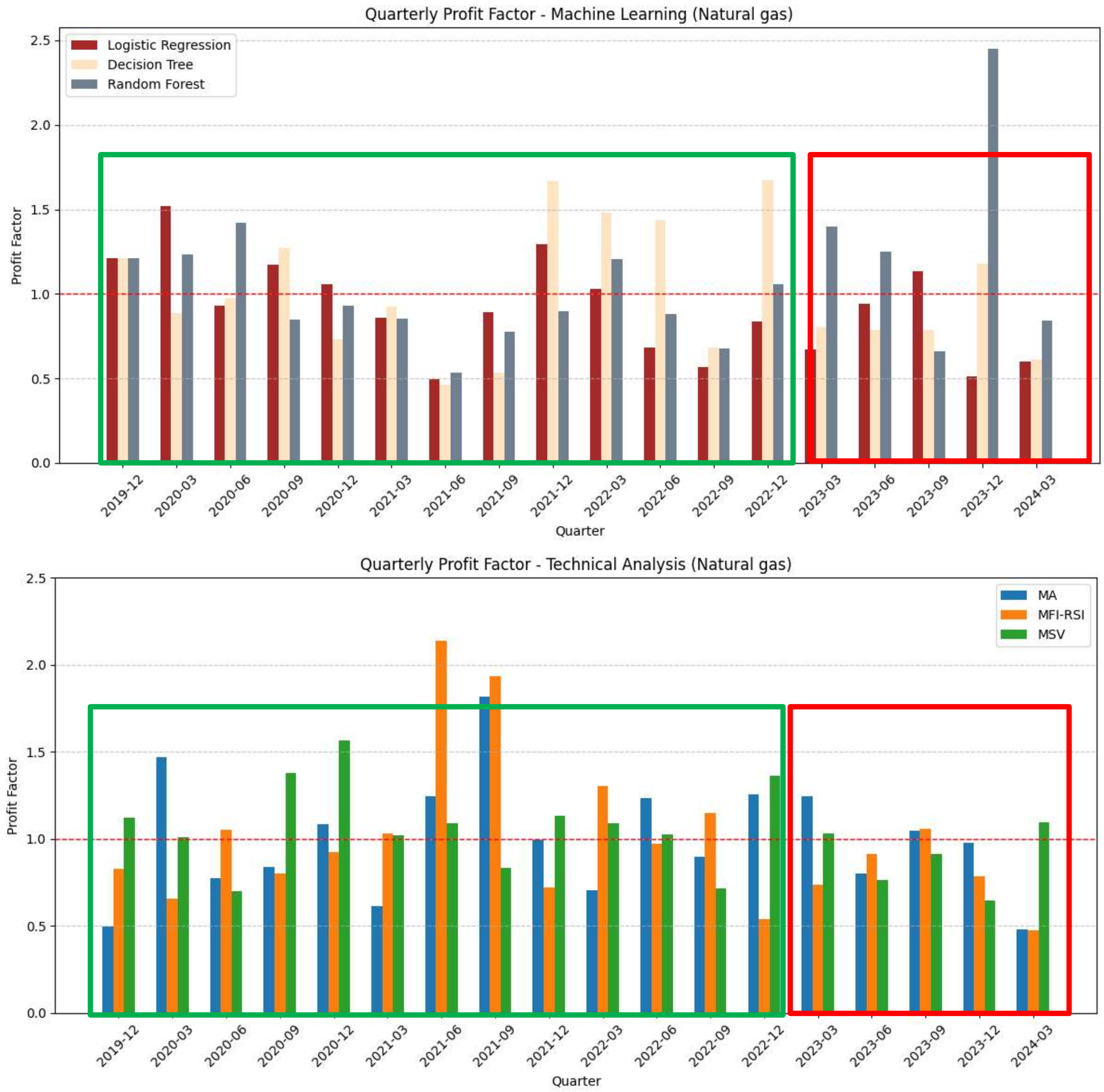
6.2.2 Natural gas

Volatility in natural gas futures was relatively steady until 2Q-3Q21 when prices more than tripled in a short timeframe. This increase was mainly due to the economic recovery and increased demand for natural gas. This can be seen in our charts, machine learning algorithms struggle to identify sudden price changes such as the increase in positive volatility that happened at this time, profit factors therefore went from around 1 to 0.5.

The situation continued in 2022 and volatility remained high when prices skyrocketed due to the Russia-Ukraine war. Decision Tree seems to better capture the market situation and deliver better performance, but we have to recall that its overall performance only reached 0.97 (profit factor).

The results for technical analysis vary and we do not see any trend except for the MFI-RSI strategy that delivered good results in 2Q21 and 3Q21.

Figure 21 Natural gas quarterly profit factor



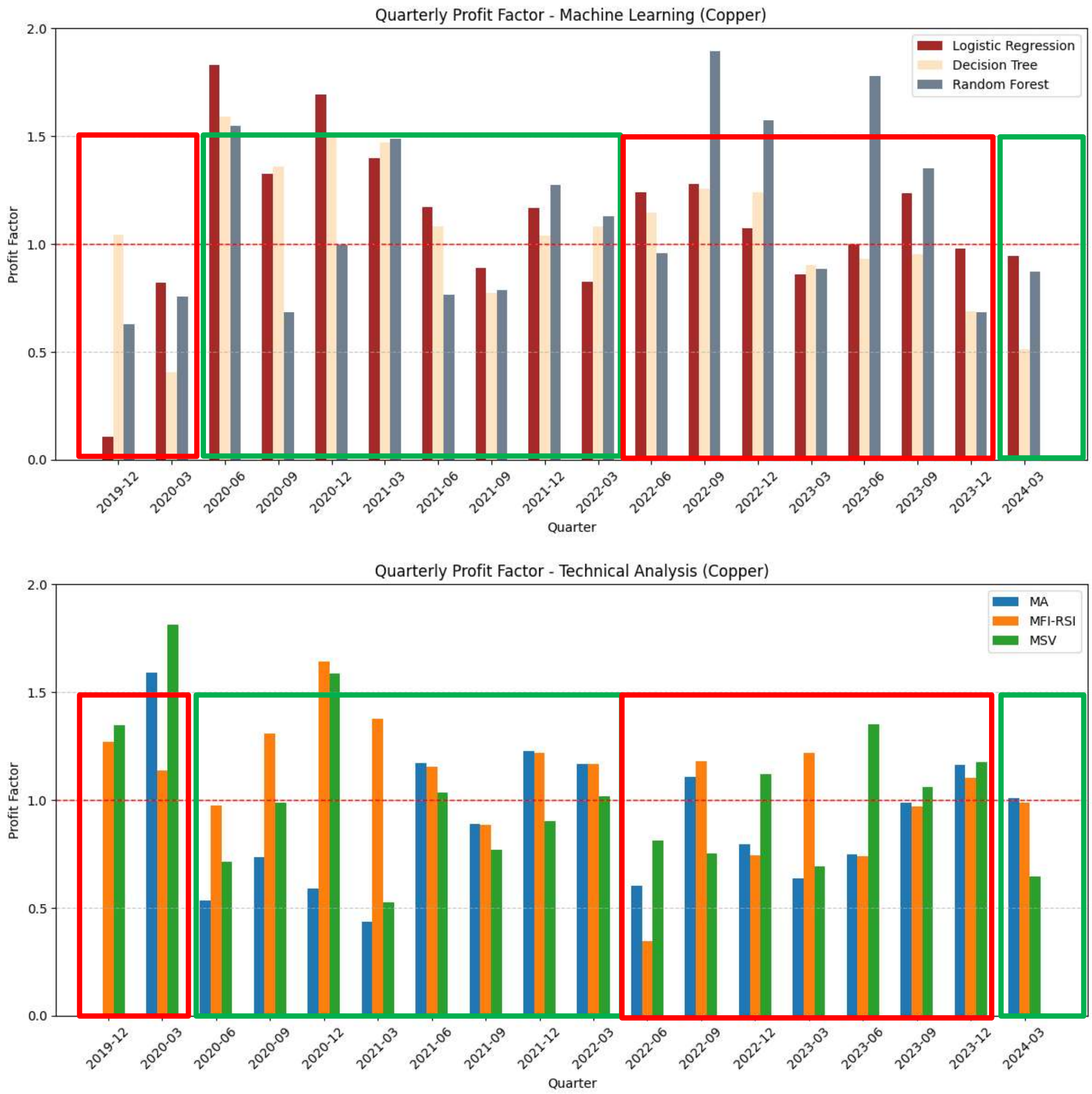
6.2.3 Copper

From 2019 to 2024, copper prices experienced many ups and downs, initially dipping due to the COVID-19 pandemic's economic slowdown but quickly rebounding due to supply chain disruptions and increased demand from stimulus measures and infrastructure projects. While fears of a global recession in 2022 caused a temporary pullback, prices surged again afterwards as China reopened its economy and the demand for copper in the green energy transition intensified.

Machine learning strategies performed relatively well in ore markets. This can be explained by the low volatility of returns of this market in the dataset used, the lowest standard deviation of returns was found in the copper market. However, we can note the significant decrease in performance for machine learning strategies during the first Covid wave which did not happen for technical analysis although the decrease was not as brutal and sudden as for other commodities. After the first covid wave, they performed relatively well overall. We can note that volatility was at its highest during 3Q22 when prices plummeted, and Random Forest delivered a profit factor superior to 1.5 during this period.

Technical analysis suffered however when the volatility was steadier, and performance increased slowly after 1Q20 (except for the MFI-RSI strategy).

Figure 22 Copper quarterly profit factor



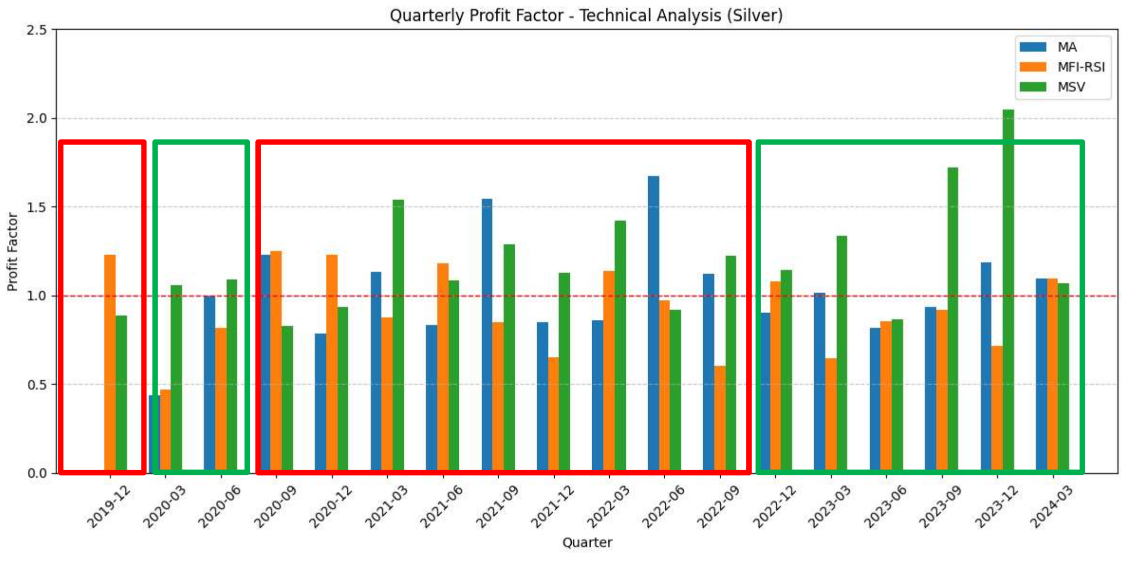
6.2.4 Silver

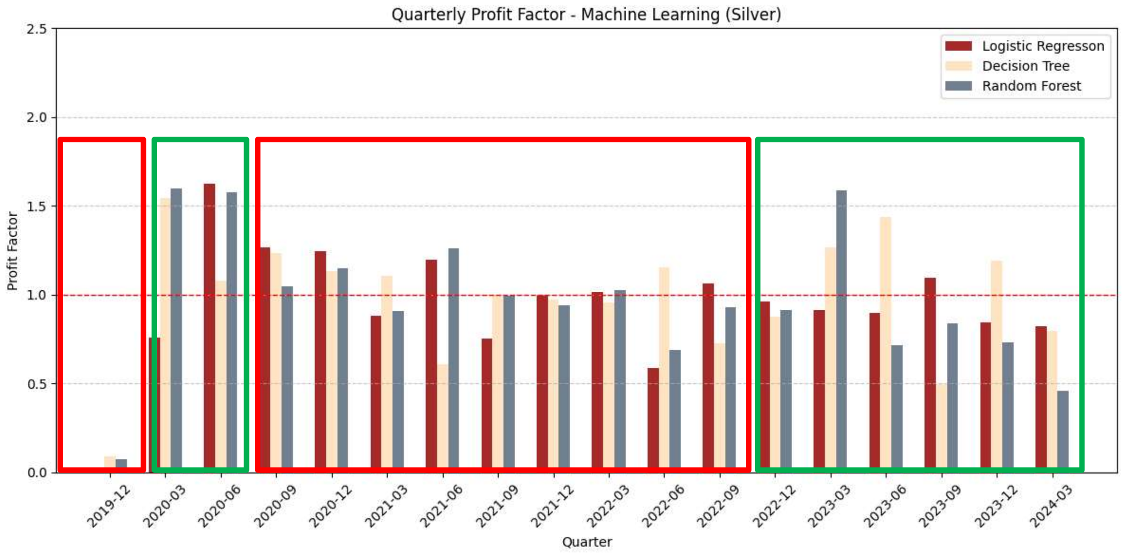
Between 2019 and 2024, silver prices experienced a turbulent journey, initially dipping due to the COVID-19 pandemic's economic fallout, then soaring to multi-year highs driven by safe-haven demand. Inflationary pressures and geopolitical risks further boosted prices in 2022, but fears of a recession and a stronger U.S. dollar led to a subsequent decline. However, a recovery began in 2024, fueled by renewed economic optimism and the growing demand for silver in green technologies.

There is an outlier for Logistic Regression during 4Q19 which led the model to reach 6 in profit factor. We decided not to show it for readability reasons. This is explained by several trades that were positive and no losses.

Random Forest often slightly surpasses other models especially in 1Q20 and 1Q23. Otherwise, performance for ML is often steady at around 1 of profit factors. The highest volatility was experienced in 2-3Q20 and 3-4Q22. This time, Decision Trees and Random Forest were able to successfully take advantage of these sudden increases in volatility which led to a profit factor above 1. On the other hand, technical strategies had varying levels of performance.

Figure 23 Silver quarterly profit factor





6.2.5 Corn

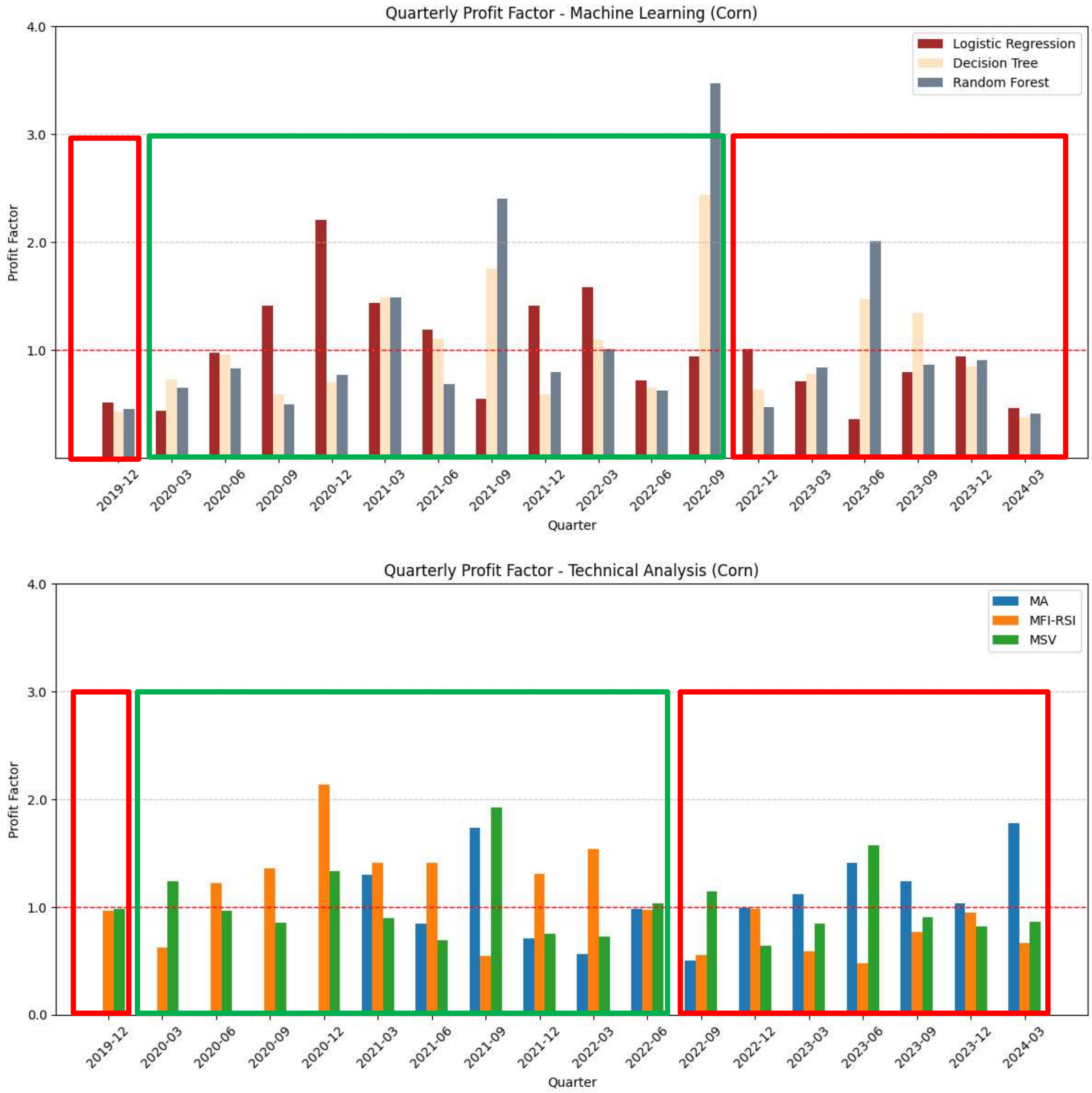
From 2019 to 2024, corn prices experienced significant fluctuations due to a series of major events. In 2020, the COVID-19 pandemic initially caused prices to slightly drop as demand from ethanol production declined. However, this was followed by a sharp increase due to supply chain disruptions, adverse weather conditions, and increased demand from China. The Russia-Ukraine war in 2022 further exacerbated the situation, disrupting global grain supplies and driving prices to record highs. While prices have moderated somewhat in 2023 and 2024, they remain elevated due to ongoing geopolitical tensions, supply chain issues, and growing concerns about climate change impacting crop yields.

For machine learning strategies, the chart shows considerable variability across different quarters. Random Forest sometimes outperforms Logistic Regression and Decision Tree, with notable peaks in quarters such as 3Q21, 3Q22, and 2Q23, where it achieves profit factors well above 1. The highest profit factor for Random Forest in 3Q22 suggests a period of strong market conditions (high positive volatility). However, there are also quarters, such as 4Q19 and 1Q20, where all ML models underperform.

In contrast, technical analysis strategies exhibit a different pattern.

The Moving Average (MA) strategy often shows higher consistency, with several quarters achieving profit factors above 1, such as in 3Q21 and 1Q24. The MFI-RSI strategy, while sometimes matching or exceeding MA's performance, is more volatile, with significant peaks and troughs, such as a high in 4Q20 followed by dips in subsequent quarters. The Moving Standard Deviation (MSV) strategy generally lags behind the other two, rarely achieving a profit factor above 1 and showing more consistent underperformance.

Figure 24 Corn quarterly profit factor



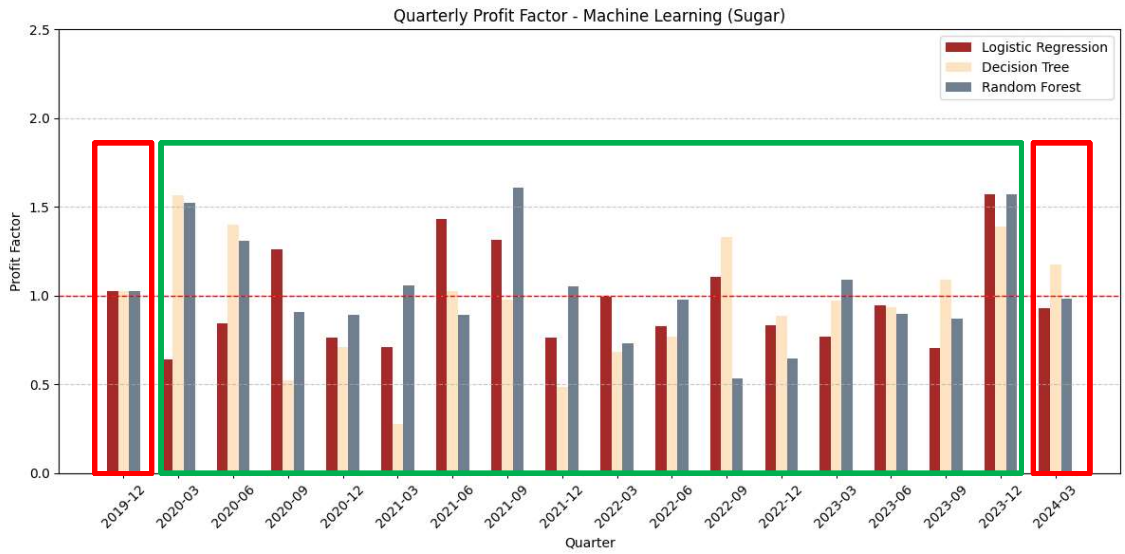
6.2.6 Sugar

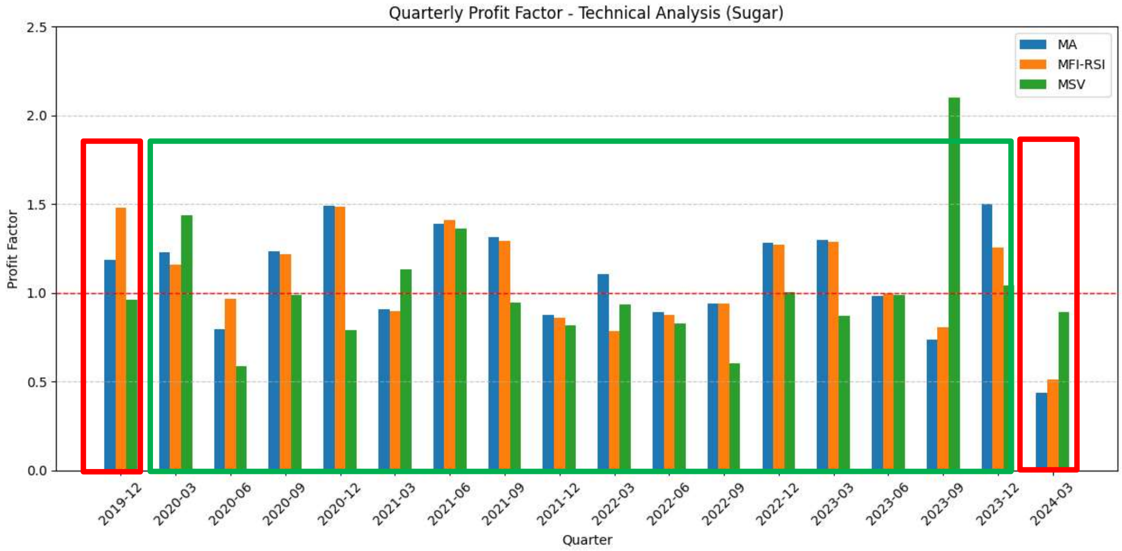
From 2019 to 2024, sugar prices experienced significant fluctuations due to a variety of factors. Initially, the COVID-19 pandemic led to a dip in prices. However, a combination of factors, including supply chain disruptions, adverse weather conditions in major producing regions like Brazil, and increased demand from China, drove prices to multi-year highs in 2023.

For machine learning strategies, the Random Forest model outperforms Logistic Regression and Decision Tree, with several quarters showing profit factors above 1. The Logistic Regression and Decision Tree models show more fluctuations.

In contrast, the technical analysis strategies display more variability across the quarters. The Moving Average (MA) strategy frequently shows profit factors around or slightly above 1. The MFI-RSI strategy is more volatile, with significant peaks in some quarters but also several low points. In addition, MSV also shows many quarters with profit factors below one, indicating inconsistent performance.

Figure 25 Sugar quarterly profit factor





Quarterly Profit Factor - Technical Analysis (Sugar)
MA
MFI-RSI
MSV
Profit Factor
Quarter
2019-12
2020-03
2020-06
2020-09
2020-12
2021-03
2021-06
2021-09
2021-12
2022-03
2022-06
2022-09
2022-12
2023-03
2023-06
2023-09
2023-12
2024-03

6.3 Robustness Analysis– Random Forest

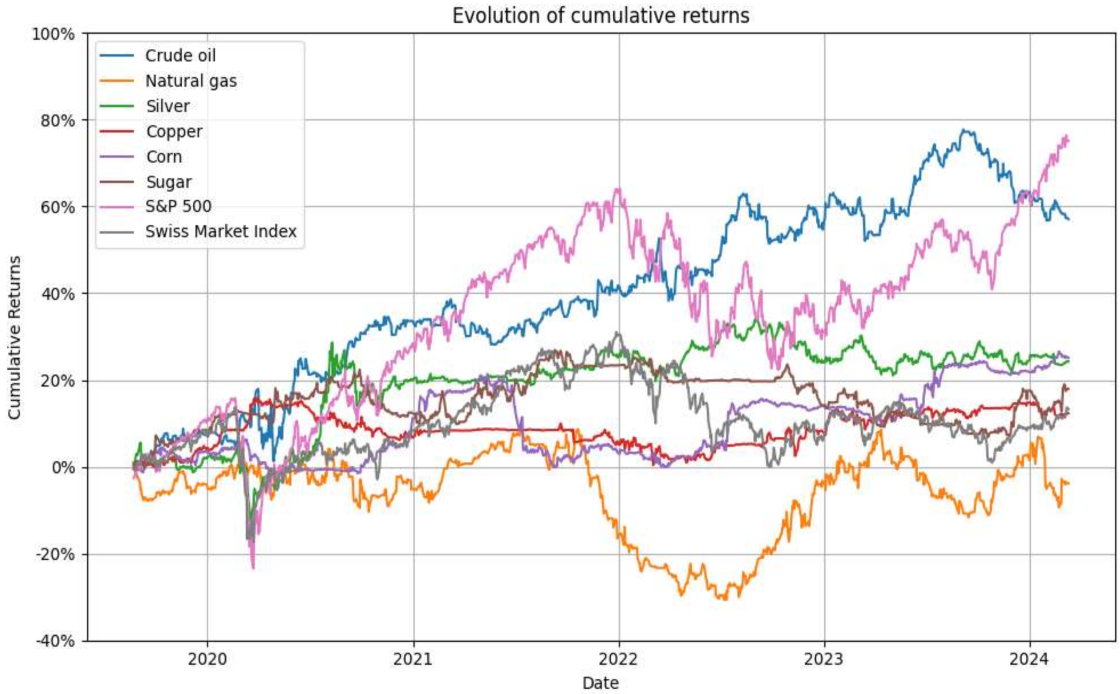
6.3.1 Comparative analysis

We analyzed the performance of Random Forest with a focus this time on comparing its returns to those of renowned equity investments which are the SP500 and the SMI (Swiss Market Index) with a buy&hold strategy.
It is evident that the Random Forest (RF) strategy exhibited great results in terms of returns. While some commodities, such as corn and silver, also demonstrated robust returns, others, such as natural gas, incurred losses. We note also the quite high returns in crude oil, almost surpassing the SP500 buy and hold strategy at the end of the period. Sugar, corn, silver, and crude oil all outperformed the SMI, but no strategy was able to outperform the SP500.

Table 11 Comparative analysis of trading performance – Random Forest

Market	Strategy	Win rate [%]	Profit factor	Sharpe ratio	Returns [%]
Crude oil	RF	**52.0**	1.0	**0.4**	**57.1**
Natural gas	RF	45.8	**0.8**	0.0	**-3.9**
Silver	RF	51.3	1.2	0.2	24.3
Copper	RF	49.8	1.3	0.2	12.3
Corn	RF	51.9	**1.5**	0.3	25.1
Sugar	RF	47.1	1.2	0.2	17.9
SP500	Buy & Hold	-	1.1	**0.7**	**75.1**
SMI	Buy & Hold	-	1.1	0.3	13.3

Figure 26 Cumulative returns of Random Forest strategy



6.3.2 Sensitivity Analysis

We will run our best performing model (Random Forest) again in a scenario where the input variables with the lowest predictive capability (i.e. that have the lowest relative importance for each commodity) are removed to understand how sensitive the model is to a change to its input variables. We are therefore removing some of the "noisiest" inputs to see if the models remain robust. The relative importance for each commodity and model can be found in the appendix section. We can see from table 12 that the overall returns remain stable for most commodities, except for natural gas which already had negative returns without any feature removal. Copper also sees a small decrease in performance when we removed two features. When we removed four features in total, the performance started to slightly decrease overall which also happened as we removed six features in total. The commodities that seemed most affected by this removal were crude oil which saw the largest decrease but remained the commodity with the highest cumulative return, and also copper but it was able to stay in positive territory. Finally, natural gas performance continued to plummet as we removed more features.

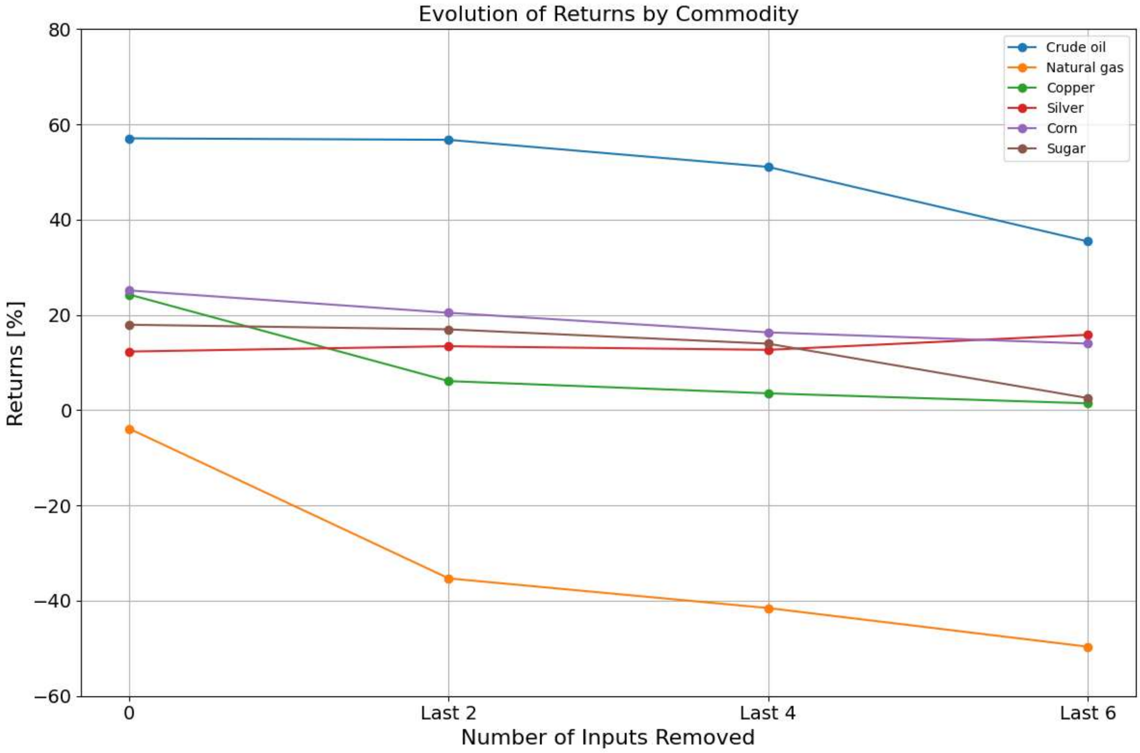
Table 12 Sensitivity analysis of Random Forest

Last four features removed				
Commodity	**Win rate**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	48.9	1.1	0.1	**51.1**
Natural gas	44.9	0.7	**-0.3**	**-41.5**
Silver	50.2	1.1	0.2	12.7
Copper	49.1	1.1	0.1	3.5
Corn	48.5	**1.3**	0.1	16.3
Sugar	49.9	1.2	0.1	14.0

Last six features removed				
Commodity	**Win rate**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	48.0	1.1	0.1	**35.4**
Natural gas	45.8	0.7	**-0.3**	**-49.7**
Silver	51.9	1.1	0.2	15.8
Copper	50.2	1.0	**0.0**	1.4
Corn	54.5	**1.2**	0.1	14.0
Sugar	53.0	1.1	0.1	2.5

Last two features removed				
Commodity	**Win rate**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	50.0	1.1	0.1	**56.8**
Natural gas	45.1	0.7	**-0.2**	**-35.3**
Silver	51.2	**1.2**	0.2	13.4
Copper	**49.3**	**1.1**	**0.2**	**6.1**
Corn	50.3	1.1	**0.3**	20.5
Sugar	47.0	1.2	0.2	17.0

Figure 27 Evolution of cumulative returns when removing features



6.3.3 Performance cross-validation

We decided to ensure the robustness of our results by running our best performing model again in different periods. We chose to show the results of separate years for better comparison. This step is crucial for validating the model.

Our results can be found in table 13, we see that the Random Forest (RF) strategy shows promising results across multiple commodities and returns consistent with those found previously.

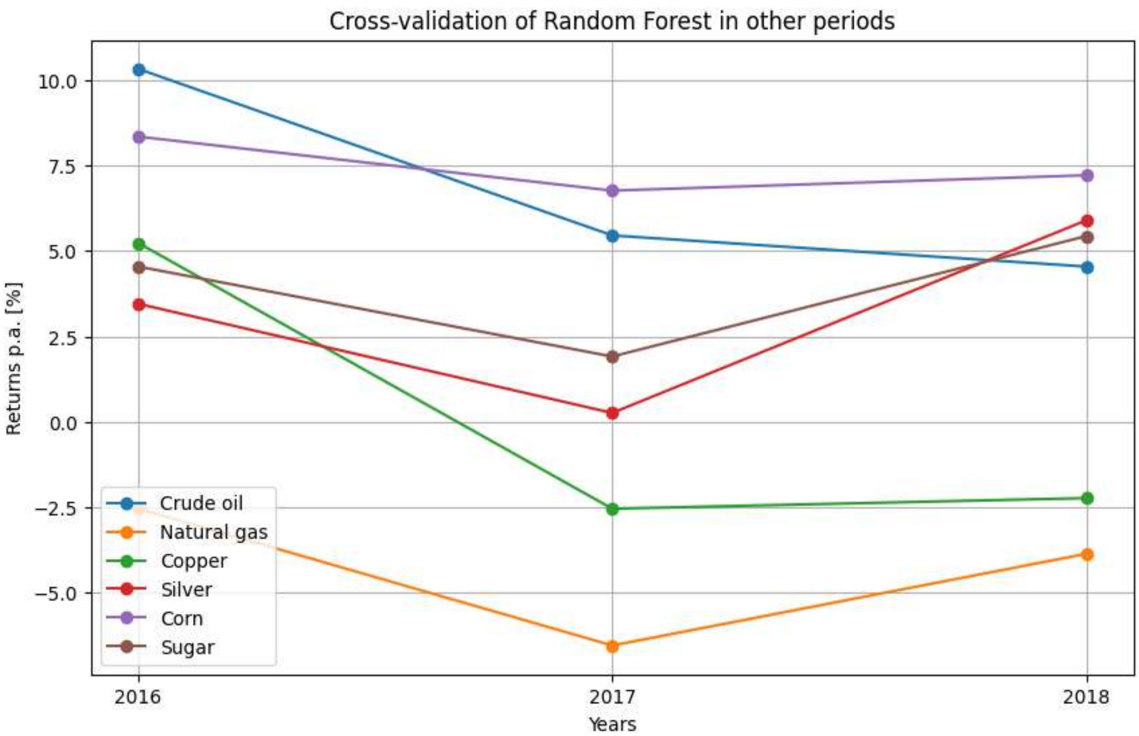
Table 13 Trading performance per year for cross-validation – Random Forest

Year 2016					
Commodity	**Strategy**	**Win rate [%]**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	RF	51.0	**1.3**	**0.2**	**10.3**
Natural gas	RF	46.2	**0.8**	**-0.1**	**-2.5**
Silver	RF	49.5	1.1	0.1	3.5
Copper	RF	50.2	1.1	0.1	5.2
Corn	RF	51.2	1.2	0.1	8.3
Sugar	RF	54.6	1.1	0.2	4.5

Year 2017					
Commodity	**Strategy**	**Win rate [%]**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	RF	50.0	**1.3**	**0.4**	5.5
Natural gas	RF	48.4	**0.8**	**-0.1**	**-6.5**
Silver	RF	49.5	1.1	0.1	0.3
Copper	RF	54.5	1.0	0.1	-2.5
Corn	RF	49.9	1.3	0.1	**6.8**
Sugar	RF	52.0	1.0	0.2	1.9

Year 2018					
Commodity	**Strategy**	**Win rate [%]**	**Profit factor**	**Sharpe ratio**	**Returns [%]**
Crude oil	RF	52.1	1.0	**0.2**	4.5
Natural gas	RF	51.0	0.9	**-0.2**	**-3.9**
Silver	RF	49.0	1.1	0.2	5.9
Copper	RF	52.5	1.0	0.1	-2.2
Corn	RF	49.5	**1.2**	0.0	**7.2**
Sugar	RF	50.7	1.1	0.1	5.4

Figure 28 Cross-validation of Random Forest in other periods



6.3.4 Portfolio Comparison

We will finally compare the investment performance that could be achieved by the strategies assessed in both categories: technical analysis and machine learning. Each commodity is equally weighted in these fictious portfolios and assessed for each model. We will also compare them to the buy&hold strategies. As we can see, the performance of the best strategy remains below the buy&hold strategy for the SP500 but is above the SMI. A portfolio composed of equally weighted commodities will hence deliver superior performance of the SMI only in the case of the Random Forest strategy. Other models delivered worse performance.

Table 14 Portfolio Analysis – All strategies

Strategy	Win rate [%]	Profit factor	Sharpe ratio	Returns [%]
RF	**53.1**	**1.1**	**0.1**	**16.3**
DT	46.6	**0.9**	**-0.3**	-19.3
LR	43.5	0.8	-0.2	-28.1
MA	49.2	0.8	-0.9	-11.2
MSV	48.8	**0.7**	-0.4	**-54.7**
MFI-RSI	**52.1**	0.8	-0.5	-42.5
SP500	-	1.1	**0.7**	**75.1**
SMI	-	1.1	0.3	13.3

7. Experts' opinion

We gathered the views of experts from the industry to better grasp the complexities of building trading strategies that are profitable in the long-term and to understand the pros and cons of building a machine learning model for companies active in commodity trading.

7.1 Data and market requirements

Respondent A acknowledged that the importance of data is crucial. He emphasized the fact that many models use different granularity of data (e.g., monthly, daily, hourly price) and input variables that can range from technical indicators and economic indicators used in this thesis to more advanced features such as market sentiment data (news, social media, etc.), market depth data (bid-ask spread, order book dynamics, liquidity levels, etc.).

Respondent C emphasized the importance of avoiding "over-validation" which he defines as the fact of having several indicators that measure the same. For example, several momentum indicators. Having data that is published on a quarterly or monthly basis is not a problem as long as the trader digs deeper and understands the consequences of each publication on the market. For example, understanding how the market responds to inflation publications and what's the impact of that. It is crucial not to assume the outcomes of certain events. In addition, respondent C believes that another component is how long the signal is valid for. For example, inflation data may be "absorbed" in a few days so this needs to be embedded in the model somehow.

Some features (economic vs. technical) also work best in some markets and less in others, which is a factor that is also influenced by the type of market participants in this specific market. Other factors that can influence the type of input that can be the most important can be how liquid the product is, if the market typically has momentum or not, etc. She also explained that if there is a solid fundamental reason for prices to increase or decrease, machine learning can accentuate the move, as they work on understanding trends and price movements.

Respondent C emphasized the importance of robust risk management systems that can use indicators such as the ATR and the importance of feature engineering and robust pre-analysis of inputs before using them in a model.

7.2 Main factors affecting machine learning profitability

Respondent B acknowledges that volatility is the main factor that could negatively affect machine learning profitability and more specifically “big market events”, because these models can understand momentum/price movements, but they cannot predict a big event such as the Ukraine war unless they account for this factor in the model. As indicated before, respondent A also explains that market news can be a variable used in some models. We believe this can enhance the probability of the model to detect a sudden change in price momentum, but this remains an important threat to the profitability of these models. Respondent C however does not think that volatility is necessarily a problem but the fact that markets are unstable. He believed the biggest threat is quick regime switches as it is difficult to train models for that. He also adds that hybrids model can provide better results.

7.3 Advantages of machine learning

Respondent A acknowledges that ML can capture complex patterns in financial data and can deliver superior performance if used appropriately. Respondent B even emphasizes that they are essential these days because that is what’s driving prices or it’s the only way to get an edge. She gives the example of electricity where the data is high frequency, and there is a lot of data (e.g. weather data, hourly pricing, etc.) and therefore the market may be more suitable for the use of machine learning models. Machine learning can also be a good method to compare and assess how markets react to a change of supply and/or demand and therefore the trader can better understand any possible future market reaction to any change in these variables. Respondent B also explained that without these kinds of tools in some markets it can be very difficult to make money.

Some markets are also highly controlled (in position sizing for instance), in order to protect the market, which increases how difficult it is to implement this type of strategy. Respondent C believes that machine learning can have an edge on technical indicator-based strategies as they suffer from slow response as they measure a particular state of the market. They aren’t predictive at all. In that context, respondent C thinks machine learning has a better chance of “fitting” the market relative to its inputs but at the risk of overfitting.

7.4 Disadvantages of machine learning

Respondent A acknowledges that the main problems lie in the implementation that can be costly for companies, as he believes the disadvantages are the complexity of building this type of strategy, as they require a deep knowledge in programming, machine learning, statistics, and they can be time-consuming and require significant resources.

On the other hand, ML models can be in some cases computationally expensive as they might require a vast amount of data and testing and some models such as neural networks may require time and power to train. This was not the case for our models.

Respondent A and B explained that the interpretability of the models can also be a concern as it can be challenging to grasp the underlying rationale behind each prediction. Respondent B explains that she doesn’t use machine learning because of lack of time and technical skills. For her, it’s more about resources and expertise.

Respondent B cited overfitting as a disadvantage given that this may lead to a significant decrease in performance in live trading if the model captures noise rather than true underlying patterns. He believes cross-validation methods are key when validating new models. Respondent C mentions the lack of quality data as being an important problem. Some particularities of the market are difficult to model such as regulations changes, countries going to war etc. The difficulty is to make the model account for specific events.

7.5 Market phases

Respondent B makes a distinction between trending versus consolidation markets. Trending involves a market where prices go up or down for consecutive days. She believes it is much more difficult to build a model that works in consolidation markets as it is more difficult for the algorithm to distinguish price patterns. Respondent A believes good machine learning models make money on both bull and bear market phases. Respondent C acknowledges that the main difficulty lies in the ability to find a model that will work across different states of the market, but a model should make money in bull and bear markets.

8. Limitations

While this study contributes valuable insights into the comparative performance of machine learning and technical analysis strategies in commodity trading, several limitations should be acknowledged.

Our analysis relies on historical commodity futures prices for a specific period. While this dataset provides a substantial historical context for our analysis, it may not capture all relevant market dynamics, especially in rapidly evolving market environments or during unprecedented events such as extreme volatility or market disruptions.

Furthermore, the performance of trading strategies is inherently influenced by a large number of variables. While our analysis encompasses various commodities and market phases, the generalizability of our findings to different market conditions or other commodities remains a subject of further exploration.

Our study employs a rolling cross-validation approach to assess the performance of trading strategies, with an in-sample dataset used for model training and an out-of-sample dataset for evaluation. While this methodology provides valuable insights into the robustness of our models, alternative validation techniques such as walk-forward validation could offer additional perspectives on model performance and stability. Also, while we tried to replicate a similar environment to live trading, certain differences remain. Slippage costs were not accounted for although we considered transaction costs. Also, illiquid markets may exhibit wider bid-ask spreads and greater price volatility, increasing the risk of adverse price movements and execution challenges.

The selection and calibration of features plays a crucial role in determining the efficacy of trading strategies. While we have chosen a set of commonly used variables that would provide a strategy that yields profits across several futures markets, we did not choose any market-specific variables. An idea for improvement would be to carry out a more in-depth analysis of the input variables needed (correlation analysis for example) and include more market-specific inputs with different granularities and with a focus on leading indicators. The usage of hybrid machine learning models and/or neural networks should also be explored.

9. Recommendations

Firms interested in using machine learning models for commodity trading could benefit from the following recommendations:

Data Strategy

Diversification and Quality: Prioritize clean, relevant, and diverse data. Incorporate fundamental data such supply/demand figures, macroeconomic indicators, weather data, and sentiment analysis. Combine fundamental and technical analysis to capture different patterns. Conduct an advanced analysis of inputs to ensure their predictability and ensure the presence of leading rather than lagging indicators.

Feature Engineering: This might involve transforming data, creating new indicators, or aggregating information across different timeframes (granularity).

Market Specificity: Tailor data collection and feature engineering to each specific commodity market, considering its unique dynamics.

Usage of Appropriate libraries: when building the models, we used several libraries to build the strategies and found that the easiest to use was named “Backtesting”.

Model Development and Selection

Knowledge of models: When building the models, we ensured previously to have good knowledge of their usage and how they work. This is crucial before implementation.

Hybrid models: hybrid models can have an edge over “unique” models, as explained by respondent C.

Adaptability: Develop models that can adapt to different markets. This may involve using different indicators, adjusting model parameters, or switching between different models altogether. Ensuring that the inputs continue to have good predictability, especially after big market events. Regular recalibration is key.

Risk Management and Trading Strategies:

Integrated Risk Management: Develop models that not only predict price movements but also incorporate risk management principles, such as position sizing based on volatility (e.g., ATR).

Implementation and Infrastructure:

Progressive implementation: after validation of a model, progressive implementation is important. This can be achieved by progressively increasing the volume traded and regularly ensuring that the inputs used continue to have predictability and do not create noise.

10. Conclusion

Through an extensive analysis of trading performance, robustness, sensitivity, and expert insights, several key conclusions emerge:

The findings reveal that the Random Forest model performed the best. Strategies using Random Forest models displayed higher Sharpe ratios and cumulative returns overall. This success can be attributed to their ability to capture intricate, non-linear relationships within market data, enabling them to outperform traditional approaches. It's important to note that while this model excelled, others such as Logistic Regression did not exhibit significant improvements compared to technical analysis strategies. In addition, it is evident that the effectiveness of trading strategies is heavily dependent on the specific characteristics and dynamics of each commodity market. While some markets, such as crude oil, corn, and silver, showed higher risk-adjusted returns for Random Forest, others, such as natural gas, exhibited lower performance. This highlights the importance of customizing trading strategies to account for market-specific factors. The analysis also revealed significant variations in the performance of both ML and technical analysis strategies across different market phases. While ML strategies demonstrated resilience in certain periods, particularly during phases of steady volatility, they encountered notable challenges amidst periods of heightened volatility or significant market events. However, it is noteworthy that the Random Forest model exhibited some capacity to deliver performance even during periods of high volatility. The ability to navigate through various market phases is crucial for effectively adapting trading strategies to dynamic market conditions, ensuring robust performance across different scenarios. The robustness analysis served to underscore the significance of validating trading strategies. We were able to confirm that our best performing model, Random Forest, continued to deliver significant performance across select commodities, even under varying market conditions. Furthermore, our sensitivity analysis highlights the robustness of this strategy across several markets despite a slight decrease in performance. Finally, insights from industry experts highlighted the importance of data quality, market-specific inputs, regular reassessment of inputs used, and advanced featured engineering. While machine learning offers opportunities for generating alpha, challenges such as data complexity, model interpretability, the potential for overfitting, and the need for robust feature engineering must be addressed. In comparison with the literature, our RF model performed worse than that of Jevtic and Délèze in terms of profit factor (1.04 vs. 5.77) and Sharpe ratio (1.15 vs. 0.40). We also reached a worse performance than Manokhin in terms of Sharpe ratio as he obtained 0.58.

References

AJMERA, Richa and KOOK, Nancy, 2012. Impact of commodity price movements on CPI inflation.

Amazon.com: Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications: 8580001070285: John J. Murphy: Libros, [online]. Retrieved from : https://www.amazon.com/Technical-Analysis-Financial-Markets-Comprehensive/dp/0735200661 [accessed 29 May 2024].

BAKER, H. Kent, FILBECK, Greg and HARRIS, Jeffrey H., 2018. *Commodities: Markets, Performance, and Strategies*. Oxford University Press. ISBN 978-0-19-065603-4. Google-Books-ID: OZNODwAAQBAJ

BALL, Ray, KOTHARI, S. P. and WATTS, Ross L., 1993. Economic Determinants of the Relation between Earnings Changes and Stock Returns. *The Accounting Review*. Vol. 68, no. 3, pp. 622–638.

Basics of Algorithmic Trading: Concepts and Examples, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/articles/active-trading/101014/basics-algorithmic-trading-concepts-and-examples.asp [accessed 5 May 2024 a].

Basics of Algorithmic Trading: Concepts and Examples, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/articles/active-trading/101014/basics-algorithmic-trading-concepts-and-examples.asp [accessed 5 May 2024 b].

BENHAMOU, Eric. Distribution and statistics of the Sharpe Ratio.

BLUESTOCK, 2023. Risk Management in Algorithmic Trading. *Medium* [online]. 11 December 2023. Retrieved from : https://medium.com/@bluestock.in/risk-management-in-algorithmic-trading-0951e502d243 [accessed 5 May 2024].

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM (US), 1999. U.S.

Dollars to Euro Spot Exchange Rate. *FRED, Federal Reserve Bank of St. Louis* [online]. 4 January 1999. Retrieved from : https://fred.stlouisfed.org/series/DEXUSEU [accessed 17 May 2024].

BODIE, Zvi, TREUSSARD, Jonathan and WILLEN, Paul, 2007. The Theory of Life-Cycle Saving and Investing. *SSRN Electronic Journal*. DOI 10.2139/ssrn.1002388.

BORENSZTEIN, Eduardo and REINHART, Carmen M., 1994. The Macroeconomic Determinants of Commodity Prices. *Staff Papers - International Monetary Fund*. Vol. 41, no. 2, p. 236. DOI 10.2307/3867508.

BROCK, William, LAKONISHOK, Josef and LEBARON, Blake, 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*. Vol. 47, no. 5, pp. 1731–1764. DOI 10.2307/2328994.

CFI. Algorithmic Trading. *Corporate Finance Institute* [online]. Retrieved from : https://corporatefinanceinstitute.com/resources/equities/algorithmic-trading/ [accessed 17 May 2024].

CHICAGO BOARD OPTIONS EXCHANGE, 1990. CBOE Volatility Index: VIX. *FRED, Federal Reserve Bank of St. Louis* [online]. 2 January 1990. Retrieved from : https://fred.stlouisfed.org/series/VIXCLS [accessed 17 May 2024].

Commodity price cycles: Causes and consequences, 2022 [online]. Retrieved from : https://blogs.worldbank.org/developmenttalk/commodity-price-cycles-causes-and-consequences [accessed 3 January 2024].

Copper Jul 24 (HG=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/HG=F/ [accessed 17 May 2024].

Corn Futures,Jul-2024 (ZC=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/ZC=F/ [accessed 17 May 2024].

Crude Oil Jun 24 (CL=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/CL=F/ [accessed 17 May 2024].

DIXIT, Avinash K. and PINDYCK, Robert S., 1994. *Investment under Uncertainty*. Princeton University Press. ISBN 978-0-691-03410-2.

EDWARDS, Robert D., MAGEE, John and BASSETTI, W. H. C., 2018. *Technical Analysis of Stock Trends*. 11th edition. New York, NY : CRC Press. ISBN 978-1-138-06941-1.

FADLALLA, Adam and AMANI, Farzaneh, 2014. Predicting Next Trading Day Closing Price of Qatar Exchange Index Using Technical Indicators and Artificial Neural Networks. *Intelligent Systems in Accounting, Finance and Management*. Vol. 21, no. 4, pp. 209–223. DOI 10.1002/isaf.1358.

FAMA, Eugene F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*. Vol. 25, no. 2, pp. 383–417. DOI 10.2307/2325486.

FEDERAL RESERVE BANK OF CLEVELAND, 1982. 1-Year Expected Inflation. *FRED, Federal Reserve Bank of St. Louis* [online]. 1 January 1982. Retrieved from : https://fred.stlouisfed.org/series/EXPINF1YR [accessed 17 May 2024].

FENG, Shizhou and DU, Jing, 2022. Design of Quantitative Trading System Based on Data Mining Method under Software and High-Performance Computing. JIANG, Xiantao (ed.), *Mathematical Problems in Engineering*. Vol. 2022, pp. 1–11. DOI 10.1155/2022/6540928.

Forward Contracts vs. Futures Contracts: What's the Difference?, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/ask/answers/06/forwardsandfutures.asp [accessed 6 May 2024].

Forward Rate vs. Spot Rate: What's the Difference?, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/ask/answers/042315/what-difference-between-forward-rate-and-spot-rate.asp [accessed 6 May 2024].

GERLEIN, Eduardo A. et al., 2016. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*. Vol. 54, pp. 193–207. DOI 10.1016/j.eswa.2016.01.018.

HANNAH BALDWIN. How a strong dollar affects international currencies & commodities. *Reuters* [online]. Retrieved from : https://www.reuters.com/plus/how-a-strong-dollar-affects-international-currencies-and-commodities [accessed 9 May 2024].

HARRINGTON, Cynthia, 2003. Fundamental vs. Technical Analysis. *CFA Institute Magazine*. Vol. 14, no. 1, pp. 36–37. DOI 10.2469/cfm.v14.n1.2789.

HARVEY, Campbell R and LIU, Yan. Evaluating Trading Strategies,

How Commodity Pricing May Correlate to Inflation, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/articles/investing/020816/importance-commodity-pricing-understanding-inflation.asp [accessed 9 May 2024].

JEVTIC, Danijel, DELEZE, Romain and OSTERRIEDER, Joerg, 2022. *AI for trading strategies* [online]. arXiv:2208.07168. arXiv. arXiv:2208.07168. Retrieved from : http://arxiv.org/abs/2208.07168 [accessed 8 February 2024]. arXiv:2208.07168 [q-fin]

LATIF, Madiha et al., 2011. Market Efficiency, Market Anomalies, Causes, Evidences, and Some Behavioral Aspects of Market Anomalies. *Research Journal of Finance and Accounting*. Vol. 2, no. 9.

Machine Learning in Finance - Overview, Applications, [online]. Retrieved from : https://corporatefinanceinstitute.com/resources/data-science/machine-learning-in-finance/ [accessed 26 May 2024].

MANE, Shraddha, 2019. Why is Transparency and Explainability Important in AI? *Persistent Systems* [online]. 9 August 2019. Retrieved from : https://www.persistent.com/blogs/thoughts-on-transparency-and-explainability-of-ai/ [accessed 26 May 2024].

MANOKHIN, Valery, 2015. *'Applications of Machine Learning Techniques to the Analysis and Forecasting of Commodity Prices"*.

MCHUGH, Catherine, COLEMAN, Sonya and KERR, Dermot, 2021. Technical indicators for energy market trading. *Machine Learning with Applications*. Vol. 6, p. 100182. DOI 10.1016/j.mlwa.2021.100182.

Moving Average (MA): Purpose, Uses, Formula, and Examples, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/terms/m/movingaverage.asp [accessed 27 May 2024].

MURPHY, John J., 1999. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Prentice Hall Press.

NAMIN, Sima Siami and NAMIN, Akbar Siami, 2018. 1. Graduate Research Assistant and Ph.D. Student, Department of Agricultural and Applied Economics.

Natural Gas Jun 24 (NG=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/NG=F/ [accessed 17 May 2024].

PARK, Cheol-Ho and IRWIN, Scott H., 2007. What Do We Know About the Profitability of Technical Analysis? *Journal of Economic Surveys*. Vol. 21, no. 4, pp. 786–826. DOI 10.1111/j.1467-6419.2007.00519.x.

PHUOC, Tran et al., 2024. Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam. *Humanities and Social Sciences Communications*. Vol. 11, no. 1, p. 393. DOI 10.1057/s41599-024-02807-x.

Pros and Cons of Automated Trading Systems, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/articles/trading/11/automated-trading-systems.asp [accessed 5 May 2024].

RUNDO et al., 2019. Machine Learning for Quantitative Finance Applications: A Survey. *Applied Sciences*. Vol. 9, no. 24, p. 5574. DOI 10.3390/app9245574.

SALKAR, Tanishq et al., 2021. Algorithmic Trading using Technical Indicators. In : *2021 International Conference on Communication information and Computing Technology (ICCICT)*, pp. 1–6. Mumbai, India : IEEE. 25 June 2021. ISBN 978-1-66540-430-3. DOI 10.1109/ICCICT50803.2021.9510135.

SCHUMAKER, Robert P. and CHEN, Hsinchun, 2009. A quantitative stock prediction system based on financial news. *Information Processing & Management*. Vol. 45, no. 5, pp. 571–583. DOI 10.1016/j.ipm.2009.05.001.

Sharpe Ratio: Definition, Formula, and Examples, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/terms/s/sharperatio.asp [accessed 17 May 2024].

SHEN, Shunrong, JIANG, Haomiao and ZHANG, Tongda. Stock Market Forecasting Using Machine Learning Algorithms.

Silver Jul 24 (SI=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/SI=F/ [accessed 17 May 2024].

SINGH, Harman, 2024. Machine Learning Algorithms for Trading: Predictive Modeling and Portfolio Optimization (Part 2…. *Medium* [online]. 11 January 2024. Retrieved from : https://medium.com/@tkaur77t/machine-learning-algorithms-for-trading-predictive-modeling-and-portfolio-optimization-part-2-7edb60da83e8 [accessed 26 May 2024].

SU, Chi-Wei et al., 2023. The impact of consumer confidence on oil prices. *Energy Economics*. Vol. 124, p. 106820. DOI 10.1016/j.eneco.2023.106820.

Sugar #11 Jul 24 (SB=F) Stock Price, News, Quote & History - Yahoo Finance, [online]. Retrieved from : https://finance.yahoo.com/quote/SB=F/ [accessed 17 May 2024].

SZAKMARY, Andrew C., SHEN, Qian and SHARMA, Subhash C., 2010. Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking & Finance*. Vol. 34, no. 2, pp. 409–426. DOI 10.1016/j.jbankfin.2009.08.004.

Trading the Death Cross [ChartSchool], [online]. Retrieved from : https://school.stockcharts.com/doku.php?id=trading_strategies:death_cross [accessed 27 May 2024].

UNIVERSITY OF MICHIGAN, 1952. University of Michigan: Consumer Sentiment. *FRED, Federal Reserve Bank of St. Louis* [online]. 1 November 1952. Retrieved from : https://fred.stlouisfed.org/series/UMCSENT [accessed 17 May 2024].

USHA et al., 2021. *Commodity and Forex trade automation using Deep Reinforcement Learning*.

What Is a Commodities Exchange? How It Works and Types, *Investopedia* [online]. Retrieved from : https://www.investopedia.com/terms/c/commoditiesexchange.asp [accessed 6 May 2024].

WILLIAM F. SHARPE, 1994. The Sharpe Ratio. *Standford* [online]. 1994. Retrieved from: https://web.stanford.edu/~wfsharpe/art/sr/sr.htm [accessed 18 May 2024].

YEN, Stéphane Meng-Feng and HSU, Ying-Lin, 2010. Profitability of technical analysis in financial and commodity futures markets — A reality check. *Decision Support Systems*. Vol. 50, no. 1, pp. 128–139. DOI 10.1016/j.dss.2010.07.008.

Use of Artificial Intelligence-assisted tools

In the context of this work, the author declares having used Artificial Intelligence-assisted tools for the following purposes:

- Improvements of form (spelling, syntax, reformulation, report structure)
  Mention of AI tools used: CHATGPT.

- Data collection and coding
  All the statistical and performance calculations were performed using Python version 3.10.8. Mention of AIs used: GitHub Copilot

Appendix 1 – Interview of respondent A

Interview protocol – Respondent A

Code	Question
QP01	Could you briefly introduce yourself and your experience with machine learning
QP02	From your experience, what do you think are the main factors that can negatively affect the performance of a machine learning model?
QP03	Would you say volatility can impact a model's performance?
QP04	And what are your thoughts on the effectiveness of machine learning compared to traditional methods such as technical analysis only?
QP05	What input data do you find to be most significant for the models, and what are the main criteria to select those variables?
QP06	Can it be a problem for the machine learning model to have input data that is published on a quarterly basis and daily prices?
QP07	What do you think are the main disadvantages of machine learning for commodity trading?
QP08	And what is the best way, in your opinion, to validate a model?
QP09	What would you recommend to a company that wants to build ML models for commodity trading?

Interview transcription - Respondent A

Code	Response
QP01A	I'm a quantitative analyst at a financial institution here in the US. I've been working in the field for about two years now, primarily focusing on developing and implementing algorithmic trading strategies for equities and foreign exchange markets. I'm still learning and exploring this field myself, but I'm happy to share what I know and hopefully learn something new from your questions as well.
QP02A	First off, bad data. If the data you feed your model is full of errors, missing info, or biases, the model's going to learn the wrong things and make bad predictions. Another issue is overfitting. The model gets so good at predicting the training data that it fails when faced with new, unseen data. Choosing the right features and transforming them in a way that makes sense for the model is also super important. In commodity trading, this could mean things like macroeconomic indicators, weather data, inventory levels, even sentiment analysis from news or social media. And last but not least, domain expertise is crucial. Machine learning is a powerful tool, but it's not a substitute for understanding the fundamentals of commodity trading. You need to know what makes these markets tick, how to interpret the model's output, and where the potential pitfalls are.
QP03A	Absolutely! Volatility can definitely mess with a model's performance. Volatility is definitely something to watch out for when building machine learning models, regardless of the asset class. If we're talking commodities specifically, my intuition tells me that volatility could be even more of a challenge. These markets tend to be more sensitive to external factors like weather patterns, geopolitical events, and supply chain disruptions, which can lead to some pretty wild price swings. The increased noise and weaker relationships between variables in a volatile market can really throw off a model's predictions. And there's also a higher risk of overfitting, where the model mistakes those random fluctuations for real trends. Now, I'm not a commodity trading expert, so I can't say exactly how to address this issue. But my guess is that you'd need to take some extra precautions when designing your model. Maybe incorporate features that capture volatility, choose a model that's more robust to outliers, or use regularization techniques to prevent overfitting.
QP04A	I think the answer is: it depends. It depends on the specific market, the quality of the data you have, the skill of the person building the model, and how much risk you're willing to take. The way I see it, machine learning and technical analysis aren't necessarily competing against each other. They can actually complement each other really well. You could use machine learning to identify potential opportunities, and then use technical analysis to confirm those signals before you pull the trigger on a trade.
QP05A	Since my experience is primarily in the stock market, I'll focus on what I've found to be most significant there. The same principles might apply to commodities to some extent, but it's important to remember that each market has its own unique dynamics. In the stock market, I'd say some of the most important input data for machine learning models include: Price and Volume Data: This is the most basic and often the most valuable information. Historical price data can reveal patterns, trends, and volatility. Volume data can indicate the level of interest and activity in a particular stock. Fundamental Data: This includes financial statements (like earnings reports and balance sheets), analyst ratings, and news about the company. These factors can give you insights into the company's financial health, growth prospects, and potential risks. Technical Indicators: These are calculated from price and volume data and are used to identify trends and patterns in the market. Examples include moving averages, relative strength index (RSI), and Bollinger Bands. Macroeconomic Data: Interest rates, inflation, and GDP growth can all affect the overall market and individual stocks. Including this data can help the model understand the broader economic context. Sentiment Data: News sentiment, social media sentiment, and other measures of

	investor sentiment can be valuable inputs. Market sentiment can play a big role in short-term price movements. As for the selection criteria, it really boils down to a few key questions: Is it relevant? Does this data point actually have a meaningful relationship with the stock's price? Is it reliable? Is the data accurate, timely, and consistent? Is it predictable? Does the data exhibit any patterns or trends that can be modeled? Is it unique? Does it offer information that isn't already captured by other variables in the model?
QP06A	I think when you aggregate daily data to a quarterly frequency, you inevitably lose some information. For example, the daily volatility of a stock might not be captured in the quarterly data. This can limit the model's ability to learn subtle patterns that might be useful for prediction. Also, If the model is not carefully designed, it might give too much weight to the higher frequency data (daily prices), simply because there's more of it. This could bias the model's predictions and lead to suboptimal performance.
QP07A	One major disadvantage could be the lack of high-quality data. Commodity markets can be less transparent than stock markets, and reliable data might be harder to come by. This could make it difficult to train a model effectively. Another challenge could be the complexity of commodity markets. They're often influenced by a wider range of factors than stocks, including weather patterns, geopolitical events, and supply chain disruptions. This could make it harder for a machine learning model to capture all the relevant information and make accurate predictions. Additionally, the black-box nature of some machine learning models could be a concern for commodity traders. Since these markets can be volatile and unpredictable, traders might be hesitant to trust a model that they don't fully understand. It's also worth noting that implementing and maintaining machine learning models can be expensive and time-consuming. This could be a barrier for smaller firms or individual traders who don't have the resources to invest in this technology.
QP08A	There are several ways. One of them is essentially splitting your historical data into 'k' different folds or subsets. You then train the model on k-1 of those folds and test it on the remaining fold. This process is repeated k times, with each fold getting a turn as the testing set. Rolling cross-validation is another really useful technique, especially when dealing with time series data like we often do in financial markets. It's a bit more computationally intensive than k-fold cross-validation, but the added benefit of mimicking the real-world trading environment makes it well worth the effort in my opinion.
QP09A	Think about it: the sheer amount of data we're dealing with is insane. And it's not just about quantity, it's about quality too. We need clean, relevant data to feed these models. Then there's the whole issue of model complexity. We need to find that sweet spot between accuracy and interpretability. It's not just about the tech though. There’s a need for the right people too – experts who understand both machine learning and the nuances of commodity markets while having a good knowledge of risk management. So, what's the bottom line? It's all about finding the right balance. Invest in the right infrastructure, build your expertise, tailor your models, manage your risks, and never stop learning. You’re probably on the right track if you do that. Another thing is also to have teams from validation and implementation that at least understand somewhat what are the objectives and have a general picture of what we are trying to achieve.

Appendix 2 – Interview of respondent B

Interview protocol – Respondent B

Code	Question
QP01	What are the primary strategies that you use and how do you decide on these strategies?
QP02	Would you say that, for example, fundamental analysis will work best in some commodities and technical analysis in others?
QP03	Do you use any backtesting method to ensure that your strategy will perform well in the future?
QP04	What do you think are the main factors that can negatively affect the performance of your strategy? Volatility?
QP06	Have you ever considered using machine learning trading strategies, and what are your thoughts on their effectiveness compared to traditional methods?
QP07	What are the main factors affecting the usage of machine learning in commodity futures
QP08	What do you think are the main disadvantages of machine learning for commodity trading?
QP09	Why don't you use machine learning?
QP10	What do you think would be the best input variables to use in generalist machine learning model?
QP11	Can you briefly discuss how different market conditions, for example, bull versus bear markets or volatile markets can affect the performance of strategies?

Interview transcription - Respondent B

Code	Response
QP01A	So I like to use a combination of fundamental and technical. First of all, I find that mixing the two allows you to express a fundamental view but with good timing. Because I find price action can be influenced by a number of different factors and even though fundamentals are a view on supply and demand, sometimes there is a lot of interference, shall we say, with other contributing factors or correlations with other macro levels. So it's very difficult to get the timing right from a fundamental view. It depends on the market as well. For example, I trade a lot of EUAs and the EUAs is a market that has fundamental data on a high frequency. So, for example, in comparison to something like the electricity market that has a very, very high frequency of data and demand and supply balancing, you don't get so much volatility in the market. Therefore, it's a little bit easier to express a longer-term fundamental view and a longer-term fundamental view I find a bit easier to express through options than necessarily through outright positioning. So it's a bit of a combination, but I would say intraday and in day-to-day momentum, I use more technical analysis.
QP02A	Yes, it depends on the market participants and the range of types of market participants. Because there are some markets that are more accessible by investment funds, hedge funds, banks, etc. than there are maybe others. For example, crude oil is one that a lot of people focus a lot of attention on in terms of money flow in and out. Similarly, something like EUAs, because they've got a lot of attraction, because it's a naturally market that's getting tight over time. So it attracted a lot of long investors to start with, but also as fundamentals have changed, you get a lot of short investors. So it really depends on whether, like I say, this inflow and general commodity investing has a momentum, but also how liquid the products are, etc.
QP03A	No, I don't implement, shall we say, systematic measures. I do it with charts, so technical analysis based on charts, because it's more about price action, it's not about news or supply and demand. You generally get a visualization of whether your decision or your strategy is appropriate or not. But physical backtesting, no. But the types of technical strategies I use are, depending on the timeframes, but for example, for more intraday or intra-week, I use point and figure. I use an actual platform called Updater, and it does a lot of those calculations for you and gives you price signals on a timeframe that you want to look at. So if you're looking short, medium or longer term. And then I look at candlestick charts with various, my own trend lines on those. And I cross-reference those to other market reports on a similar thing.
QP04A	So volatility is obviously one, and I think for me it's big market events. Because price action can, all of these more technical analysis that are just purely looking at momentum or price movements, can work, but it cannot predict a big event. Like, for example, when we had the Ukraine war, or when you get news about, in some commodities, when you get news about interest rates, you just get a market reaction. Also, obviously, the big market reactions, for example, in the EUAs, we have a daily auction, and an auction either clears close to spot price or not. If there's a big premium or a big discount, you get immediate price action. And all of these systematic strategies, unless they embed this behavioral element into them, I don't think they're very good at predicting how, because you can get a very, very different change in momentum very quickly from a bit of data like that.
QP05A	So I think in some markets they're essential these days, because in some markets that is what's driving prices, or that's the only way you can get an edge in those markets. For example, I've traded in the past a lot of electricity. Electricity, as I

	mentioned, is high frequency, it's very, very high data. You've got a lot of weather data, you've got hourly pricing, you've got so many factors that require a lot of data crunching, and therefore I think is more attuned to machine learning models. Also, because you have a lot of history of how these markets balance, machine learning can be a very effective way to actually compare, let's say, the supply and demand outlook versus a historical outturn of price, and therefore you can get a good proxy of where this market could outturn, and therefore you can position yourself based on that. So that, for me, if you trade in the spot electricity market, you have to have that minimum. And for example, without that, it becomes a very, very difficult market to follow, or to make money out of. And there's a lot of new, MAller commodity trading houses that focus on this type of approach, that are making a lot of money, especially in gas and power.
QP06A	Yes, exactly. Especially also if you have very difficult physical, or if you have very strict physical, not just physical delivery options, but for example, some markets are very highly controlled in terms of position sizing and things like this, obviously to protect the market, and that becomes even more difficult for a lot of these strategies to be pushed in, or to be used in. However, for example, the cocoa market is a soft market that isn't a particularly big market, but a lot of people are now talking about the fact that people have used a bit more, it's attracted this type of trading, because there's a story behind it. And the thing is, if there's a fundamental story that is driving price, what you then find is some of these machine learning models accentuate the move, because they work on understanding trends, or they work on understanding this sort of price move. And I think this is where sometimes you get, if the machine learning approach can be adapted to different markets, and not just focusing on one market at a time, but move to the markets where it's working at the time, or is irrelevant, then I think it can make a lot of, I think they're very profitable.
QP07A	I think for traditional commodity traders, I think the issue is, is you get this complete disconnect between fundamentals and price action, from machine learning and automated trading. Because they could be following price action, rather than fundamentals, you find that sometimes you get this big disconnect, which then fundamental traders, purely fundamental traders, either have to absorb and hold positions through big drawdowns, or you then, also then discourage people to trade longer term fundamentals, and it has to be based on shorter term, where the actual fundamentals are coming into play. The other disadvantage, I think, is you get crowded trades, meaning you get a situation where the whole market is long, or the whole market is short. And then when that happens, and then something does change, something fundamentally change, or something really changes that direction, or move in that market, I think that then creates a huge amount of volatility and risk, for all of the types of traders, because it's like everyone rushing for the door at the same time, when you're trying to get out. That's the only thing I would say, you get this crowded trade situation, which is not always optimal either.
QP08A	Because I don't have the, one, I don't have the time or the technical skills. It's the programming, it's to be able to use that data, to be able to implement it. I currently have a quantitative trading assistant, who's working on some of this, but again, we're focusing on fundamentals to start with, and then we will look to layer something else. What can we learn from auction, price outturn versus price action? That sort of thing is another strategy that we can implement. But right now, like I say, even though there's a lot of volatility, and there's a lot of opportunity in the market, the main play at the moment, is more longer term fundamental strategies, that are more option based. But I would like to be able to implement that as well. But again, you need the resources

	and the expertise.
QP09A	I think it's quite interesting, because it depends what rules you put around these. Because for example, Bollinger Bands, I know certain traders that trade in a very technical way, that will only reallty look when it starts to hit the upper band. Because it's almost squeezing out, or it's breaking out. And they look for breakouts, rather than saying, okay, we've got to the extreme of the band, therefore we should maybe sell it. Whereas I know some traders almost do contra trading and saying, well, we're at the Bollinger Band, we're really pushing out, we're actually going to squeeze out of this, we're going to break out of this range, and therefore we'll go long. So I mean, yes, I would say MACD, which has got the stochastic oscillators, MACD, RSI, Bollinger, all stuff I use. So I think they are good. The problem you also have is a lot of these indicators work in a trending market, and some of them work better in range bound markets. And I think for me, and at times I have looked at more, shall we say automated trading strategies, or I have tried to almost find an indicator that says, are we range bound, or are we trending? Because I think if you're in a different regime, maybe it could help in terms of determining which ones you give more weight to, for example. I think it's interesting that you've used the Baltic Dry Index. While you have, I would say, that is, you have to understand that's dry commodities, and not oil, etc. And also not all commodities as in the futures, are not representative of including a freight price. Yeah, because most of them are a location where it's FOB, effectively. The spot price feeds into a FOB. So I'm not sure. I don't know if you've seen something different there. I think inflation is a good one. I think also though, by using US versus maybe Europe might help. Okay. But yeah, dollar definitely affects things. And interestingly enough, the VIX Yeah, the VIX. is, I would say, is probably more of a delayed indicator.
QP10A	Yeah, like I say, you know, we've had a lot of events driving a lot of consequences, consequences of these events driving markets and market flows and changes in this, which has drastically created whether you're in a bull, bear, and as I said, more of a trending versus consolidating range bound market. I would say we're getting to more of a consolidation market, lower vol market in some markets. But then again, you know, the geopolitical situation can flare up at any point and can create more. But and also a lot of these markets then generate a lot of premium in the markets based on these potential events or this potential risk premium they're putting in the market. And as soon as it is removed or there's a headline that says this is not really, you know, it's not really an issue anymore. You can see big volatility as well. I mean, I think if you can find a model that works in a range bound market, I think you'll, you know, a lot of traders very happy because I think in that environment we've coming out of an environment now where traders have been used to making a lot of money because it's been big news. I think people going into this more consolidation time, they're struggling to make the money they were used to. They're looking at putting spreads on that used to maybe make \$20 and now make two. Yes, I think machine models can help on that. But also, though, I think machine learning models should also help on the risk management, meaning the sizing of the trades and that type of thing, because obviously if you're used to making \$20, you maybe only have 100 lots on. But if you're making \$2, you're probably going to, you

	know, you're looking to put a thousand lots on. Is that the best strategy? Because you're trying to still make the same sort of money. So I would say machine learning alongside how to, you know, how your positioning and your risk can be adopted according to that, I think is super important. Like ATR, for example, I think is a really good measure of understanding what your drawdown has to be to be able to try to make money in that market. Your average true range gives you an indication of how volatile that market is. And therefore, if you cannot hold your position for at least two ATR, then you shouldn't be entering into that. So, for example, if your ATR is 2.2, you've got to realize you've got to be able to hold a position through a 4.2 move of that, whatever that price is. Otherwise, you will get stopped out too quickly. Yeah, because these sorts of things, I think machine learning can be used for a lot more than just, you know, price indication.

Appendix 3 – Interview of respondent C

Interview protocol – Respondent C

Code	Question
QP01	Could you briefly introduce yourself and your experience in commodity trading?
QP02	From your experience, what do you think are the main factors that can negatively affect the performance of a machine learning model?
QP03	And what are your thoughts on the effectiveness of machine learning compared to traditional methods such as technical analysis only?
QP04	What input data do you find to be most significant for the models, and what are the main criteria to select those variables?
QP06	Can it be a problem for the machine learning model to have input data that is published on a quarterly basis and daily prices?
QP07	Do you think it is possible to have a generalist model that trades on all commodities? Or is it better to have one model for each commodity?
QP08	Can you discuss how different market conditions, for example, bull, bear markets, trending markets affect the performance of machine learning models?
QP09	What would you say are the main disadvantages of machine learning for commodity trading?
QP10	And what is the best way, in your opinion, to validate a model?
QP11	Do you see overfitting happens very often in your models, or is it something quite rare?

Interview transcription - Respondent C

Code	Response
QP01A	Sure. I've been in commodities trading since 2008. Mostly in energy, so oil, power, gas, carbon, coal. I've looked at other commodities like side, like gold, metals, a little bit of soybeans and other commodities. I've done both physical and financial derivatives on commodities. I worked for a long time on optimization problems in commodities, which is kind of like the backbone for trading and on a physical basis. And then a few years ago, I started up a more quantitative trading desk in which we first deployed more like typical indicator strategy or indicator-based strategies. We played around with some machine learning techniques on top of that and separately as well. And then kind of also traded discretionary in that space as well. And yeah, so I've done both physical and financial sides.
QP02A	No, I don't think high volatility is necessarily a problem. I think it's the fact that markets are not really, markets are stable. I think the regime switches can occur quite quickly and it's very difficult to train models for that because it's, you don't have enough data for each of those periods. So that makes it, I think, quite difficult to train machine learning models on that.
QP03A	Well, there's a difference between technical analysis that's indicator-based and technical analysis that's not indicator-based. So I think indicator-based trading suffers from slow response, because ultimately it's not predictive. Indicators are not predictive at all. So they're basically trying to measure the particular state of the markets. For example, if you use, I don't know, you're using like an RSI, it will tell you, okay, this market is now overbought or oversold. But that's just detection based on how the market has moved. It doesn't really tell you what's going to happen next. And markets can continue to be overbought or oversold longer than the indicators will indicate. And sometimes indicators will because they recalculate and recalibrate themselves. So I think they don't work very well all the time, unless the market is in a particular state. So it's almost the opposite way. An indicator doesn't tell you what the market is going to do, but it can detect and calculate that the market is doing something. And if the market continues to do the thing that is kind of, I don't know, like precedes what normally the indicator says, and if that continues to happen, then the market or the model will make money. I guess simplest example is if you have a moving average crossover, if you detect that crossover, if the market actually does trend and the trend lasts long enough so that you can get in at the crossover and stay in until the next crossover, if that period in between the two crossovers is long enough to get some money out of the market, then the model will work. But that has nothing to do with how the market is behaving now. That's a feature of the market itself. Markets tend to trade. And therefore, whatever detection algorithm that you're using is going to effectively be profitable. And I think machine learning has a little bit better of a chance of fine-tuning your sensitivities versus inputs, but because it has more of that sensitivity nature, it's also prone to overfitting. So I think that's kind of where the difficulty sits.
QP04A	I think when you use, like, because you gave me a list of some of the indicators, I think it's important that you don't over-validate. Like, you should have one indicator that measures momentum. You should have one indicator that measures volatility.

	You shouldn't do, like, five indicators that all roughly calculate the same thing but in different ways, because that doesn't work very well, because it makes your model more confused. And so, I think calibrating what you put in and why you put in the input is more important. I mean, the more diversified your scope can be of inputs, the more interesting it can be. And you mentioned a few things like dollar index or the dollar currencies that you use. I think that's relevant, because it kind of has a macroeconomic impact. But they all have different timescales that they operate.
QP05A	It doesn't have to be, but you have to be absolutely sure about when the data became available. Like, saying this inflation number came out on, I don't know, the 14th of May, that's not good enough. You have to say it came out at 2.30 on the 14th of May. Like, you actually have to dig deeper. So I think you can still use that data, but you will have less data points that you have to try and disentangle what the effect is. Because ultimately, you're trying to find a signal and a response. So you might have inflation data coming out, how does the market respond to that? And that will depend on what state the market is in and what the release of the information is that comes out. Because you get bullish data, bearish data, neutral data. And if the market is like overheated, and bullish data comes out, what does it do then? Like, so your model should, you should give it room and kind of scope to work and to fine tune that. But I don't think it's necessarily a problem. You just can't assume. And I think the other component of a model should be how long is this signal valid for? You get inflation data coming out, what's the impact? How long does it last? Is it an hour? Is it a day? Is it a week? Is it a month? And or anything in between, but that has to be measured. Because without that, it becomes very difficult to do anything with that. This inflation data tends to be absorbed in like two days, if it's like a heavy, heavy point, most of the time. So I think these kind of things need to be embedded within the model somehow, like, how long is your memory? What kind of timescale do you work on?
QP06A	Not all commodities trade in the same way. So it depends on the timescale. So if you work on really fast timescales, I think you can probably have a model that's similar as long as it's adjusted to volatility, which sounds like you have right into volatility goes into your input. And I think you can train that across. But on longer timescales, I don't think that works very well.
QP07A	That's like, they are the big differentiator. Of course, of course, it affects that, right? It's, if you've got a very big bull market that just keeps going and going and going, like almost all models will work. It's very difficult to create a model that doesn't, at some point, get long in a bull market. Okay. Even if you're late to the party, the party keeps going, you'll still have a good time. So I think finding models that work across different segments or different states of the market is very difficult. Mm-hmm.
QP08A	I think that they don't have enough data. So I think the reliability of those models can be quite poor. And commodity markets also change rapidly. I don't know, Ukraine is now closed, so there's no more Russian gas coming in. This has affected the grain market as well. These things are not necessarily obvious in your model.

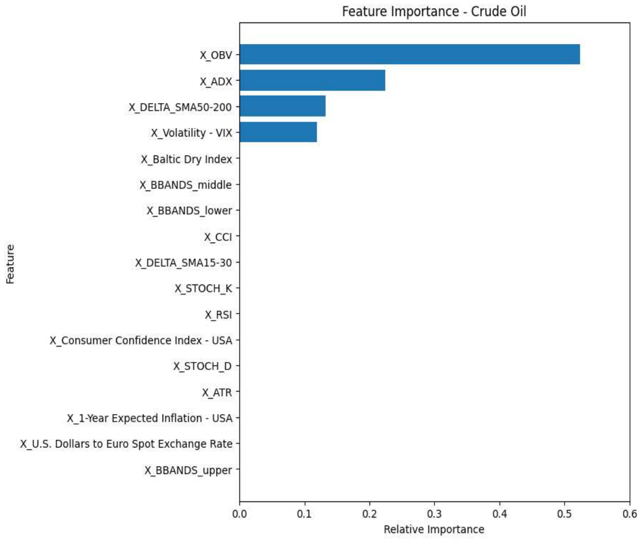
	Models are better with the adjustments, but I think commodity markets are continuously changing, regulation changes, quality specs change in your energy commodities. You run out, countries go to war. I think that in general, the energy landscape can change, or the commodity landscape can change quite a bit. You have to somehow find a way to teach your model that as well, which is not easy. Because how often have you had data where there was a Russian war that affected Europe for two years non-stop, where we ended up building new terminals to import gas into Europe? How often has that happened in the past?
QP09A	Take some out of sample, take some in sample data, and then keep rolling that forward. So you test, or you train, you test, you run, and then you keep rolling that forward. I think that probably makes the most sense if you want to do it structurally. Choosing those periods is also quite an art. I think that's why discretionary trading still works well, because humans are still better at determining whether today was a good representation for tomorrow or not.
QP10A	I think it happens all the time. It is important when a model is validation to regularly increase the volume of trade and regularly reassess the inputs to ensure there isn’t anything new going on. After big market events it is even more important to ensure that the importance we put in each input is assessed.

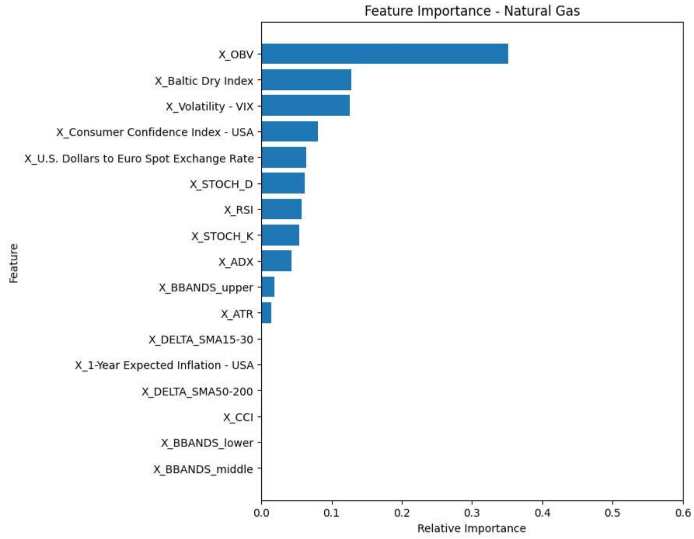
Appendix 4 – Summary of parameters

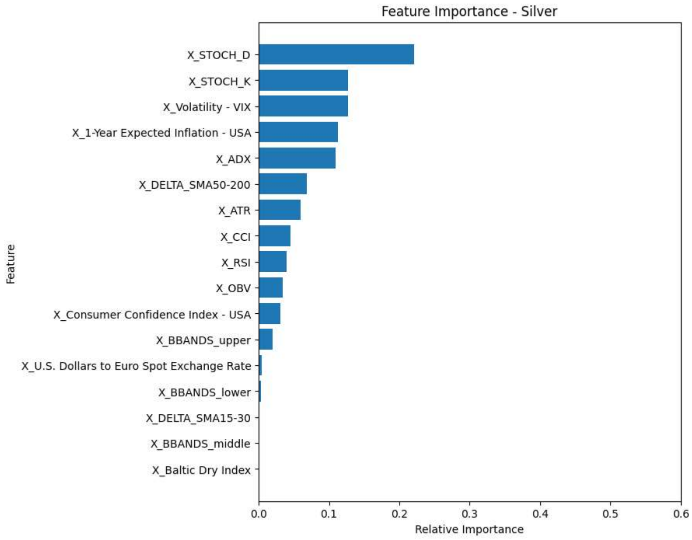
Strategy	Parameter*	Value
MA	N_1 N_2 b	9 21 0.05
MFI – RSI	N_{RSI} N_{MFI} SO_{MFI} SO_{RSI} BO_{MFI} BO_{RSI}	14 9 30 30 60 60
MSV	N_1 N_2 e b g	9 21 6 0.01 10

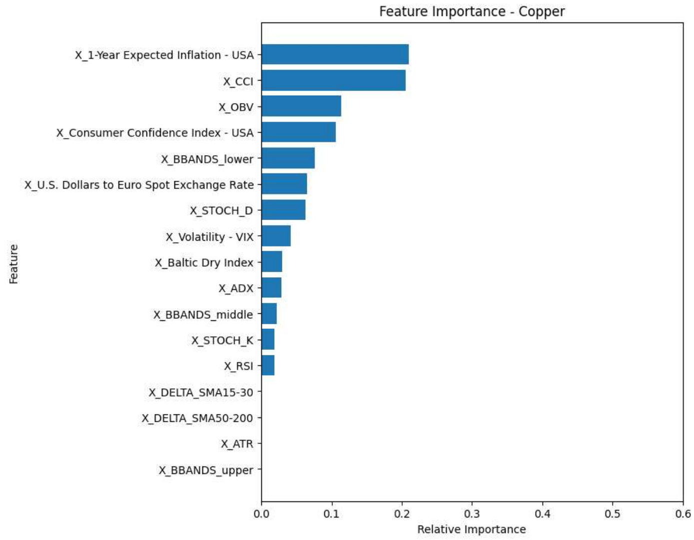
*The exact definition of parameters can be found in the methodology section and the respective strategies section.

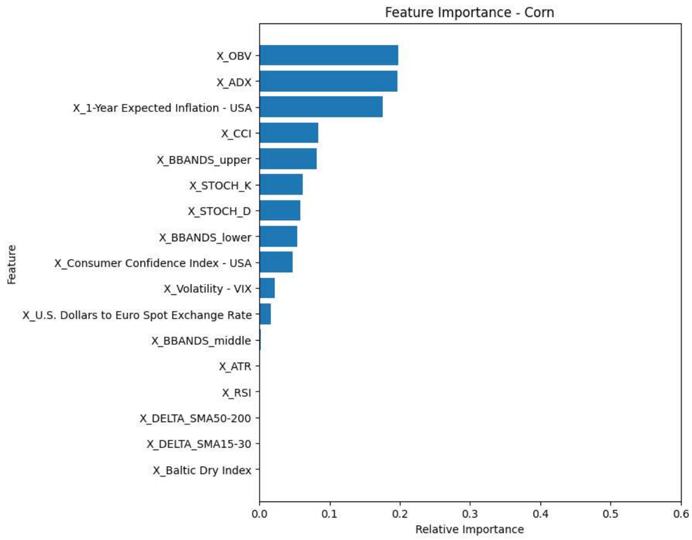
Appendix 5 – Feature Importance (Logistic Regression)

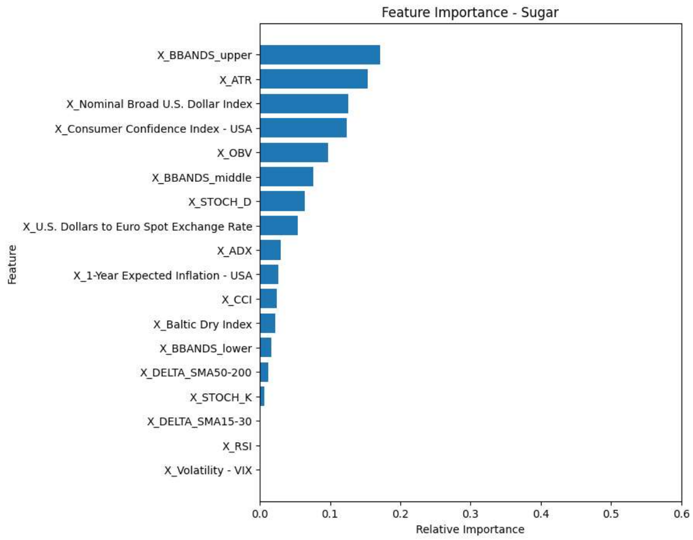




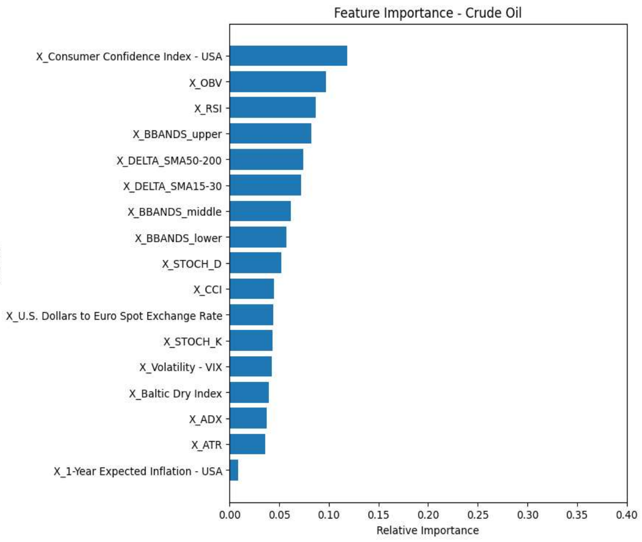


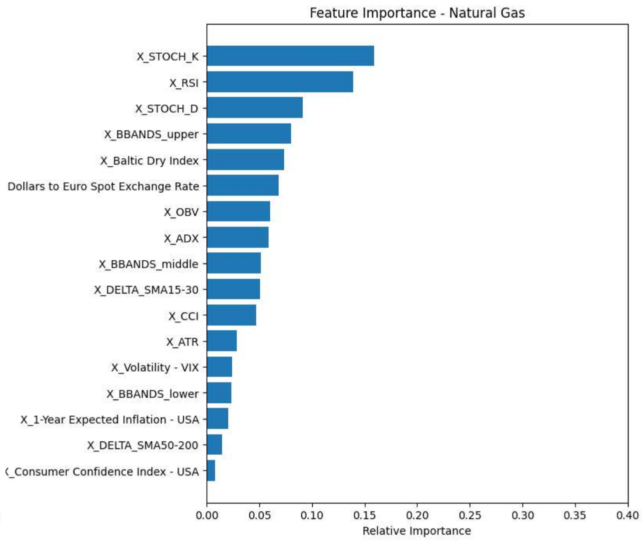


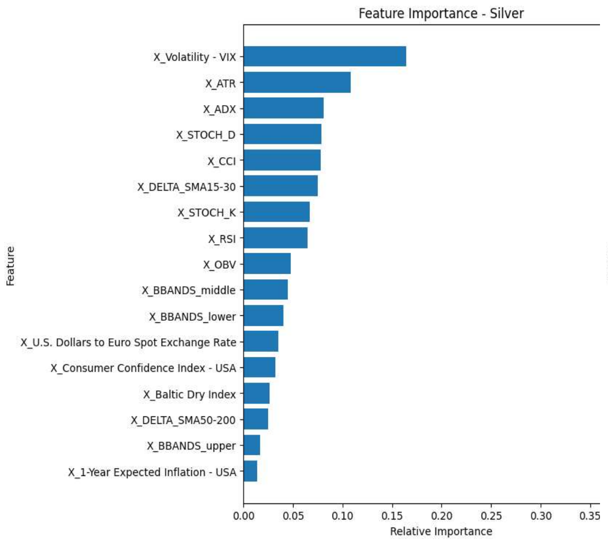


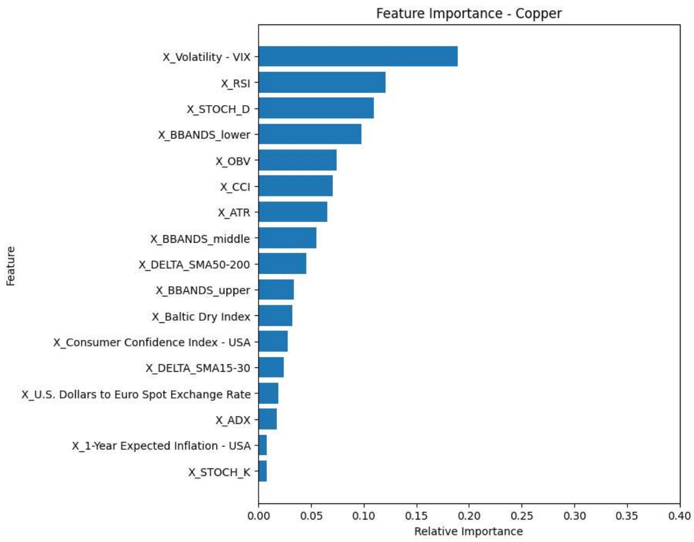


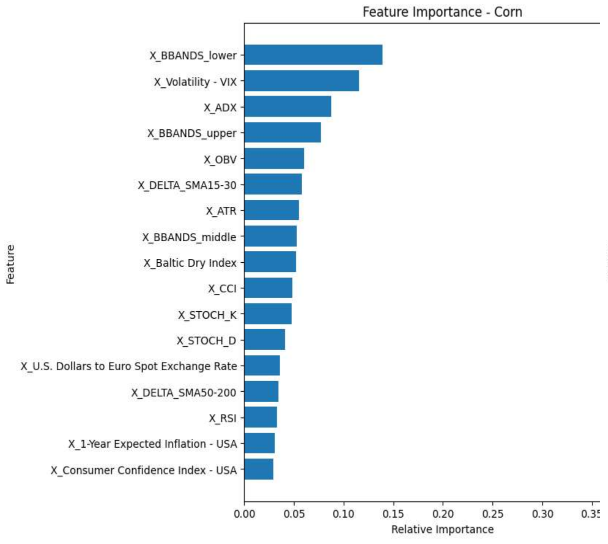
Appendix 6 – Feature Importance (Decision Tree)

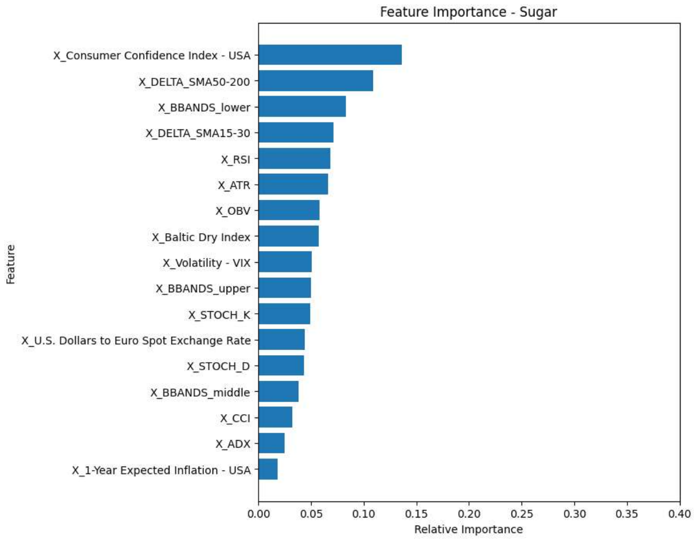












Appendix 7 – Feature Importance (Random Forest)