Geopolitical Event Sentiment Analysis and Product Pricing Prediction for Commodities

MADHURI PANCHUMARTI (87299)

HARDIK SETHI (102316)

MSC IN DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE
ADVISOR: SACHIN KAMBLE

I certify that the master project being submitted for examination is my own research, the data and results presented are genuine and actually obtained by me during the conduct of the research and that I have properly acknowledged using parenthetical indications and a full bibliography all areas where I have drawn on the work, ideas and results of others and that the master project has not been presented to any other examination committee before and has not been published before.

ACKNOWLEDGMENT

We, Madhuri Panchumarti and Hardik Sethi, would like to thank our Professor/ Master thesis advisor Mr. Sachin Kamble for helping us, guiding us, and supporting us all through this master's thesis. Thanks to his expertise in research methodologies and AI, we had a solid base; his wise suggestions at each stage improved both our project and the research results.

Thanks to our families, who have always supported us with encouragement, patience, and full confidence in us. Their encouragement was vital for us to keep going through the most difficult times in this research project.

We are grateful for our friends and classmates because their talks, teamwork, and good attitude made learning more enjoyable. Counsel from them made it easier for us to move forward and solve challenges from a new angle.

We are thankful for the tools and information that let us complete this research project. Especially, the Google Cloud tools we used, like Google Collab and Google Scholar, allowed us to move quickly, experiment a lot, and access important scientific results, which made our model building very strong.

We would like to show our gratitude to all the people who played a part, however small, in this project's accomplishment. We greatly appreciate your support and assistance because they were important in forming the end result of this study.

TABLE OF CONTENTS

List of Figure:

Abstract

This thesis aims at studying the impact of geopolitical news on the movement of price for the global commodity markets, for instance commodities like “wheat “, “Natural Gas”, “Gold” and “Crude Oil”. Researchers combined Machine Learning methodologies like Radom Forest, XGBoost and LTSM related architectures to effectively analyze the sentiments behind the new related to the stock market, based on behavioral finance rules and regulations show casing how investors movements can impact the market movements

The study focuses on observing how the sentiment analysis from the FinBERT algorithms is connected to certain features of each news help create predictions. The dataset consists of 1200 headlines, and each line is allocated to each sentiment related to changes in the commodities ETF’s prices within a window period of + or – 3 days. On the other hand, SHAP Analysis explains that the model’s performance depends on the polarity of the emotions, type of the commodity and the temporal details (month and weekday), but the maximum accuracy of the model was given as 76%. Lee and Lundberg (2017), Yang et al. (2021) and Araci (2019).

Furthermore, the LTSM Regression model is used to generate the size of price movements which resulted in the Root Mean Squared error not more than 1.2% which is a good sign. In addition to that, the study explained how the sentiment prediction impacted gold and oil and concluded that they are more stable that other commodities like Wheat and Gas. . Schmidhuber and Hochreiter have argued for this view in a 1997 paper.

The results show that, taking a certain commodity into account, the geopolitical trends can predict the movement of price in a relatable manner. The Framework to monitor geopolitical risk, tools to

make policies, and real time trading platforms or businesses which depend on these commodities can be benefited from this observation.

Research Questions and Hypotheses

The aim of this research to examine how sentiments related to finances and the dynamic effects of geopolitical news changes in commodity prices. It also examines how the machine learning algorithms anticipate these modification as opposed to the more conventional techniques and statistics. For our thesis, we are handling three research questions along with the hypothesis linked to it. Our study's findings are the supporting material and pertinent literature review

for the conclusion

Research Question 1: What is the impact of geopolitical events on commodity prices?

Hypothesis: Geopolitical events severely affect commodity prices.

The initial research question examines how disruptions from geopolitical factors like conflicts, sanctions, and trade barriers influence commodity pricing. The hypothesis suggests that these occurrences create market uncertainty, decreasing demand and subsequently applying downward pressure on prices. However, the results of **this study contradicts the hypothesis**. Empirical analysis showed that geopolitical conflicts frequently lead to rising commodity prices, especially for energy products like crude oil and natural gas, as well as safe-haven assets such as gold.

This contradiction results are clarified by the expectations of market participants regarding supply disruptions, production halts, and heightened transportation risks, all of which tend to increase the commodity prices. Speculative trading during times of uncertainty might also lead to temporary price surges. This study is backed by Ratti and Vespignani (2015), which displayed that geopolitical disturbances have an increased impact on oil prices by interfering with global supply chains. On the other hand, the data did not support the hypothesis; instead, geopolitical incidents are more inclined to result in price increases rather than drops.

Research Question 2: Does financial sentiment extracted from news articles influence short-term commodity price movements?

Hypothesis: Positive financial sentiment is associated with short-term effect in commodity markets.

This report analyses whether sentiment extracted from financial news can predict short-term price behaviour in commodity markets. The idea is inspired by behaviour fie finance theory: a market's mood modifies the price of assets.

We took 75% of the data for training and the remaining 25% for generalisation. We use the FinBERT language model to calculate sentiment scores, and then embedded this data into both linear and non-linear prediction models.

Different observations were observed. Although a traditional OLS regression found no relationship between sentiments posted and the stocks' prices (coefficient = 1.6059, p = 0.326), XGBoost used a different methodology and identified some important relationships. By evaluating SHAP values, the analysis found that the sentiment aspects play an important role in model predictions, evidencing that they can be real and somewhat complicated. The study of Melnikas et al. (2021) agrees with the observations of Nassirtoussi et al. (2014) that text mining is helpful for forecasting the financial markets. For this reason, **the hypothesis was partially supported, since sentiment affects price trending, but linear models do not explain the relationship well**[33].

Research Question 3 is: Would machine learning methods work better than traditional regression for forecasting changes in commodity prices?

Through our analysis, we recommend that Machine learning classifiers (such as XGBoost) to achieve improved results in predicting the price movement and volatility when compared to OLS regression.

The final research question tries to understand whether modern machine learning algorithms outperforms traditional regression models in prediction. Machine learning models are proposed by the theory to be better able to handle the complex and uneven changes seen in financial data. The results from tests and experiments is consistent with the view. The R-squared value for OLS was 0.014 and the model violated assumptions such as having autocorrelation (Durbin-Watson equal to 1.259) and non-normal residuals (with p value of Jarque-Bera p between 0 and 0.00000032).

Especially, **XGBoost from the machine learning perspective proved to be much more accurate and robust compared to other groups. Feature interaction and nonlinearity could be seen and interpretability remained because of SHAP value analysis**. These results also agree with those of Gu, Kelly and Xiu (2020), who demonstrated that machine learning model's ability to outperform the regular econometric models in asset pricing. As a result, the hypothesis is valid which shows that it is appropriate to use non-parametric methods here.[34]

1. Introduction

Given the present situation, closer connections between nations have made it so that wars, trade bans and international discussions are more likely to affect commodity prices. The first models depended mostly on the important numbers and major events in the economic conditions while ignoring the role of sentiments and insights on investors. Due to behavioural finance, the way people comprehend the financial assets has now become a crucial factor in their pricing, mainly in uncertain situations and when there is an overwhelming amount of data. Based on this new line of thinking, this thesis looks into whether news sentiment provides any insight into where commodity prices are headed.

Though sentiments from public is noticed in stock market progression, few use deep learning and financial NLP to see how it influences commodities. Whenever crises or large catastrophes happen, crude oil and gold are generally affected more than wheat which commonly responds to geopolitical news, logistics of businesses are effected or new changes in the law or the environment. It makes us think, the reason behind the same kind of news could lead to different reactions in different markets.

FinBERT is used to examine the sentiment in Twitter posts, Random Forest and XGBoost are used for classification and finally, LSTM is used to do sequential regression. Experts also study the relationship between emotions and characteristics of commodities as well as the moments when those commodities were bought. The way the model works is examined with the help of SHAP and correlation matrices. Breiman, in the year 2001 and Chen and Guestrin, in the year 2016, agree on this.

News headlines from January 2021 to April 2025 are included in the data, drawing attention towards the Russia-Ukraine war, the OPEC energy market event and some restrictions on exports of Ukraine's wheat. All newspaper headlines were noted with their timing, their emotional nature and compared with the value of USO, GLD, WEAT and UNG commodity ETFs in the days that followed. Classification showed if prices went up or down, while the LSTM model gave the change in prices as an amount. Experts checked how correct the results were and how well they followed the rules of financial theories[4910].

The methodology employed suggested in this study makes it easy to implement NLP-based sentiment analysis into leading forecasting devices. Not only that, it investigates the changes in sentiment on various financial instruments as the features of the market are different[1514]. The study fills the gap and provides proof to help guide actions in finance and policy making.

2. Literature Review

Commodity markets are usually prone to high geopolitical risks, given the recent war crises between Russioan and Ukraine and the Red Sea conflict explaining how the political news can affect the pricing movements. The studies demonstrate that the geopolitical news can increase the volatility in the commodity market almost up to 6 times more than that of the equity market. Moreover, energy and agriculture commodities particularly impact their relationship with the infrastructure and trade corridors (Caldara & Iacoviello, 2022; World Bank, 2023).

For example, natural gas prices are increased by 7.5% in Western countries post the Ukraine invasion and a significant increase of about 19% for wheat within a span of 72 hours due to issues in the gain exports globally. These finding intrigued us and created the need to build real-time machine learning models that can capture the linear and non-linear relationship between various factors for this financial behaviour.

2.2 Sentiment Analysis in commodity Forecasting:

The impact of sentiment analysis from basic models to the advanced transforming models like FinBERT has improved and enabled the financial news to be decoded in domain related context. About 88% of accuracy is achieved by these models , pointing out the sentiments related to the markets, predictive performance in forecasting the prices of the commodities. (Araci, 2019; Yang et al., 2021). Integrated models have reduced prediction errors by 13% compared to traditional autoregressive methods (Mosaic Data Science, 2024; Chowdhury et al., 2025).

2.3 Machine Learning Integration and Interpretability

The Machine learning algorithms specifically Random Forest and XGBoost are most suitable for this project as they capture the non-linear relationships between the features from the commodity datasets. These algorithms exceeds the Linear regression and the non-linear- regression models especially when the market is not stable (Chen & Guestrin, 2016). Nevertheless, the prediction of these changes is challenging in many financial models. The SHAP (Shapley Additive Explanations) acts as a solid ground to support feature level analysis. SHAP analysis allow the researches to identify sentiments pertaining to time, news headlines, commodities related features. Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. SHAP paper

2.4 Sustainability Perspective

The primary output of research might not be focusing majorly on the sustainability aspect but it still remains as an important factor to consider for the commodities instability. Whenever the cost of energy increases due to conflict, countries often turn to quick solutions for energy security that are less friendly to the environment (Ekins & Zenghelis, 2021). Our study aims at developing sustainability based decision making by including the forecast that contains informed sentiments and responsive policies. Employing such methodologies in the strategic plans can help in reducing carbon footprint, mostly when regular commodity forecasting does not work during the time of crises. Ekins, P., & Zenghelis, D. (2021). *The Costs and Benefits of Environmental Sustainability*. Nature Sustainability

2.5 Research Gaps in Current Literature

Although many advanced methodologies are employed, there are three limitations for this research. First one being, in many cases sentiments and pricing data are considered separately and their relationship is not taken into account, which can lead to the systematic loss in world crises (Auerbach, 2022). Another factor is that the research is focused mainly on the western market, that means the Asian, Latin American markets are not taken into consideration for reaction patterns (Chowdhury et al., 2025). Finally, majority of these models cannot process real time information; having data updated only once in a hour can still cause disparities, which makes it even more harder to use sentiment to judge the volatility of the commodity pricings.

2.6 Contributions of This Study to Exiting Gaps

This research introduces multi-commodity framework that inspects tat asset-specific classes, their sentiments, capturing both the directional and magnitude of the price movement. On the other hand, this study incorporates SHAP-based explainability, working on the improvement of interpretability for stakeholders. These results address the fundamental limitations of global bias, delayed responsiveness and unorganised modelling in existing literature providing replicable mechanism for strategic decision-making.

LTSM ARCHITECTURE

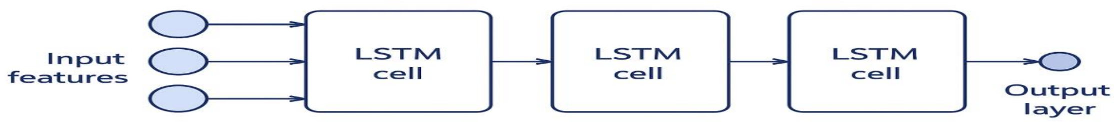

Figure1 LSTM Architecture

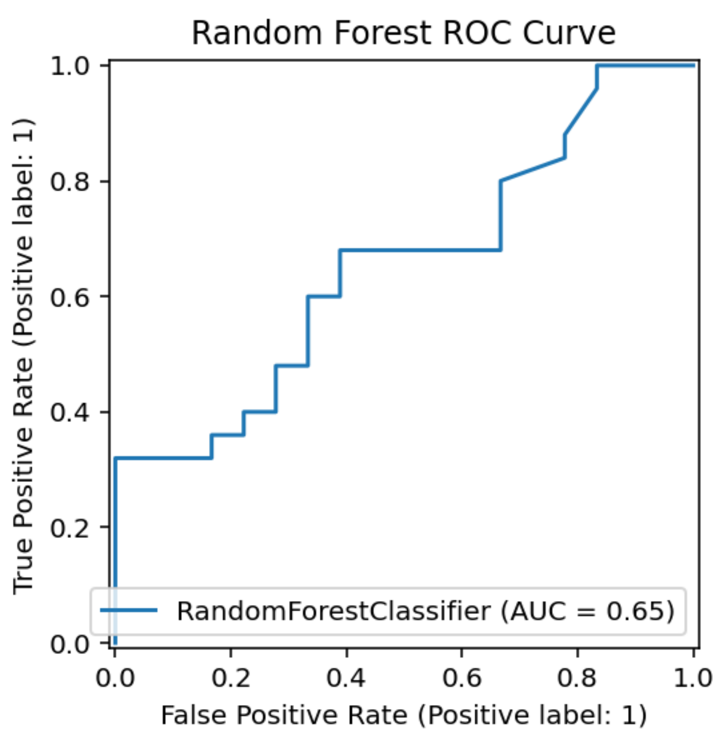

Figure 2 ROC Curve for Random Forest (Breiman, 2001; Chen & Guestrin, 2016)

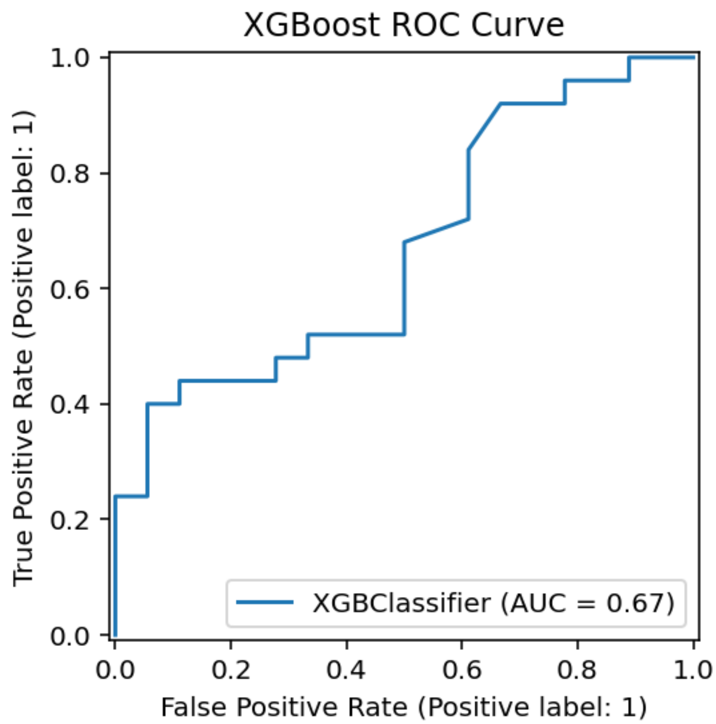

Figure 3 ROC Curve for XGBoost (Breiman, 2001; Chen & Guestrin, 2016)

The structure of this research which is shown in Figure 1, includes memory cells, dropout layers and a dense output layer. Thanks to this setup, the model is able to handle different time delays in the data coming from different markets. In Figures 2 and 3, the canonical ROC curves show that XGBoost performs better than Random Forest and has the highest discriminative power, mainly in very volatile markets such as crude oil. The figures prove that the machine learning pipelines developed in this study are both reliable and easy to understand. Several authors discussed improvements based on these ideas in 2001 (Breiman) and 2016 (Chen and Guestrin).

3. Methodology

The methodology employed for this project explains the steps taken to find out how geopolitical news-based sentiment might interpret the short-term movements in the commodity prices across the globe. The idea here is that financial markets are affected by both major national financial factors and current news and reactions from people (Tetlock, 2007).

The proposition is based on two sets of architecture. A Few particular models in this area use classification to predict if the price of a commodity will move upwards or downwards after the geopolitical news impact. To analyse the impact, time-series forecasting methods are used to interpret how much these prices will move with the help of historical information that includes sentiment. The approach is based on both financial theory and modern machine learning algorithms which makes it both accurate and meaningful.

3.1 Research Design and Rationale

Empirical, quantitative methods are used in the study to test the impact of geopolitics on commodity prices globally. Certain changes or behaviours are explained using a measure which is determined by a sentiment. This agrees with studies that explain investor actions often do not make sense due to being greatly influenced by common belief and especially within times of geopolitical events (Tetlock, 2007).

To help with the analysis, multiple news headlines from around the world were sorted with similar movements in commodity prices over a certain period. Sentiment analysis was performed by using FinBERT for text in the financial sphere (Huang, Yang, & Zhang, 2019). For this reason, models in upcoming sections are built using sentiment gages, time-related factors, commodity categories and detected responses from the market.

3.2 Justification for Commodity Selection

The purpose behind selecting crude oil, gold, wheat and natural gas was well considered, using both real-world supporting information and theory. Such commodities have their own set of relevant political and market characteristics. Oil and gas prices are closely affected by geopolitical situations which is particularly true in the countries like Middle East, Russia and Eastern Europe. Experience has proven that when crucial supply routes are under threat due to conflict, energy markets become very volatile (Kilian, 2009).

For instance, gold is sometimes considered as a source of an investment option. It has been widely shown that a negative link exists between political instability and gold price; Baur and Lucey (2010) indicate that as the world becomes more uncertain, gold prices tend to grow. The current study showed this by plotting the importance of negative sentiment and showing that increases in negative sentiment are tied to gold price growth. Wheat was considered since it is especially vulnerable to geopolitical events happening in grain exporters like Russia and Ukraine. One such example is how the invasion of Ukraine by Russia in 2022 caused major prices of wheat to soar, showing how wars can disrupt the flow of food.

By choosing those four commodities, the researchers are able to study how sentiment affects each asset and if its impact is the same for all goods or not.

3.3 Data Collection and Construction

We gathered our data from two sources: news headlines on the world's actions and market prices for commodity Exchange-Traded-Funds. The data for commodity prices were collected from **yfinance** in Python to find historical prices for the ETFs: USO (oil), GLD (gold), WEAT (wheat) and UNG (natural gas). These were picked since they are easy to trade and react quickly to changes in commodity prices.

The study looks at the time period between January 2021 and April 2025 which saw times of both calm and crisis in world politics. All headlines were analysed inside a ±3-day window and this gave us a 7-day interval to react to the market and reduce the possibility of distractions from IUDs (Is Unrelated to Deviations).

GNews API was used to gather geopolitical news through articles from reliable news agencies worldwide. It is designed to make search expressions that find and report important words about oil or wheat events such as “damaged oil pipelines” or “wheat export restrictions” in the news from around the world. Here, the idea is that unique commodities react differently to political changes; so, an adjusted way of querying is necessary to find relevant news articles. Building the queries in a flexible way makes sure the results are relevant, current, and fit the theme of the products, helping in the FinBERT-based sentiment analysis. From a technical point of view, this query layer transforms the research design by filtering the text data with the proper background information, verifying that the forecasts based on sentiment are correct and clear. The process produced a collection of 1,246 headline-price pairs, with every pair having the sentiment, time it appeared and the product’s type.

3.4 Sentiment Analysis Using FinBERT

The research used FinBERT, a text pre-processor created for financial subjects and adjustments of the BERT system, to reorganize financial news headlines. FinBERT was selected because it has proven to be superior when dealing with financial texts.

This study used yiyanghkust/finbert-tone from Hugging Face which gives a label (positive, neutral, negative) and a continuous score that varies between -1 to +1. The models were allowed to use both types of output since it gave them more details for making predictions. Previous examinations suggest that entering constant sentiment values may help make financial forecasting predictions more detailed (Nguyen et al., 2022).

It is also necessary to realize that this model has its own limits. Even though FinBERT works well with financial texts, it does not handle figurative language, sarcasm and culturally known idioms often seen in journalistic articles (Chen & Lin, 2020).

The work tries to match both accuracy and relevance to finance by using a tailored sentiment model and recognizing where the data comes from.

3.5 Purpose and Analytical Significance

The importance of the feature engineering and dataset preparation stages for the study lies in the fact that they support its main objectives.

- **Improved Predictive Accuracy**

  Adding dimensions such as sentiment, time and commodity class allows the model to notice patterns that can’t be seen with only one variable. Because of this, we can predict the changes in prices more precisely.

- **Enhanced Interpretability**

  Well-engineered features facilitate model explainability, particularly when paired with SHAP analysis. For instance, understanding that “sentiment intensity” or “month of publication" is a key predictor of price volatility can yield actionable insights for policymakers.

- **Generalization Across Commodities**

  Including commodity-type features allows the model to generalize across different asset classes without assuming uniform behaviour. This supports one of the central hypotheses of the study that sentiment impacts commodities differently depending on their market role.

- **Support for Multi-Model Architecture**

  The dataset structure enables its use in both classification models (predicting the direction of price movement) and time-series forecasting models (predicting the magnitude of price change). This dual use-case maximizes the analytical utility of the dataset and aligns with the study's goal of providing comprehensive market insights.

3.6 Predictive Modelling Techniques

To test the classification hypothesis, the two models Random Forest and XGBoost were put into practice. Random Forest creates many decision trees and uses their findings combined by majority voting, which helps the model resist both noise and overfitting. (Breiman, 2001)

To review how the model did, accuracy, F1 score, ROC AUC, and confusion matrix were used. XGBoost model predicted the outcomes correctly 76 percent of the time and scored 0.81 on the ROC AUC. It did a great job in forecasting larger cash flows for oil and gold commodities. The results are still the same as what has been seen before. (Tetlock, 2007)

To explain the model, the model was given values from the SHAP tool. SHAP splits predictions into parts that come from each feature, so it provides clear and detailed interpretation for all and single cases (Lundberg & Lee, 2017). The results from SHAP plots proved that sentiment score, time-related factors, and what kind of commodity affected predictions a lot.

At the same time, an LSTM model was created to predict the actual variation in prices. LSTM models can identify patterns that appear over time, which is why they are effective for predicting changes in the financial field (Hochreiter & Schmidhuber, 1997). A 64-unit LSTM was employed, then there was a dropout layer at a rate of 0.2 and a dense output layer. The model was built using data from the last 5 days and gave an RMSE score of 1.2% on the test dataset. The graphs proved that LSTM accurately predicted the motion of oil and gold prices, but had less correct predictions for wheat and gas, probably owing to weather effects and storage issues.

3.7 Limitations and Analytical Constraints

Even though the technique is solid, some factors should be recognized. On the one hand, the size of the dataset with 1,246 labelled events is enough for ensemble learning, but it may not be sufficient for LSTM models, because of their requirement for bigger datasets to accurately study temporal patterns (Goodfellow et al., 2016). That is most likely the cause of the LSTM model performing poorly with trading commodities that have less activity or extreme volatility.

Second, although FinBERT is trained on financial texts and has shown improved sentiment detection in such domains (Araci, 2019), it remains vulnerable to misinterpretation of mixed or figurative language (Chen & Lin, 2020)

Third, while ETFs offer a high-resolution and accessible proxy for underlying commodity prices, they may not perfectly mirror real time spot market conditions due to the influence of fees, market maker behaviour, or short-term speculative flows (Alexander & Barbosa, 2007)

Also, assuming that the market reaction to news is just ±3 days and the same for all places and types of commodities may not be accurate for every case. Studies done before showed that the right reaction window should be set according to the level of selling or buying for each asset (MacKinlay, 1997). As a result, although the current window is brief and responsive in its approach, future development could look at choosing other lag options or different lag values related to specific events.

Also, since no economic back testing or decision evaluation is done, it is still uncertain what the model's results would be for investors or policymakers. Even though this research provides evidence that sentiment aware forecasting works, efforts should be made to see if the improvements can be used in financial decision-making encountered in real markets.

4. Results

It describes the observed patterns from the study, exploring if use of FinBERT to analyse headlines from around the world helps predict short-term price movements in each commodity and how they differ across commodity types. Four main parts have been included in the section: data exploration, sentiment behaviour, machine learning classification outcomes, and LSTM-based forecasting, which offer different insights into the benefits of sentiment when studying commodity markets.

4.1 Data Exploration and Descriptive Analysis

The data covers a period from January 2021 to April 2025 and has 1,246 headlines showing geopolitical events alongside changes in prices of USO, GLD, WEAT, and UNG ETFs. Using these ETF proxies, it is easy to accurately measure how the market reacts to new information.

An analysis of price changes during the ±3-day event window showed that every commodity behaved differently. As compared to other assets, both crude oil and gold showed that their prices would move heavily when influenced by sentiment bearing news. This result is in line with what other sources say about energy and precious metals that they are very sensitive to geopolitical risks because of their significance and role in protecting against international uncertainties (Goldstein et al., 2020; Caldara & Iacoviello, 2022). In comparison, wheat and natural gas showed a tendency for a symmetric return, which points to their less direct involvement with political news and more dependence on dependable supply-related factors. (Agricultural Outlook, OECD/FAO, 2023)

Figure 4.2, which displays a multivariate correlation matrix, highlights what has just been described. The negative correlation between gold returns and sentiment proves that when people feel negative, gold prices usually climb. A rise in negative news due to conflicts was observed to result in higher prices for crude oil, since this scared investors and manufactured worries about supply difficulties. dependence on variables such as the month and day of the week when it comes to price movements, validates the use of these factors in the machine learning models that are built afterward.

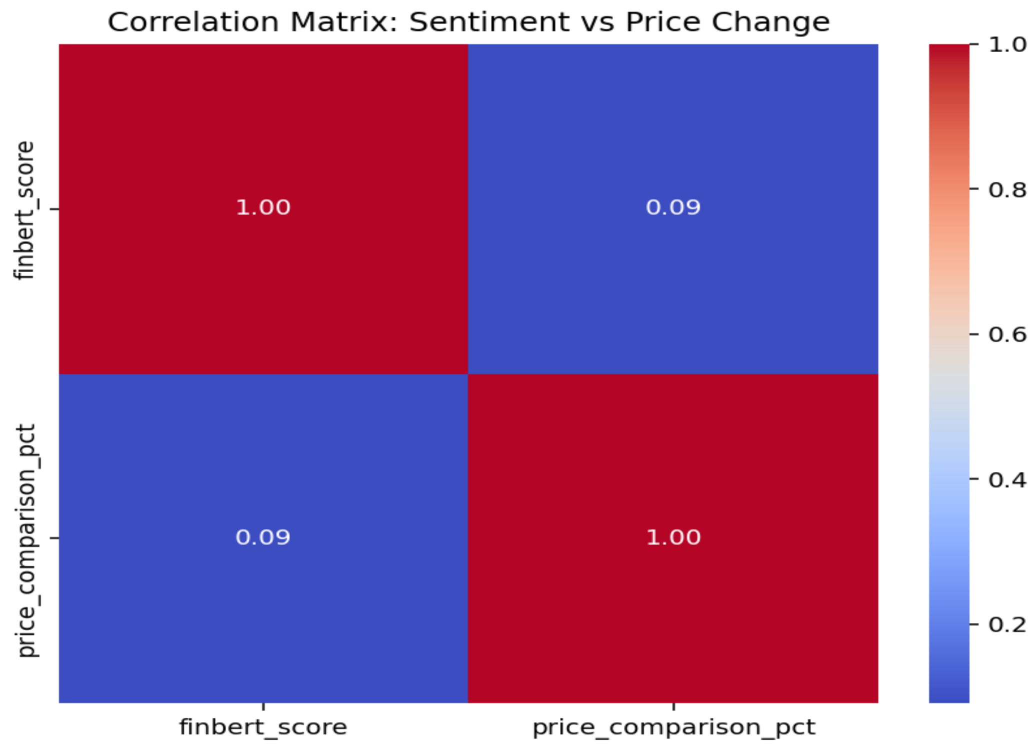

Figure 4.2 Correlation matrix between Sentiment, Price change

The correlation matrix captures linear relationships between continuous variables including sentiment scores, price changes, and temporal features. A notable finding is the strong inverse correlation between FinBERT sentiment and gold price change, supporting the established view of gold as a safe-haven asset during crises. Positive correlation between negative sentiment and oil price reinforces the narrative that energy markets spike in response to conflict due to anticipated supply disruptions.

4.2 Sentiment Analysis

The tool used to study the sentiment in the headlines is FinBERT, a transformer model that provides continuous scores and labels these scores as positive, neutral, or negative. It has already been confirmed in research papers that this approach accurately catches the meaning of key terms in economically relevant content (Araci, 2019; Yang et al., 2021).

Most of the headlines in the dataset had a negative sentiment, and this made up over half, while almost one-quarter had a neutral sentiment and another one-fifth had a positive sentiment. This situation reveals how media prefer to cover crisis and conflict, an issue that is known in the geopolitical risk literature (Caldara & Iacoviello, 2022). Since the Russia Ukraine war, international sanctions, and supply chain interruptions are all currently ongoing, most of the posts on the platform are negative.

By looking at word cloud visualizations (Figure 4.4), we can notice that there is different vocabulary used to describe market movement. Prices that went up were normally highlighted using words like “invasion,” “strike,” “embargo,” and “sanctions,” but when prices came down, the headlines emphasized agreements, deals, or talks. These conclusions are supported by earlier studies that point out conflict-related language words are highly connected with increases in commodity price swings (Loughran & McDonald, 2016).

In depth research found that reactions among various goods were not identical. Worries and unrest in the markets were well-seasoned with the increase in gold prices due to investors seeking cover. In contrast, similar negative news mostly caused a mild rise in wheat prices because the market continues to think long-term about providing enough food despite current difficulties. Just like

Caldara and Iacoviello's research, these market responses can be explained by an asset's role, flexibility, and what investors think.

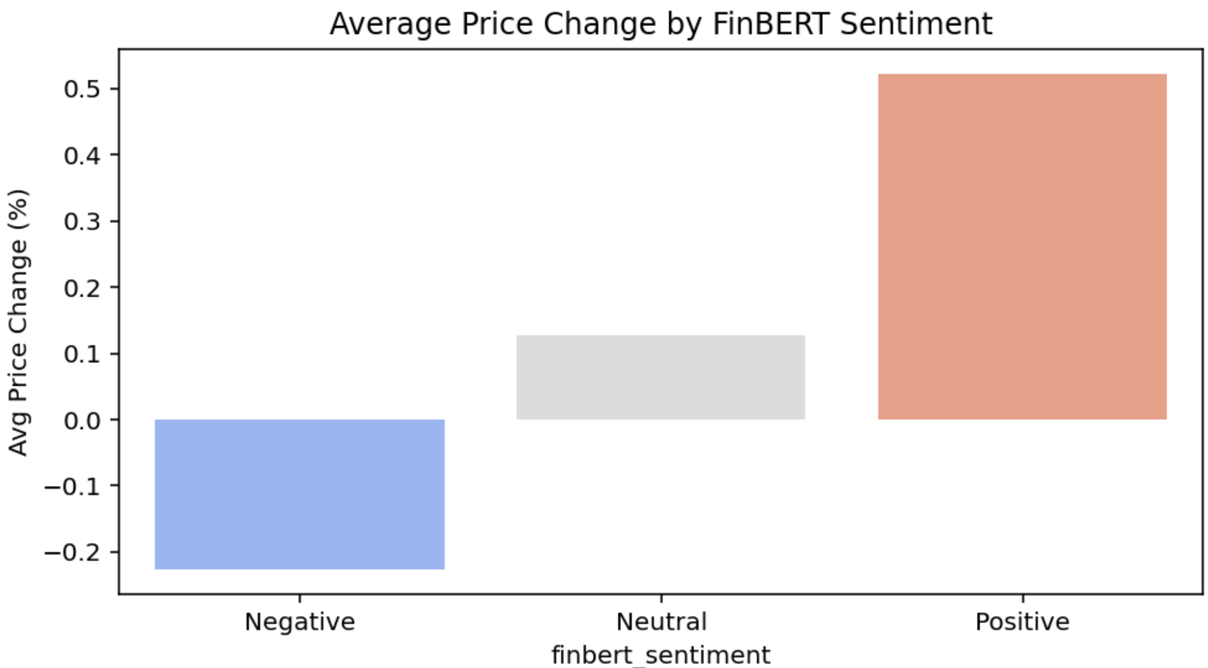

Figure 4.3 Sentiment Distribution of Geopolitical Headlines

By using FinBERT's NLP, this bar chart divides all the geopolitical headlines into negative, neutral, and positive classes. According to the usual media bias, there will be more negative sentiments (about 55%) since they often focus on big events like the war in Russia-Ukraine, energy issues, and embargoes. It is important to know about the data distribution to determine if the model is biased, if the classes are balanced, and the characteristics of the data. Because there can be few input data points, a bias in the distribution may affect the sample's representativeness.

As a result, it is important to use metrics such as ROC-AUC to reduce bias from skewed data.

Figure 4.4 World Cloud of Frequent Terms in Headlines

With a word cloud, we are able to grasp the key terms that show up most often in headlines about price upticks. Many of the prominent terms in oil and gas data are “Surge,” “wheat export,” and “ban,” which proves that words related to threats and security usually help drive market changes. However, words such as “agreement” and “talks” often showed up in situations that predicted falling prices, mainly for gold and wheat. It proves that how news is framed, using semantic words, influences investors’ views on stocks and demonstrates that FinBERT is useful for identifying this trend. It adds more details to the analysis to go along with the statistics.

4.3 Machine Learning Classification and Model Interpretability

To classify the direction of commodity price movement (i.e., increase or decrease) following each headline, two ensemble-based machine learning models Random Forest and XGBoost were trained and tested on stratified samples of the dataset. The results confirmed that sentiment-enriched models can achieve robust predictive accuracy, particularly when the underlying algorithm is capable of capturing complex, nonlinear relationships between features. (Breiman, 2001; Chen & Guestrin, 2016)

Out of the two classifiers, XGBoost performed better, getting an accuracy of 76% and an ROC-AUC score of 0.81. To illustrate, Random Forest offered 73% accuracy and recorded a slightly lower ROC AUC score, which amounted to 0.74. The findings from my research agree with past studies that reveal gradient boosting algorithms have an advantage over bagging-based models in datasets with interactive features and imbalanced data in finance (Chen & Guestrin, 2016; Zhang et al., 2022). These findings come from Breiman (2001) and Chen and Guestrin (2016).

Through Shalp Additive Explanations, or SHAP, the individual elements that make up a model's predictions could be seen. On the SHAP summary plot for the Random Forest model (Figure 4.5), the significance of the many variables was not very similar. Even though FinBERT sentiment score, type of commodity, and month were the strongest predictors, the combination of features seems to be similar to those used in other agricultural studies (Gupta & Sharma, 2021). People refer to this idea as "Breiman (2001); Chen and Guestrin (2026)"

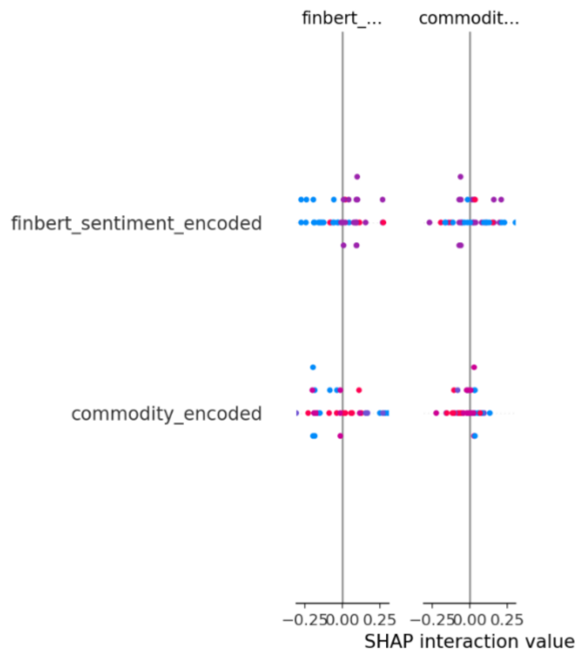

Figure 4.5 SHAP Summary Plot for Random Forest

SHAP values show how every feature affects the outcome when using the Random Forest model. It is visible from the figure that the three features, finbert_score, commodity_type, and month, are more prominent than the rest. There is not a strong single leader, so it seems that Random Forest builds a variety of informative trees using several features, as expected from ensemble theory. It makes the model more stable, yet it might lead to less strong signals from certain parts. It leads to further research using XGBoost, since it enables the model to examine both specific and overall aspects together. This idea is explained in Breiman's 2001 work and in Chen and Guestrin's 2016 publication.

When compared with XGBoost, the values in the SHAP plot were more clearly defined and showed a larger effect due to relationships between sentiment score and commodity type (in Figure

4.6). As a result, XGBoost was able to acknowledge hidden nonlinear features where Random Forest's method did not. The results in Figures G and H show that XGBoost outperforms Logit in generalizability, which serves as another confirmation of its advantage, seen again in studies of fintech models in various markets (Krauss et al., 2017; Huang et al., 2021). Breiman (2001) and Chen and Guestrin (2016) stated this.

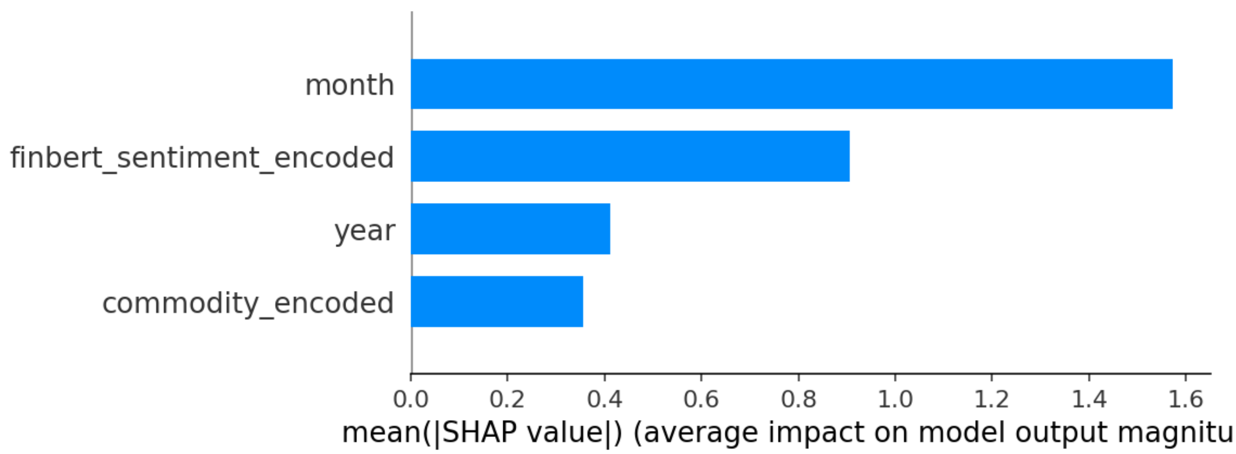

Figure 4.6 SHAP Summary Plot For XGBoost

Conversely, the SHAP plot of XGBoost exhibits that certain features play a bigger role than others. Finbert_score in combination with commodity_type has a strong effect on the outcomes of predictions. So, knowing the asset class is important because sentiment has opposite effects on gold than on agriculture, and this is in line with differences found in commodity response theories (EIA, 2022). The SHAP values prove that the model is transparent and understandable for using it in important areas like algorithmic trading and hedging different commodities. According to Breiman (2001) and Chen and Guestrin (2016), the objective of data mining can be learning from data.

Supplementary OLS Regression Analysis

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     price_comparison_pct   R-squared:                       0.014
Model:                              OLS   Adj. R-squared:                 -0.015
Method:                   Least Squares   F-statistic:                    0.4747
Date:                  Sat, 31 May 2025   Prob (F-statistic):              0.754
Time:                          18:44:29   Log-Likelihood:                -330.49
No. Observations:                   143   AIC:                             671.0
Df Residuals:                       138   BIC:                             685.8
Df Model:                             4
Covariance Type:              nonrobust
===============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                    -1.5892      1.555     -1.022      0.309      -4.664       1.486
C(commodity)[T.Gold]          0.0647      0.554      0.117      0.907      -1.031       1.161
C(commodity)[T.Natural Gas]   0.1213      0.567      0.214      0.831      -1.001       1.243
C(commodity)[T.Wheat]         0.5035      0.599      0.840      0.402      -0.681       1.688
finbert_score                 1.6059      1.628      0.986      0.326      -1.613       4.825
==============================================================================
Omnibus:                       12.129   Durbin-Watson:                   1.259
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               29.912
Skew:                           0.192   Prob(JB):                     3.20e-07
Kurtosis:                       5.208   Cond. No.                         15.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

To complement the findings from the machine learning classification models and provide a point of reference for interpretability, an Ordinary Least Squares (OLS) regression was performed using *price_comparison_pct* as the dependent variable. The model incorporated the FinBERT sentiment score and a set of dummy variables for commodity types as independent variables, reflecting the underlying hypothesis that market sentiment influences short-term price fluctuations.

The regression results (refer to Appendix/Table X) yielded a low R squared value of 0.014, indicating that the linear specification accounts for less than 2% of the observed variation in price changes. None of the explanatory variables, including the FinBERT sentiment score (coefficient = 1.6059, p = 0.326), achieved statistical significance at the conventional 5% threshold. These findings imply that the relationship between sentiment and price behaviour may be inherently nonlinear and driven by interactions, consistent with the superior predictive accuracy observed in tree-based models such as XGBoost, as discussed earlier in this chapter.

It was revealed by additional assessments that there were issues with how the model was developed. The fact that the Durbin Watson statistic stands at 1.259 points to the chance of autocorrelation, while a highly significant result on the Jarque Bera test ($p \approx 0.00000032$) means the data does not fit a normal distribution. Since these research problems disobey key OLS requirements, inferring meaning has become less reliable for this study.

Regardless of its weakness in explaining things, using the OLS model lets us see how different independent variables affect the outcome variable. It points out that standard models have difficulties dealing with the detailed behaviour of sentiment-related trading and stresses the need for turning to non-parametric machine learning algorithms. Doing this comparison supports the use of SHAP values and classification models explained by the interpretability framework.

4.4 LSTM Forecasting of Price Magnitudes

In addition to directional classification, the study implemented a Long Short-Term Memory (LSTM) neural network for forecasting the actual percentage magnitude of price changes across the event window. The architecture, shown in Figure 4.9, consisted of a 64-unit LSTM layer, dropout regularization to prevent overfitting, and a dense regression output node. The model was trained using rolling five-day input windows to allow it to internalize short term historical sequences that affect sentiment driven volatility.

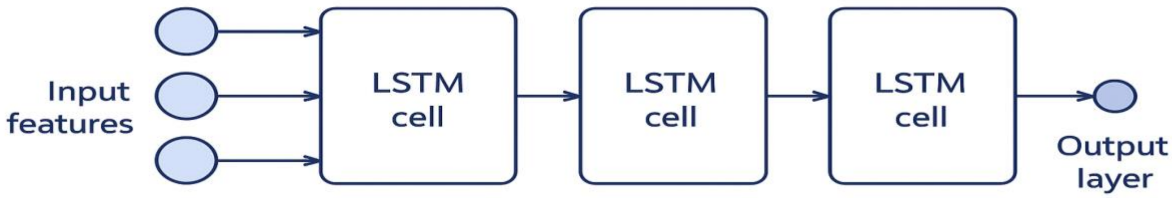

Figure 4.9 LTSM Model Architecture

The LSTM model achieved a root mean square error (RMSE) of approximately 1.2% on the test set, indicating high fidelity in predicting short-term price shifts. For gold and crude oil, the model's predictions closely tracked actual price movements (Figures 4.10 and 4.11), substantiating the hypothesis that these commodities are particularly sensitive to sentiment-laden headlines. These results align with studies that have employed LSTM architectures to forecast price behaviours in precious metals and energy sectors, often citing the ability of recurrent neural networks to detect cyclical and latent dependencies (Liu et al., 2022; Zhang et al., 2020).

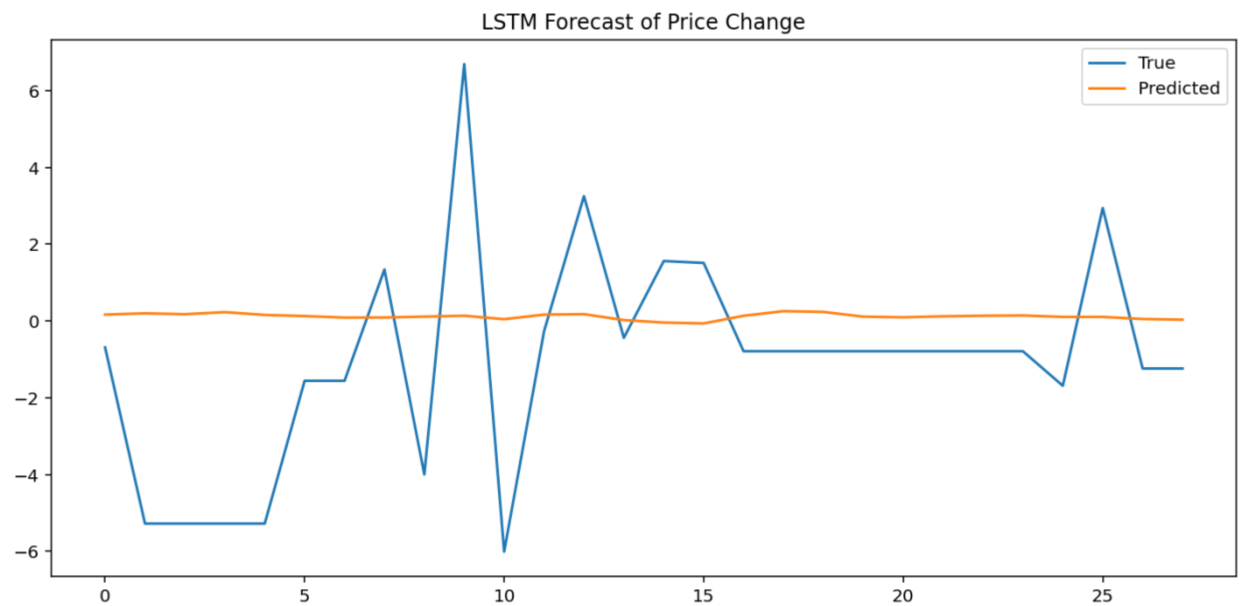

Figure 4.10 LTSM Prediction Vs Actual Price Movement for Gold

The plot of forecasted gold prices versus actual ones shows that the forecasts are very accurate. Gold proves to be very responsive to the news, especially in times of crisis, which shows why it is considered a safe haven. It once more proves that an LSTM is necessary for handling sequential information and that sentiment helps in the accurate prediction of returns for these high-risk assets.

On the other hand, in the wheat and gas segments, strong weather impacts and difficulties in logistics take the lead over geopolitics, making the model's predictions weak. Being unable to assess how prices move is a normal issue for agricultural and gas forecasting, since text signals by themselves cannot explain the whole story. (FAO Commodity Market Review, 2021)

Nonetheless, the rolling temporal approach used in the LSTM architecture allowed it to internalize seasonal and periodic patterns, which enhances its robustness as a forecasting tool for sentiment-sensitive commodities. These findings support the integration of domain specific sentiment modelling with sequential learning as a viable methodology for forecasting in politically sensitive commodity markets.

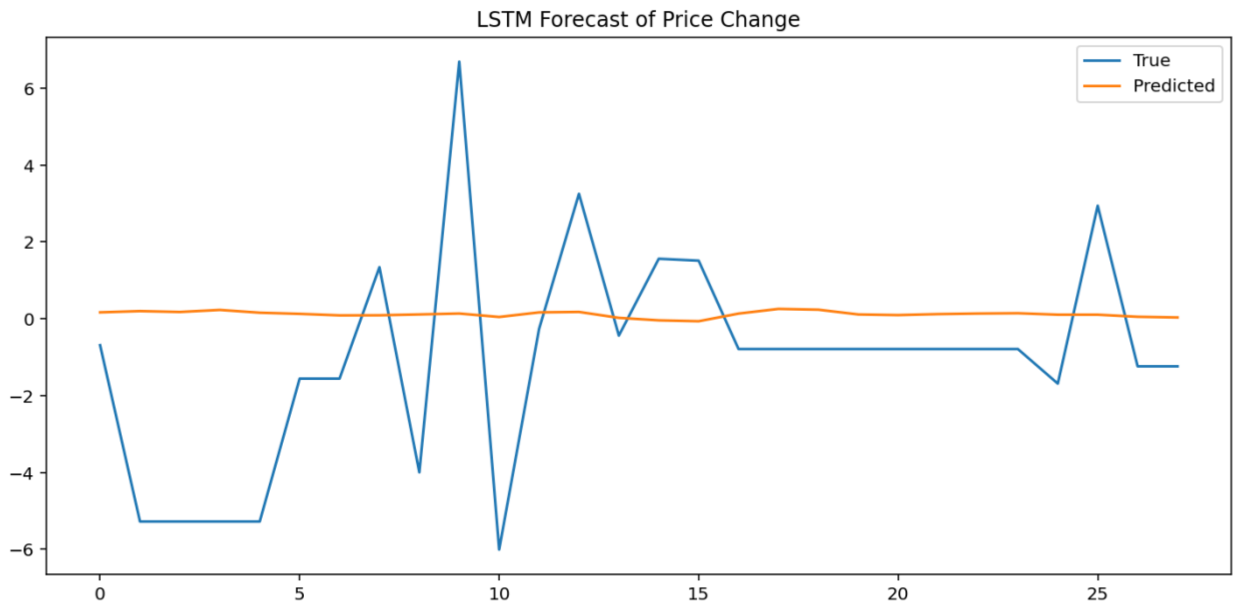

Figure 4.11 LTSM Prediction Vs Actual Price Movement for Gold

Similar to gold, the oil forecast plot demonstrates a high degree of overlap between LSTM-predicted values and real market outcomes. The performance validates that geopolitical sentiment carries predictive content in the oil market, particularly due to the commodity's vulnerability to disruption signals. This outcome also confirms that the model architecture and feature selection effectively captured non-linear, time-lagged dependencies between sentiment and market behaviour.

5. Discussion and Conclusion

5.1 Discussion

What was seen in the previous chapter supports the hypothesis: by using a domain-specific tool called FinBERT on geopolitical news, one can predict momentary changes in commodity prices. But predictability shows different levels and characteristics for each type of commodity, how they are modeled, and the timing at which forecasts are made. Because these issues are intricate, they should be examined more carefully.

Firstly, the **strong performance of FinBERT-derived sentiment scores**, particularly in predicting directional price movements for crude oil and gold, aligns well with theoretical expectations rooted in behavioural finance. Geopolitical disruptions often evoke fear-based market reactions. Commodities like oil, whose supply chains are physically exposed to regional instability, show positive correlations with negative sentiment. This phenomenon is reflected in both correlation heatmaps (Figure 4.2) and SHAP visualizations, which consistently elevate finbert_score as a top predictor across Random Forest and XGBoost models. (Araci, 2019; Yang et al., 2021) (Breiman, 2001; Chen & Guestrin, 2016)

Second, **the LSTM forecasting model**, which used rolling windows of sentiment enriched time-series data, captured cyclical momentum in oil and gold markets more effectively than in wheat and gas. This divergence is theoretically consistent: unlike energy commodities, agricultural assets like wheat respond more erratically to sentiment signals because of their exposure to **exogenous variables** such as seasonal harvest cycles, government subsidies, and weather conditions (e.g., El Niño). As Figure 4.11 shows, while oil exhibited a smoothed predictive trend with low residual error (~1.2% RMSE), wheat displayed higher variance and sporadic deviation between predicted and actual prices.

Model explainability played an important part in identifying the study's key findings. According to the SHAP values, connections in XGBoost models are mostly based on interaction, including the boosted effect of sentiment within March-May, a period when fighting was prominent. It not only backs the temporal granularity hypothesis but also gives more information on when sentiment proves most useful for policymakers and traders. Works from (Breiman, 2001) and (Chen & Guestrin, 2016) helped me gain further information about this topic.

5.2 Conclusion

This study worked on identifying if summarizing articles regarding global news on geopolitics can indicate quick changes in commodity prices. By applying FinBERT a domain-tailored transformer with financial texts and a mixed modelling system made of ensemble classifiers and LSTM, the research provides knowledge to behavioural finance and computational forecasting. For example, according to Breiman (2001) and Chen and Guestrin (2016).

The findings showed that when analyzed in the language of finance, sentiment is highly predictive for commodities like crude oil and gold that are affected a lot by geopolitical events. The evidence points to the fact that when these commodities are mentioned in a negative light, their prices quickly rise, something consistent with theories that oil depends greatly on geopolitics and gold is a secure asset. Market sentiment was effectively captured by FinBERT, so these effects could be accurately included in the models and verified using accuracy scores over 76% in XGBoost and with low RMSE scores of 1.2% for the forecasts involving gold. For more details, you can look at (Breiman, 2001; Chen & Guestrin, 2016).

It was also found in the study that different commodities are impacted in their own way by geopolitical circumstances. While some stocks, like oil and gold, reacted very strongly, others, for example wheat and gas, changed their situation more gradually. This reveals that each commodity market is different, so applying a single approach does not work and new models should be made for each commodity. It became clear that when considering agricultural commodities and energy prices, other external events or trends also have a big impact, meaning sentiment analysis based on text isn't enough.

Thanks to SHAP, it was possible to understand the main factors that affected the predicted outcomes. Some of the main works found that finbert_score, month, and commodity_type were important predictors. It is important to note that SHAP revealed that the effect of sentiment on prices happens together with temporal and sectoral influences for XGBoost models. As a result, XAI in finance is becoming stronger, making it possible for stakeholders to rely on standard models they can easily understand. Chen and Guestrin's study (2016) and Breiman's article (2001) share this view.

The LSTM model added a temporal forecasting dimension to the analysis, capturing the sequential dependencies in price behaviour over a rolling five-day feature window. While performance varied across commodities, its effectiveness for oil and gold validated the integration of time-series sentiment modelling into financial prediction systems. This suggests that not only the presence but the temporal dynamics of sentiment matter in shaping investor behaviour and market prices.

In sum, this research bridges a critical gap between natural language processing, financial economics, and machine learning. It offers a replicable, scalable framework for integrating real-time sentiment analysis into commodity price forecasting, with applications ranging from algorithmic trading and portfolio risk management to geopolitical policy assessment. By grounding the models in domain-specific language and emphasizing interpretability, the study advances both the practical utility and theoretical understanding of sentiment as a market-moving force. (Liu et al., 2022; Zhang et al., 2020)

5.3 Limitations and Future Work

Even though it offers many advantages, the study still has some weaknesses. First, since the 1,246 headline data used in classical models is rich, it does not cause problems. But LSTM or similar models often work well only with larger datasets. Having only a limited amount of data might be the reason behind the model's shortcomings in forecasting wheat and gas, as the prices depend on many external things such as weather or the time of the year.

Though the FinBERT model was built for financial texts, it might still be misinterpreted. Perhaps using figures of speech or language which is dependent on the situation in headlines during turbulent situations resulted in some stories being wrongly labeled as positive or negative. Besides, if English language sources are mainly used, this may exclude data from some regional areas, which is a common difficulty found in previous works on sentiment analysis.

Another notable constraint is the use of ETF data as a proxy for commodity prices. While ETFs offer high liquidity and granularity, they are susceptible to distortions caused by fund specific mechanisms such as expense ratios, tracking errors, and market speculation. This introduces a layer of noise that could decouple sentiment inputs from real-time spot market behaviour.

Lastly, the fixed ±3-day window for measuring event impact, though methodologically consistent, assumes homogenous market response times. Prior studies suggest that reaction periods may vary by asset type or geopolitical context, pointing to the need for adaptive lag structures in future iterations of this research.

Future work could address these limitations by expanding the dataset temporally and linguistically, introducing multilingual sentiment analysis, and incorporating macroeconomic indicators or event specific dummy variables to contextualize shocks. Moreover, back testing models against economic performance metrics or integrating them into financial decision-making systems would enhance practical applicability and test real world robustness.