

Geopolitical Event Sentiment Analysis & Commodity Pricing Prediction

MADHURI PANCHUMARTI
HARDIK SETHI

INTRODUCTION

Motivation: Understanding how geopolitical disruptions affect global commodity markets

Commodity Focus: Crude Oil, Gold, Wheat, and Natural Gas

Objective: Integrate financial NLP with ML for better market insights



Research Questions & Hypotheses

RQ1: Do geopolitical events impact commodity prices?

Hypothesis: Yes, verified as prices tend to increase with conflict due to supply concerns.

RQ2: Does sentiment from financial news influence short-term price movement?

Hypothesis: Partially true, supported in non-linear models like XGBoost.

RQ3: Are machine learning models better than linear regression for forecasting?

Hypothesis: Yes, ML models captured complex interactions and outperformed OLS.



Data Collection and Preprocessing

- Time Frame: Jan 2021 – Apr 2025
- Data Sources: GNews API (headlines), Yahoo Finance (ETF prices)
- Queries: Focused on keywords like 'pipeline explosion', 'export ban'



- Preprocessing: +3/-3 day window alignment, headline-token cleanup
- Sentiment Scoring: FinBERT model from HuggingFace, scores in $[-1, 1]$

OLS with price_change as dependent variable

- $R^2 = 0.014$, p-value of sentiment = 0.326 → Not statistically significant
- Conclusion: Sentiment-price relationship not well captured linearly
- Durbin-Watson = 1.259 → Autocorrelation present
- Jarque-Bera $p \approx 0.00000032$ → Residuals not normal



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			

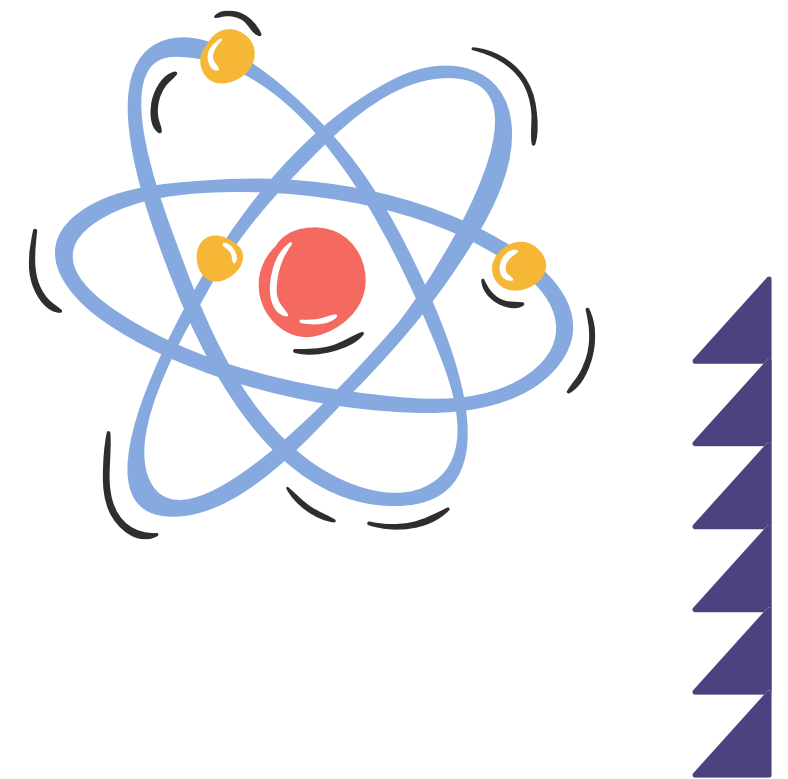
Notes:

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

ML Models: XGBoost & Random Forest

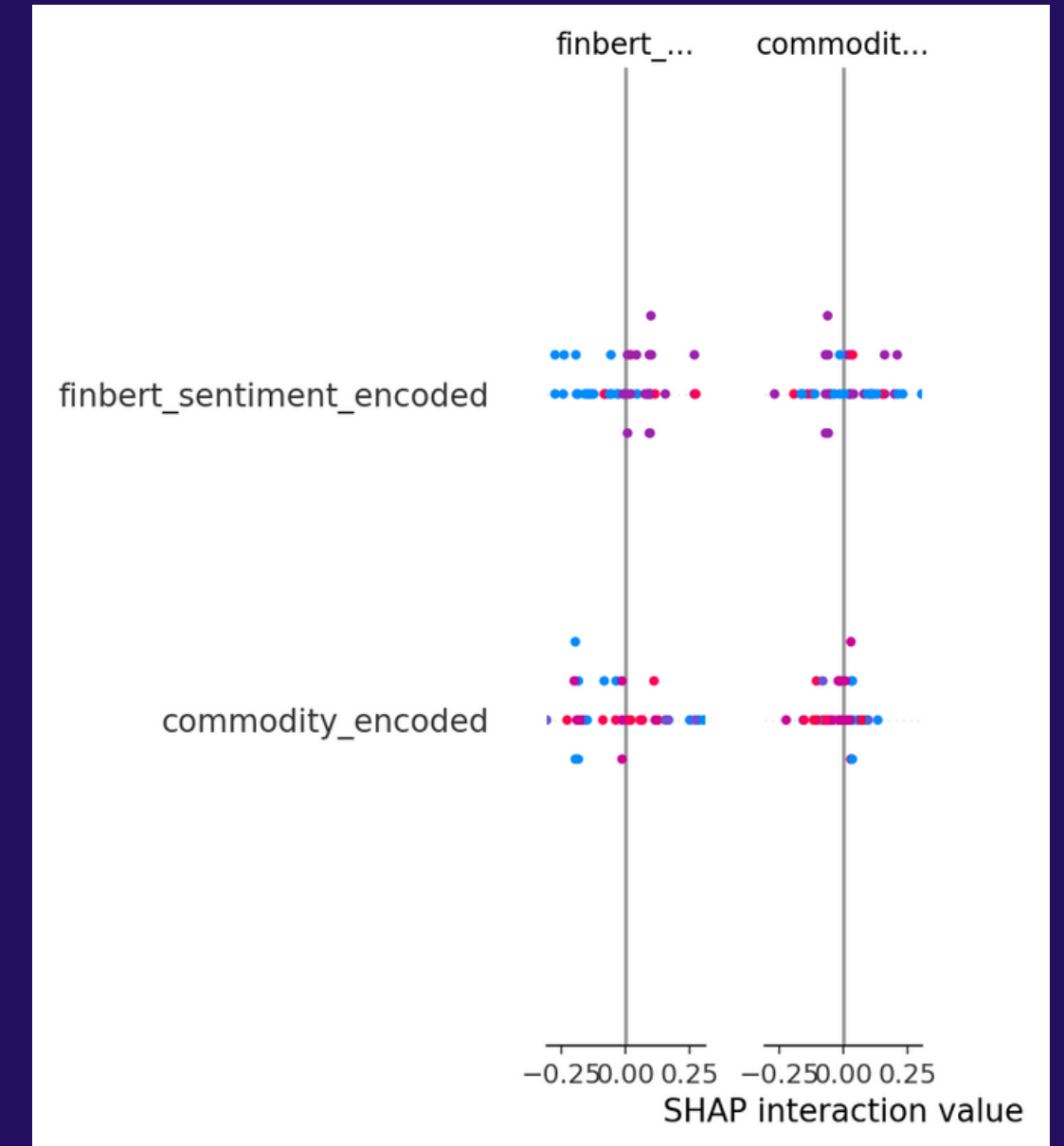
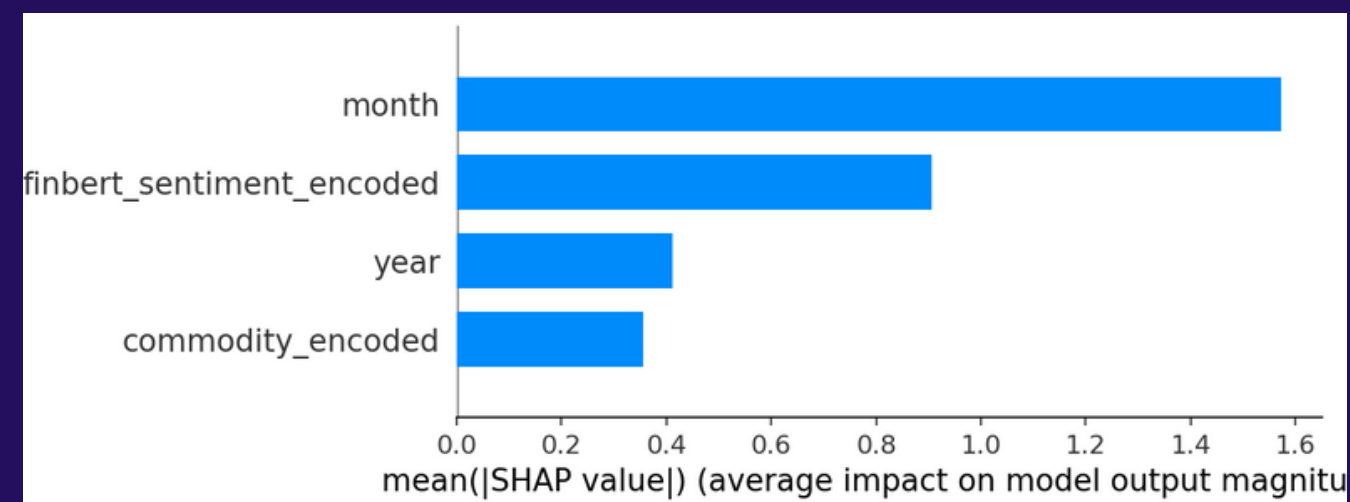


- Target: Predict price direction (up/down)
- Random Forest: Accuracy = 73%, ROC AUC = 0.74
- XGBoost: Accuracy = 76%, ROC AUC = 0.81
- Why ML? Captures non-linearities and interactions
- Features: Sentiment score, commodity type, temporal features



SHAP Interpretation

- Purpose: Explain model predictions feature-by-feature
- Top Contributors: FinBERT score, commodity type, month
- XGBoost SHAP values showed stronger influence than Random Forest
- Helps in understanding model behavior for policy makers and traders



Forecasting Price Magnitudes: LSTM



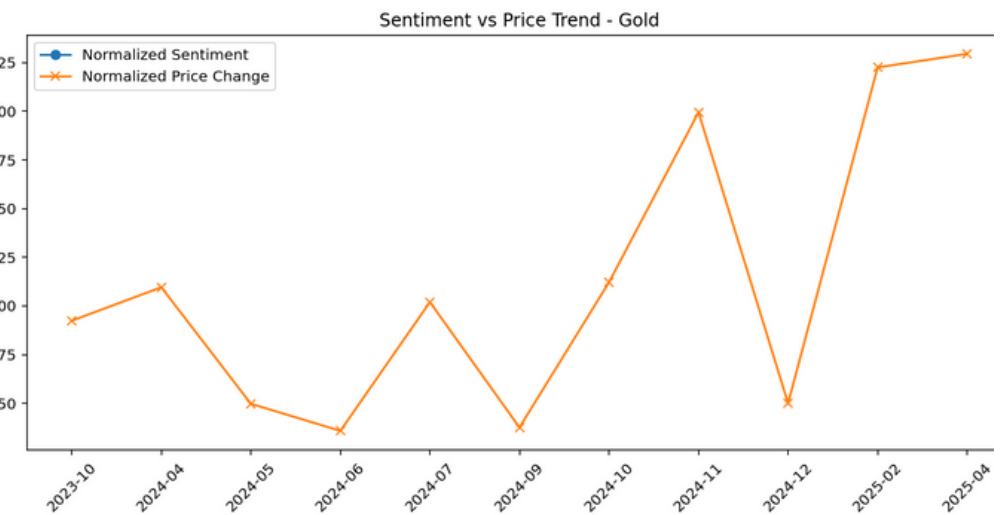
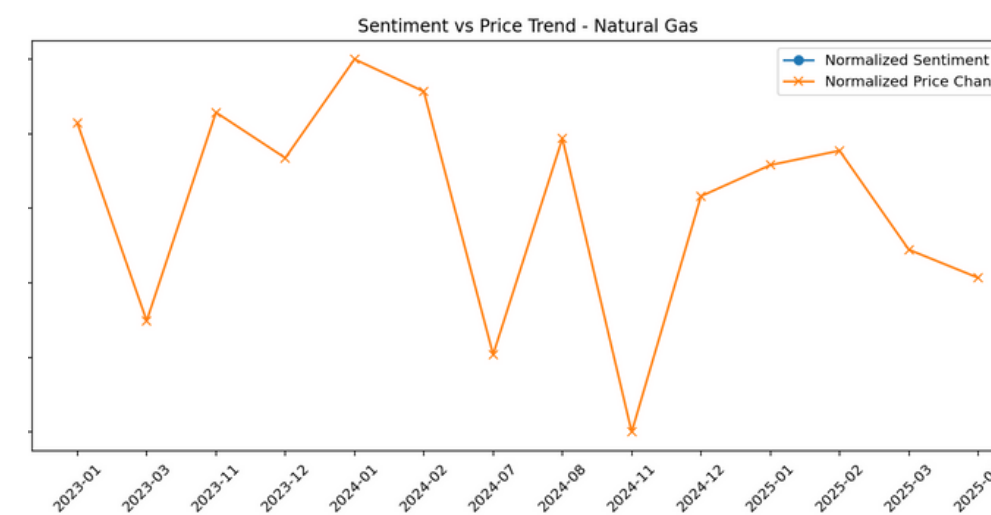
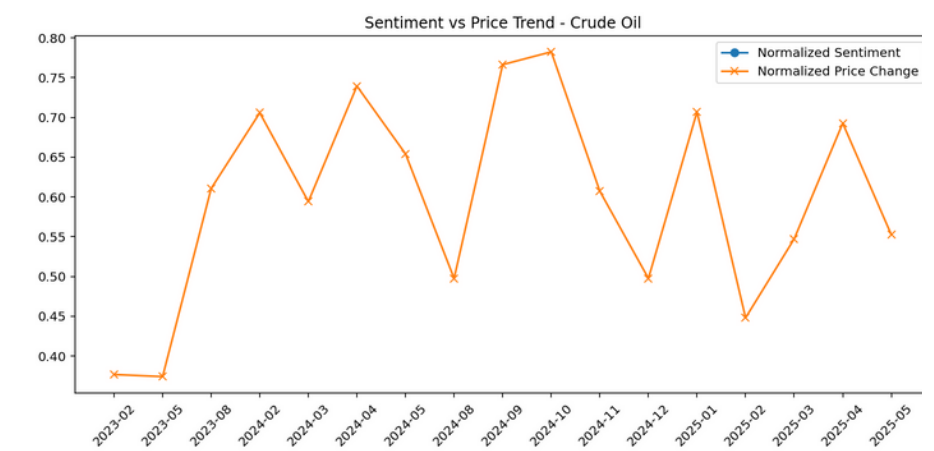
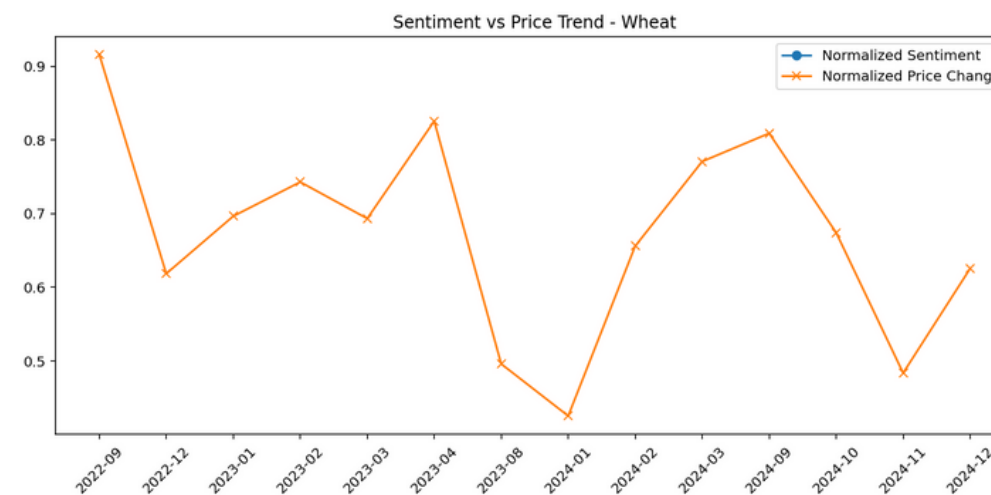
- Model: LSTM with 64-unit hidden layer and 5-day rolling window
- RMSE: ~1.2% for gold and oil
- Architecture: Includes dropout layer to prevent overfitting
- Wheat and gas show less predictability due to external factors
- Sequential dependencies captured better in stable commodities



Commodity-wise Forecasting Results



- Gold: High predictability, reacts positively to negative sentiment
- Crude Oil: Price spikes during conflict, strong sentiment impact
- Wheat: Mixed results due to supply-chain complexity
- Natural Gas: Noisy and difficult to forecast due to storage/weather



Results and Summary

- Geopolitical sentiment impacts commodity pricing
- Gold & oil most responsive to negative news
- ML models outperform traditional regression
- Sentiment and time-based features improve accuracy

Limitations & Future Work

- LSTM performance limited by small dataset size
- FinBERT may misinterpret sarcasm/figurative speech
- ◆◆◆ • ETF proxies may differ from spot prices ◆◆◆
- Future: multilingual NLP, macroeconomic variables, adaptive windows

Conclusion & Impact

- Introduced replicable ML-NLP pipeline for forecasting
- Explained sentiment impact on commodity markets
- Useful for traders, policy makers, and risk managers
- Bridges behavioral finance with interpretable AI models



Thank You

