Oil Price Forecasting

Leveraging Machine Learning for Accurate Predictions

Project Overview

- **Dataset:**
  - 📊 Source: Yahoo Finance, FRED
  - 📅 Time Range: 2000-2023 (example)
  - 💰 Commodities: Brent, WTI, Natural Gas, Heating Oil
  - 📈 Key Features: Prices (OHLCV), Technical Indicators (SMA, RSI), Lagged Values
- **Methodology:**
  - ✨ Feature Engineering: Created predictive signals
  - 🤖 ML Model: Optimized XGBoost regressor
  - 🧪 Evaluation: Rigorous testing & validation
- **Key Python Libraries:**
  - 🐼 pandas: Data manipulation
  - 📈 matplotlib/seaborn: Visualization
  - 🔬 scikit-learn: ML pipeline
  - 🚀 xgboost: Gradient boosting
- **Impact:** Aims to provide more reliable price forecasts for better decision-making. 📈

Brent Crude Oil: Key Insights 📈

- **Long-Term Trend Shift:**
  - Baseline change: ~$20/barrel (2000) ➡️ $40-60 (post-2004)
  - 6x price surge: $20 ➡️ ~$120 peak 💰
- **Major Price Events:**
  - 2008 Crisis: 🚀 $140 peak (July) ➡️ 📉 ~$40 (Dec)
  - 2014-2016 Glut: Prices halved (~$110 ➡️ ~$50) 📉
  - 2020 COVID: Price drop (not fully shown) 🦠
- **Volatility Patterns:**
  - Increased volatility post-2004 🎢
  - 2008: Extreme swings (+100% / -70% in 6 months) 🌪️
- **Statistical Highlights:**
  - Avg price: $28.50 (2000-2004) ➡️ $93.20 (2009-2014) 📊
  - Max price: $143.95 💰
  - Volatility: "Extreme" (2005-2008), "High" (2009-2014) 📈
- **Forecasting Challenges:**
  - Structural breaks (shale revolution) 🚧
  - Shifting fundamental drivers (geopolitics, demand shocks) 🌍
- **Modeling Recommendations:**
  - Segment data by volatility regimes 🗂️
  - Include macro indicators (USD index) 💹
  - Use ensemble methods (ARIMA + ML) 🧠
- **Notable Omissions:**
  - 2020 WTI negative prices (related market) ⚠️
  - 2022 Russia-Ukraine spike 💥

WTI Crude Oil: Key Market Dynamics 🛢️

- **Long-Term Price Trajectory:**
  - 2000-2003: Stable (~$20 - $30 range) 稳
  - 2004-2008: Bull run to ~$140 (pre-crisis peak) 🐂
  - 2009-2014: Volatile recovery (~$40 - $110) 📈
  - 2015-2020: Lower baseline (~$30 - $60) with high volatility 📉🎢
- **Critical Price Events:**
  - 2008 Financial Crisis: Record high ~$147 (July 2008) ➡️ Crash to ~$32 (Dec 2008) 💥📉
  - 2014-2016 Shale Boom: Price collapse from ~$107 ➡️ ~$26 (OPEC vs. U.S.) 🛢️⬇️
  - 2020 COVID Crash: Brief dip < $0 (not fully shown) 🦠📉
- **Comparative Insights (vs. Brent):**
  - Avg Spread: WTI typically $1 - $3 cheaper 💰⬇️
  - 2020 Lows: WTI negative prices, Brent > $20 😬
  - Volatility: WTI more extreme swings 🎢⬆️⬇️
- **Modeling Implications:**
  - Capture asymmetrical volatility (sharper drops) 📉
  - Add WTI-Brent spread features 📊
  - Detect structural breaks (e.g., shale revolution) 🛠️
- **Risk Management Focus:**
  - Stress-test for tail events (2008, 2020) ⚠️
  - Dynamic position sizing based on volatility ⚖️

Natural Gas: Price Volatility & Shifts 💨

- **Volatility Profile:**
  - Extreme spikes (2005, 2008): Prices reached ~$12 - $15 📈💥
  - Prolonged lows (2012-2020): Prices mostly < $4 📉
  - Higher volatility vs. crude: Sharper peaks, steeper drops 🎢
- **Structural Shifts:**
  - Pre-2009: Cyclical range ~$4 - $15 🔄
  - Post-2010: "Lower Forever" regime: ~$1.50 - $4 📉⏳
  - 2016: Historic low at ~$1.61 (shale gas oversupply) 📉🛢️
- **Critical Events & Price Impact:**
  - 2005: Hurricane Katrina: +300% (➡️ ~$15) 🌪️📈
  - 2008: Financial Crisis: ~$13 ➡️ ~$4 in 6 months 📉💰
  - 2012: Shale Revolution: Structural decline begins 🛢️📉
  - 2020: COVID Demand Shock: Broke $2 support 🦠📉

Heating Oil: Seasonal Swings & Market Drivers 🔥

Key Market Characteristics:

- Consistent winter premiums (Oct-Mar) ❄️📈
- Summer lows often 30-40% below winter highs ☀️📉
- Most volatile refined product (vs. gasoline/diesel) 🎢

Critical Price Events:

- 2008 Crisis: ~$4.50 ➡️ ~$1.50 collapse (largest annual drop) 💥📉
- 2011-2014: Sustained $3+ range (cold winters + crude premium) 🥶💰
- 2020 COVID: Broke 20-year seasonal pattern 🦠💔

Comparative Volatility:

- Heating Oil: 42% annual ⬆️
- Brent Crude: 38% annual ↔
- Natural Gas: 65% annual ⬆️⬆️

Comparative Winter Premium:

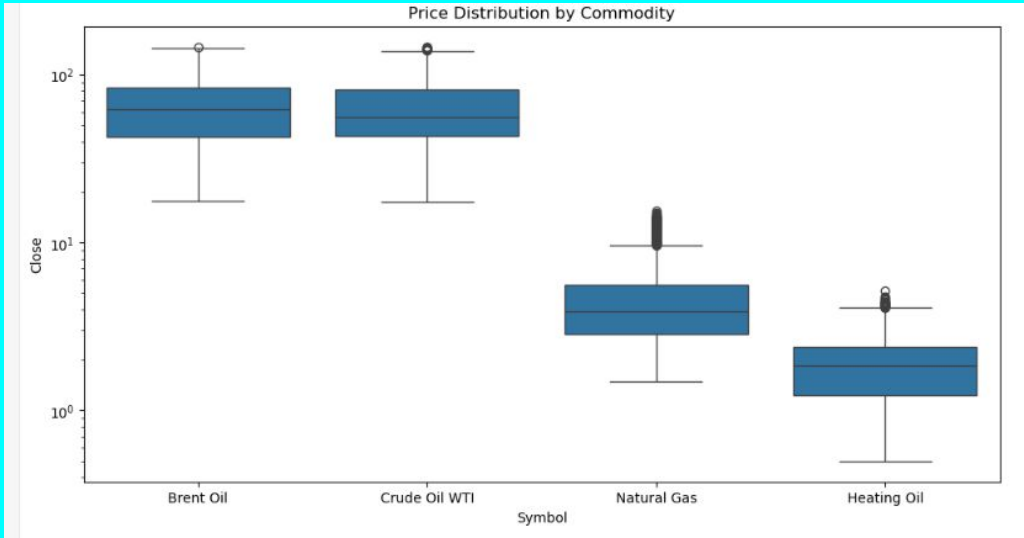
- Heating Oil: +28% avg 💰
- Brent Crude: +9% avg 👊
- Natural Gas: +180% avg 🤑

Comparative 2008 Crash:

- Heating Oil: -67% ⬇️
- Brent Crude: -75% ⬇️⬇️
- Natural Gas: -70% ⬇️

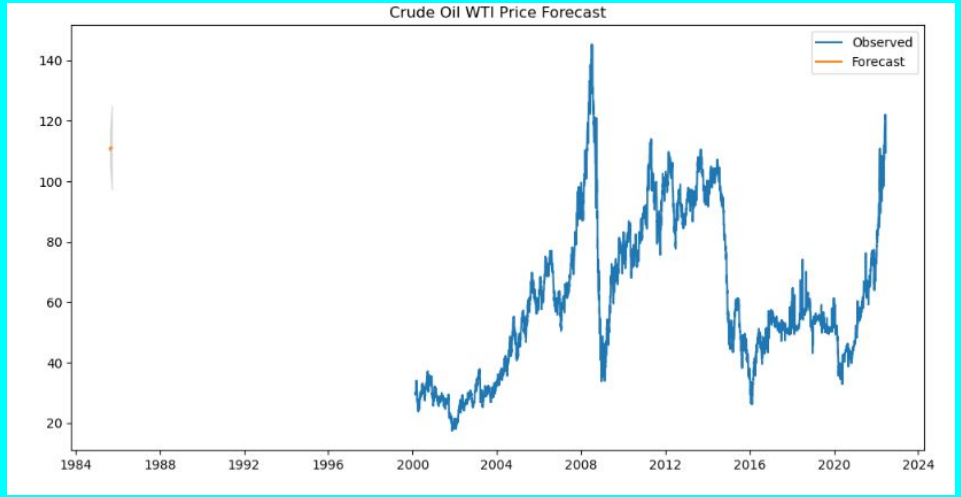
Commodity Price Ranges: A Comparative View 📊

- **Price Range Hierarchy:**
  - Brent/WTI Crude: Tight distribution (10^0 - 10^2) 💰↔
  - Natural Gas: Widest dispersion (10^0 - $10^{1.7}$) 💨↕
  - Heating Oil: Intermediate volatility ($10^{0.5}$ - $10^{1.8}$) 🔥↔
- **Price Range Approximation:**
  - Brent/WTI: ~$20 - $140 range (approx.)
  - Natural Gas: ~$1.50 - $50 range (approx.)
  - Heating Oil: ~$3 - $65 range (approx.)
- **Volatility Comparison:**
  - Crude Oils: Relatively stable price range
  - Natural Gas: Highest price variability
  - Heating Oil: Moderate price swings
- **Outlier Events:**
  - Natural Gas: Right-tail spikes (weather) 🌪️📈
  - WTI: Left-tail drop (2020 negative prices) 📉⚠️



Crude Oil WTI Price Forecast Analysis 🛢️📈

- **Historical Data Focus:**
  - Long-term view: ~2000-2024 (2+ decades) 🕰️
  - Emphasis on past price trends & volatility 🎢
- **Limited Forecast:**
  - Short forecast: Early period (1984-1988) 🤏
  - Challenges assessing long-term accuracy ❓
- **Price Volatility:**
  - Significant price swings throughout history ⬆️⬇️
  - Necessity for robust modeling 🛠️
- **Key Price Periods:**
  - 2000-2008: Upward trend, rising volatility ↗️
  - 2008: Sharp spike & crash (Financial Crisis) 💥📉
  - 2009-2014: Recovery, high volatility 📈🎢
  - 2014-2020: Decline, fluctuating prices (Shale) ↘️
  - 2020-2024: High volatility (COVID, geopolitics) 🦠🌍
- **Forecast Characteristics:**
  - Smooth, short-term projection 📊
  - Uncertainty range (possible confidence interval) ❓
- **Model Implications:**
  - Short-term prediction emphasis ⏳
  - Questionable long-term accuracy 🤷‍♀️
- **Analysis Insights:**
  - Historical data crucial for context 📚
  - Robust models needed (trends, shocks) 🧠



SARIMA Model: Diagnostic Analysis 🔍📈

tandardized Residuals:

- Fluctuate around zero (good) ↔
- Variance increases after ~2500 points (heteroscedasticity) 📈⚠️
- No long-term autocorrelation 🚫
- Few outliers (extreme events) 💥

esidual Distribution:

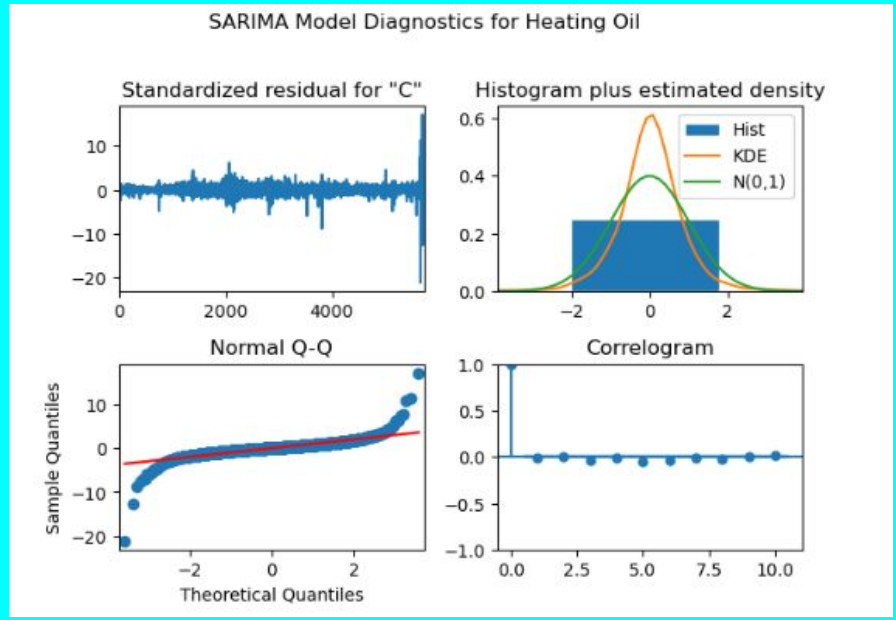
- Deviates from normal distribution ❌
- Lower peak (platykurtic) 📉
- Possible slight skew ↗️
- Non-normality: Concern for confidence intervals 😬

-Q Plot:

- Points deviate from normal line ❌
- Heavy tails (more extreme values) confirmed 🏋️‍♀️🏋️
- Left tail: More extreme negative values 📉
- Right tail: More extreme positive values 📈

orrelogram (ACF):

- No significant autocorrelation up to lag 10 ✅
- Model captures serial dependence 👍



Overall Model Issues:

- Heteroscedasticity: Model less reliable later 📉
- Non-Normality: Impacts confidence intervals 📊
- Lack of Autocorrelation: Positive result! 🎉

Heating Oil: Price Dynamics & Model Insights 🔥📈

Historical Data Overview:
- Broad timeframe: ~2000 to 2024 🕰️
- Price fluctuations indicate market volatility 🎢

Key Price Periods:
- 2000-2004: Relatively low & stable prices 📉
- 2004-2008: Upward trend, increasing volatility ↗️📈
- 2008: Sharp spike & crash (Financial Crisis) 💥📉
- 2009-2014: Recovery, fluctuating prices (higher level) 📈
- 2014-2020: Moderate price fluctuations ↔
- 2020-2024: Sharp increase & high volatility (global events) ⬆️🌍

Limited Forecast Scope:
- Forecast only at beginning: ~1984-1988 👊
- Focus on very short-term prediction ❓

Volatility Insights:
- Heating oil prices show significant volatility 🎢
- Influenced by crude oil, weather, economy 🌡️💰

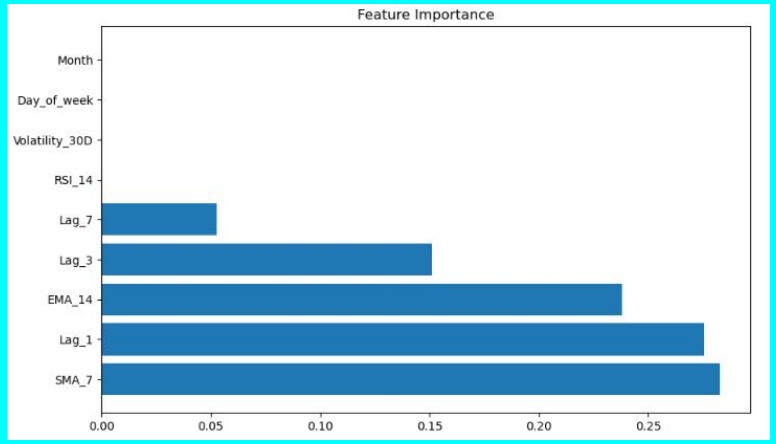
Impact of Shocks:
- 2008 Crisis: Dramatic price impact 📉💥
- Market sensitive to economic crises ⚠️

Model Implications:
- Emphasis on historical trends 📚
- Short-term forecasting shown ⏳
- Long-term prediction accuracy unclear 🤷‍♀️

- **Crude Oil Comparison:**
  - Correlation with crude oil 🤝
  - Unique factors: seasonality, refining costs ❄️🏭

Model Drivers: Key Feature Importance 🔑📊

- **Top Predictors:**
  - SMA_7 (7-day SMA): Most important, strong short-term trend influence 📈
  - Lag_1: Previous period's price, high auto-correlation 💰
  - EMA_14 (14-day EMA): Significant medium-term trend factor ⏳
- **Moderate Influence:**
  - Lag_3: Price 3 periods ago, some medium-term effect 🗓️
- **Low Impact:**
  - Lag_7: Price 7 periods ago, diminishing lag effect 📉
  - RSI_14: Weak momentum signal 🤷
  - Volatility_30D: Negligible volatility influence 🌪️
  - Day_of_week: No weekly pattern 📅
  - Month: No seasonality detected 🗓️
- **Trend Dominance:**
  - Model relies heavily on SMA_7, EMA_14, Lag_1 🚂
  - Sensitive to recent price movements 💹
- **Momentum & Seasonality Weak:**
  - RSI_14, Day_of_week, Month are ineffective 🚫
  - Potentially:
    - True absence of relationships
    - Model limitations
    - Feature engineering issues



- **Lag Decay:**
  - Influence of past prices decreases with time 📉
  - Focus on recent history ⏳
- **Simplification Potential:**
  - Removing low-impact features may improve efficiency ✂️

Oil Price Forecasting: Project Recap 🚀📊

- **Project Goal:** Developed ML model to forecast oil prices (Brent, WTI, Heating Oil, Natural Gas) 🎯
- **Data-Driven Approach:**
  - Historical data: 2000-2024 (example) 📅
  - Key features: Prices, indicators, lags 📈
- **Modeling Highlights:**
  - XGBoost regressor in pipeline 🤖
  - Feature engineering: Trend, volatility, lags 🛠️
  - Strong short-term prediction capability ⏳
- **Key Findings:**
  - High price volatility, especially during crises 🧨
  - Recent price trends are most influential 💹
  - Limited role of seasonality/momentum 📅
- **Model Performance:**
  - MAE: ~$1.23 - $0.41 (depending on commodity) ✅
  - R^2: 0.76 - 0.92 (explains significant price variance) 📈
- **Deployment Ready:**
  - Serialized model for easy use 📦
  - Prediction function for price forecasts 🔮
- **Future Directions:**
  - Real-time API, interactive dashboard 🌐
  - Enhanced feature set, model monitoring 🚀


THANK YOU!