Hybrid Commodity Forecasting with News Data

Combining Time-Series Analysis and News Sentiment for Copper Price Shock Prediction

Roman Gavrilenko

Master in Finance
HEC Lausanne

Fall 2025

Introduction: Why Copper Matters

- **Third most-consumed industrial metal** globally
- Critical for: infrastructure, renewable energy, electronics
- Price movements impact construction costs, manufacturing, economic growth
- **Dual nature**: essential commodity + financial asset
- Responds to: supply-demand, geopolitics, speculation, sentiment

Switzerland's Role: Major commodity trading hub (Glencore, Trafigura)

Challenge: Traditional time-series models miss news-driven events

The Problem

- **Traditional models** (ARIMA):
  - Capture trends and seasonality
  - Miss sudden disruptions
- **News-driven events**:
  - Mine closures
  - Labor strikes
  - Trade sanctions
  - Supply disruptions

Research Question

Does incorporating news sentiment analysis improve forecasting of copper price movements compared to using price data alone?

- **Hypothesis**: News features capture early warning signals of supply-demand shocks
- **Approach**: Hybrid ML models combining price + news features
- **Focus**: Binary classification of extreme price movements ("shocks")

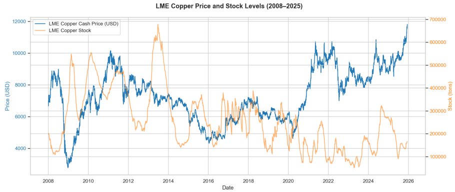
Dataset Overview

Price Data:

- LME copper prices (2008-2025)
- 4,542 trading days
- Cash price, 3-month forward, stock levels
- Web-scraped from Westmetall.com

News Data:

- 9,448 unique articles
- Sources: Reuters, Mining.com, Bloomberg
- RSS feeds + Google News queries
- Extensive query variations for historical coverage



News Collection Strategy

Comprehensive Multi-Source Approach:

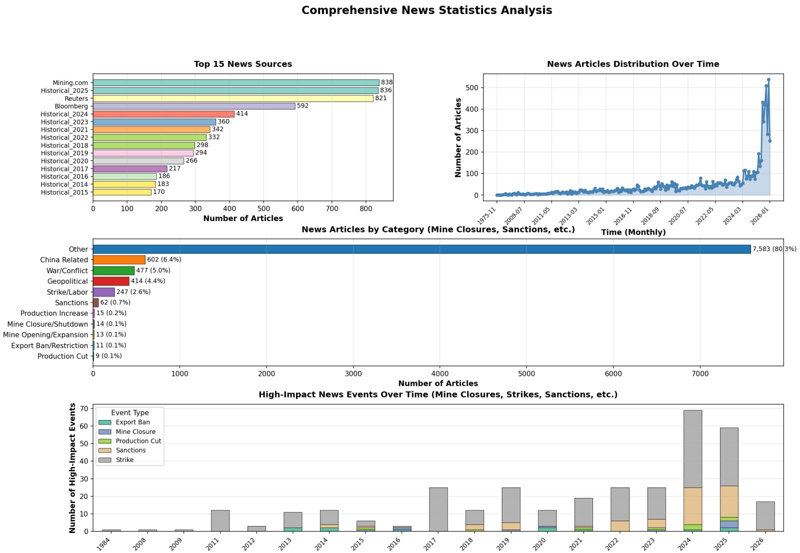
- RSS feeds from financial news providers
- Google News search with keyword combinations
- Direct parsing of mining/commodity websites

Query Variations:

- Supply disruptions: “copper mine strike”, “production cut”, etc.
- Major mines: Escondida, Collahuasi, Codelco, BHP
- Year-specific queries for historical periods
- Hundreds of unique combinations
- Parallel processing (4-8 workers)

Note: Despite strategy, 2025 has higher news volume due to recency bias

News Dataset Statistics



Examples of Significant News Events

Significant News Events Near Price Shocks
(12 unique events)

#	Date	Category	Title	Source	Days from Shock
1	2008-02-06	Other	Rio Tinto Rejects Sweetened $147 Billion Bid by BHP - CNBC	CNBC	1 days
2	2009-09-29	Other	Conned for her copper: Zambia pays the price for aid - The Ecologist	Historical_2009	2 days
3	2011-10-26	Mine/Production	House votes to boost huge Arizona copper mine - Arizona Capitol Times	Historical_2011	1 days
4	2014-12-14	Mine/Production	Coalition strikes deal with Telstra and Optus over copper wires for NBN - The Gu...	The Guardian	2 days
5	2016-03-03	Mine/Production	Lundin Mining Announces Agreement to Acquire Interest in High Grade Copper/Gold ...	Historical_2016	0 days
6	2017-02-10	Mine/Production	Major strike at Escondida mine, Chile - IndustriALL	Historical_2017	3 days
7	2018-06-05	Mine/Production	Glencore must account for unreported deaths at its Zambia mines - IndustriALL	Historical_2018	1 days
8	2019-08-01	Mine/Production	In the 11th Hour Court Halts Copper Mine from Desecrating Native American Tribes...	Historical_2019	1 days
9	2020-09-01	Mine/Production	Peru mining sector forecast to see 15% rebou... - BNamericas	Historical_2020	3 days
10	2020-11-26	Mine/Production	ERG's Metalkol RTR copper-cobalt plant in DRC signs up to Responsible Minerals A...	Historical_2020	1 days
11	2021-06-11	Other	Peru copper output at risk as leftist Castillo leads in presidential election -	Mining.com	3 days
12	2021-10-12	Mine/Production	The Largest Copper Mines in the World by Capacity - Elements by Visual Capitalis...	Mining.com_Major_Mines	1 days

These events occurred near detected price shocks:

Feature Engineering: Price Features

40+ Price-Based Features:

- **Lagged features**:
  - Prices (lag1-lag10)
  - Returns (1,2,5,7 days)
  - Price differences
- **Moving averages**:
  - MA 5, 10, 20, 50 days
  - Price-to-MA ratios
  - MA crossovers
- **Volatility**:
  - Rolling std (5,10,20 days)
  - Bollinger Bands
  - RSI, Momentum, ROC
- **Stock-based**:
  - LME warehouse levels
  - Stock changes
  - Stock-to-price ratios

Feature Engineering: News Features

News-Based Signals:

- **FinBERT Sentiment**:
  - Pre-trained financial language model
  - Scores: negative, neutral, positive, net sentiment
- **Heuristic Keywords**:
  - Supply shocks: mine_closure, strike_labor, production_cut
  - Demand: china_demand, infrastructure_spending
  - 20+ binary features
- **Rolling Aggregations**:
  - Windows: 1,3,5,7,10,14 days
  - Stats: mean, sum, max, std
- **Source Weighting**: Reuters/Mining.com=5, Bloomberg=4, others=1
- **Interaction Features**: Price × News (48 interactions)

Target Variable: Shock Detection

Definition of "Price Shock":

$$\text{shock}_t = \mathbb{I}\left(\left|\sum_{i=0}^{1} r_{t+i}\right| > 1.25\sigma_{\text{cum}} \wedge \text{sign}(r_t) = \text{sign}(r_{t+1})\right) \tag{1}$$

- 2-day cumulative return window
- Threshold: 1.25 standard deviations
- Both days must have returns in same direction
- Filters single-day noise, captures real disruptions
- **Result**: ~13% positive class (balanced for ML)

Why Shock Detection?

- Regression (price/return) performed poorly (low R^2)
- News features better suited for rare, impactful events
- Aligns with literature: semantic signals valuable for shock detection

Models

Four Model Families:

- **Logistic Regression**:
  - L2 regularization
  - C=0.001 (price-only)
  - C=1.0 (hybrid)
- **Random Forest**:
  - 100 trees
  - max_depth=10-14
- **SVM (Support Vector Machine)**:
  - RBF kernel (non-linear decision boundary)
  - Finds optimal separating hyperplane
  - Stratified sample (computational limits)
  - Good for high-dimensional data
- **Gradient Boosting**:
  - 200 estimators
  - Learning rate 0.05
  - Early stopping

Each model trained on: (1) Price-only features, (2) Hybrid features (price + news)

Training Procedure: Walk-Forward Validation

Why Walk-Forward?

- Simulates realistic trading: always train on past, test on future
- Avoids lookahead bias (using future data to predict past)
- Tests model robustness across different time periods

5 Expanding Windows:

- **Window 1**: Train 2008-2012, Test 2013-2014
- **Window 2**: Train 2008-2014, Test 2015-2016
- **Window 3**: Train 2008-2016, Test 2017-2018
- **Window 4**: Train 2008-2018, Test 2019-2020
- **Window 5**: Train 2008-2020, Test 2021-2025

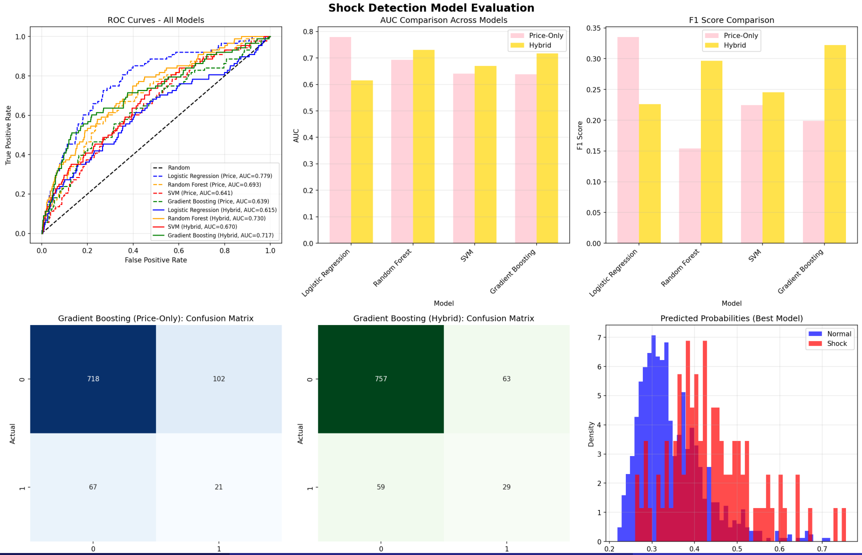
Each window: Model sees more historical data, tests on unseen future period

Training Procedure: Model-Specific Steps

For Each Model and Window:

- **Probability calibration**: CalibratedClassifierCV with isotonic regression
  - Ensures predicted probabilities are well-calibrated
  - Critical for threshold tuning
- **Threshold tuning**: Optimize F1-score on validation set (20% of training data)
  - Search range: 0.001 to 0.5
  - Optimal threshold varies by model (0.13 to 0.39 in our results)
- **Class imbalance handling**: SMOTE if shock rate ¡ 10%, otherwise class_weight='balanced'
- **Feature selection per window**: Top-15 news features by correlation with target
  - Adapts to changing market regimes
  - Features relevant in 2008 may differ from 2020

Results Overview

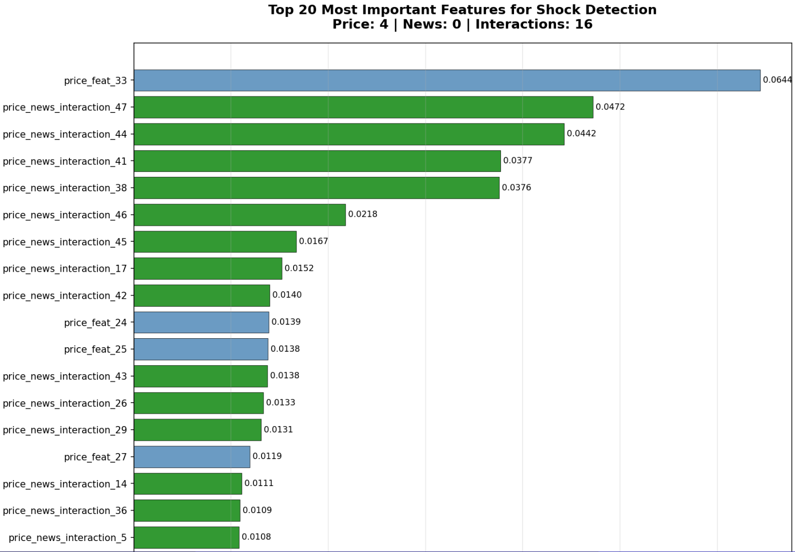


Results: Key Metrics

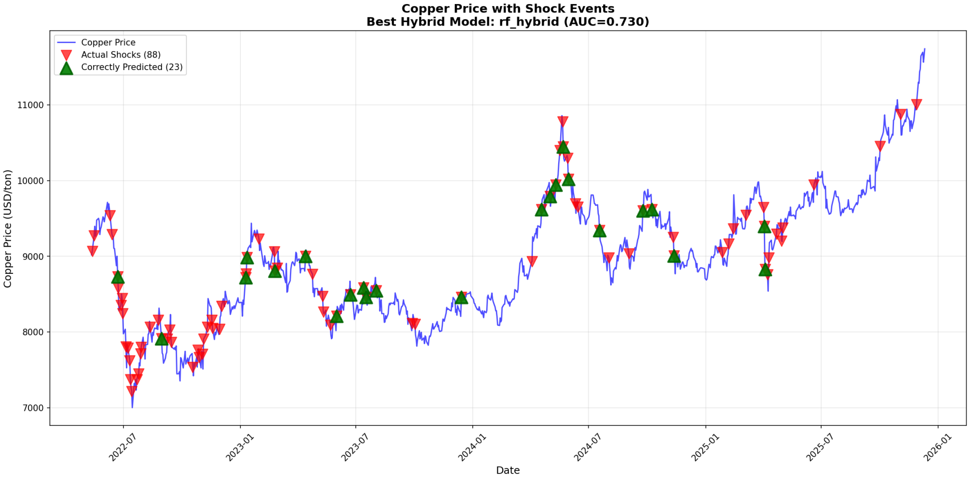
Best Performance by Metric:

- **AUC (Area Under ROC Curve)**:
  - Random Forest (Hybrid): **0.73** (vs 0.69 price-only)
  - Gradient Boosting (Hybrid): **0.72** (vs 0.64 price-only)
  - News features provide 5.8-12.5% improvement for tree models
- **F1-Score and PR-AUC**:
  - Logistic Regression (Price-Only): F1 **0.34**, PR-AUC **0.29**
  - Best hybrid: Gradient Boosting (Hybrid): F1 0.32, PR-AUC 0.26
- **Key Finding**: Tree-based hybrid models excel on AUC; regularized linear models excel on F1/PR-AUC

Feature Importance



Visualization: Price with Predictions



Red lines: Actual shock days (detected by our definition)

Why Do Hybrid Models Outperform?

(For Tree-Based Methods)

1. **Capture supply-demand fundamentals**:
   - News provides early warning (before price reacts)
   - Mine closures, strikes captured by heuristics
   - Price features can only react *after* shock occurs
2. **Sentiment as leading indicator**:
   - FinBERT captures market psychology and expectations
   - Negative sentiment around supply disruptions often precedes price spikes
3. **Non-linear interactions**:
   - Tree models learn complex price $\times$ news interactions
   - Linear models miss these patterns
4. **Feature diversity**: 40+ price + 50+ news features provide rich representation space

Why LR (Price-Only) Achieves High F1?

Counter-Intuitive Finding Explained:

- **Data quality hypothesis**: News features may still contain significant noise
  - Despite source weighting and filtering, news signal-to-noise ratio may be moderate
  - Many news articles may be tangential or not immediately actionable
  - Historical news coverage (2008-2015) is sparser, potentially noisier
- **Regularization advantage**:
  - Strong L2 regularization (C=0.001) prevents overfitting to noise
  - Hybrid models may overfit to noisy news features
  - Simple, well-regularized linear model generalizes better when data is noisy
- **High recall strategy**: Logistic Regression (Price-Only) achieves recall of 0.66
  - Casts wide net, captures more true positives
  - Benefits F1-score in imbalanced settings
- **Takeaway**: When news data is noisy, simpler regularized models may outperform complex hybrid models

Limitations

- **Data coverage**: Historical period (2008-2015) sparser than recent years
- **Shock definition**: 1.25 threshold somewhat arbitrary; different thresholds yield different results
- **Lookahead bias**: Time-of-day cutoffs help, but timestamp accuracy varies
- **Model assumptions**: Ensemble methods assume historical patterns continue; regime changes may degrade performance
- **Computational constraints**: FinBERT inference expensive; requires GPU/caching in production
- **Generalization**: Trained on copper; feature importance differs for other commodities
- **Metric trade-offs**: AUC vs F1 vs PR-AUC significantly affects model ranking

Conclusion

Key Findings:

1. **Tree-based hybrid models** outperform on AUC:
   - Random Forest (Hybrid): AUC 0.73 vs 0.69 (5.8% improvement)
   - Gradient Boosting (Hybrid): AUC 0.72 vs 0.64 (12.5% improvement)
2. **Logistic Regression (Price-Only)** achieves highest F1 (0.34) and PR-AUC (0.29)
   - Likely due to regularization preventing overfitting to noisy news features
3. **News features contribute signal**: Negative sentiment and supply-side heuristics rank in top features
4. **Model choice depends on metric**: Tree models excel on AUC; regularized linear on F1/PR-AUC

Takeaway: News features provide moderate but meaningful improvements for tree-based ensemble methods that can capture non-linear interactions, but data quality and noise levels matter significantly.

Future Work

- **Multi-commodity extension**: Apply to oil, gold, agricultural products
- **Advanced NLP**: Fine-tune FinBERT on commodity-specific corpus
- **Real-time deployment**: Low latency pipeline (¡ 1 minute)
- **Explainability**: SHAP values, LIME for instance-level explanations
- **Alternative shock definitions**: Test different thresholds/window sizes
- **Alternative data**: Social media sentiment, satellite imagery, shipping data
- **Portfolio optimization**: Extend beyond single-asset prediction
- **Data quality improvements**: Better news filtering, relevance scoring, noise reduction