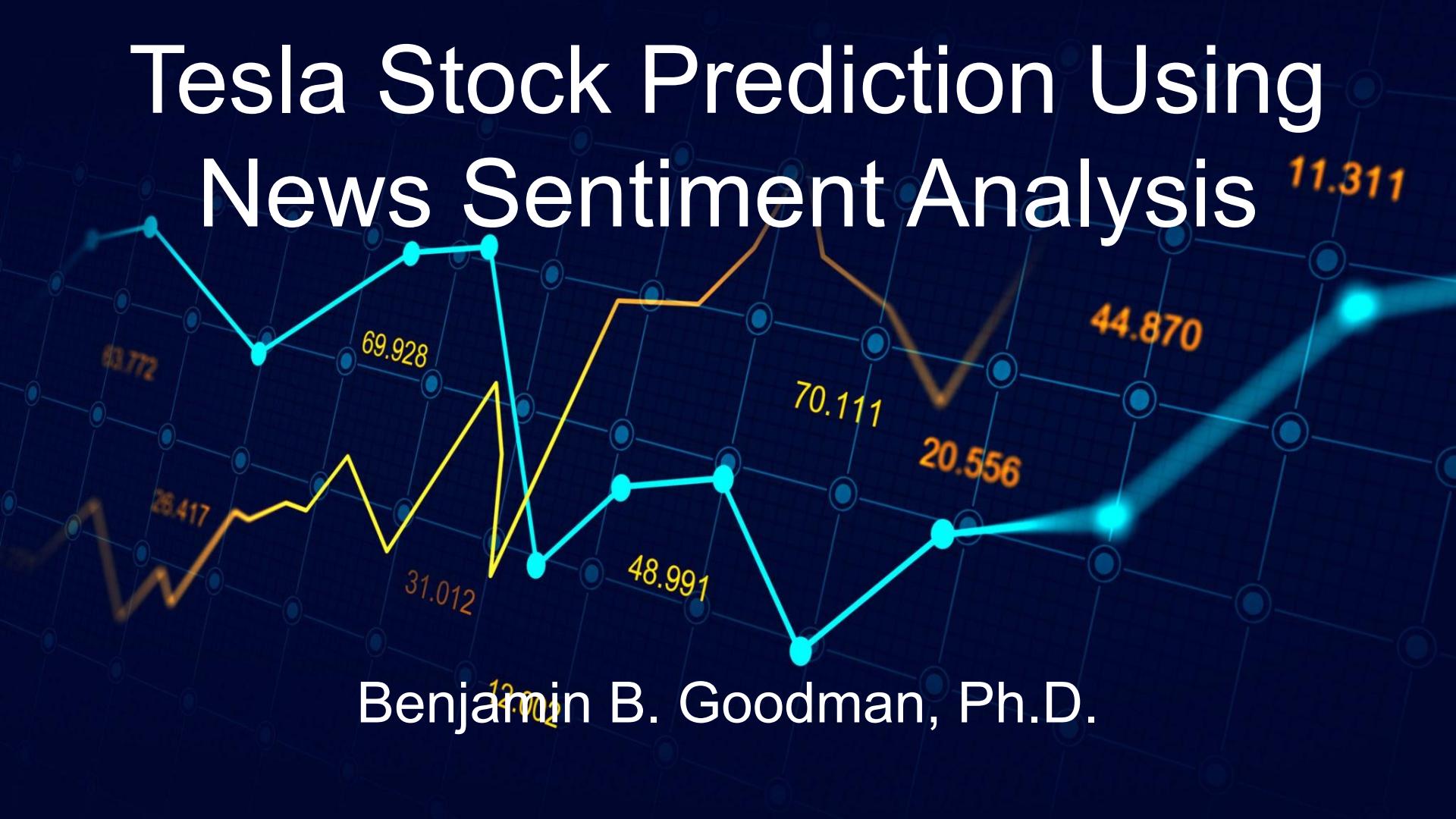


Tesla Stock Prediction Using News Sentiment Analysis



Benjamin B. Goodman, Ph.D.

Problem Statement

Research Question:

- Does daily news sentiment contain predictive signals for Tesla's short-term stock price direction?

Approach:

- Predict 5-day forward trend (up vs. down) using:
 - NLP sentiment analysis
 - Feature engineering
 - Traditional ML methods

Why Predict 5-Day Trends (Not Next-Day)?

Daily movements are **dominated by noise**:

- Random market fluctuations, highly volatile
- Initial model attempt: **AUC ~0.47** (essentially random)

5-day trends help **filter volatility** and capture signal:

- Sentiment takes time to propagate through market
- Improved model: **AUC 0.57** (meaningful signal)

Result: More realistic, actionable predictions

Data Sources

Stock Prices:

- Yahoo Finance, Daily OHLCV data (2024-2025)

News Articles:

- Google News feeds
- 25,398 articles across, 732 days (2024-2025)

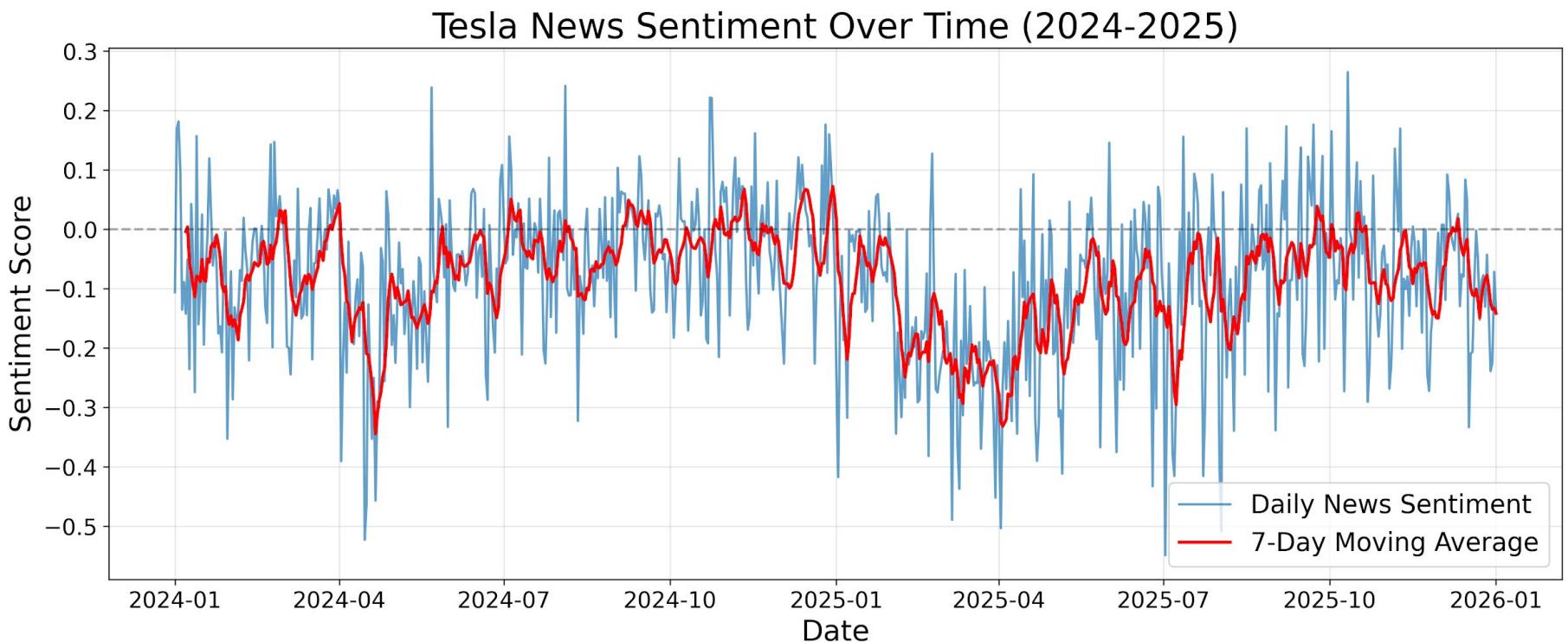
Earnings Calls:

- 9 quarterly transcripts (Q4 2023 - Q4 2025)

SEC Filings:

- Collected from SEC EDGAR but not analyzed (time scope)

News Sentiment Over Time



Sentiment Analysis Tool

FinBERT: Financial domain-adapted BERT model

- Pre-trained on earnings calls, analyst reports, financial news
- Understands finance-specific language

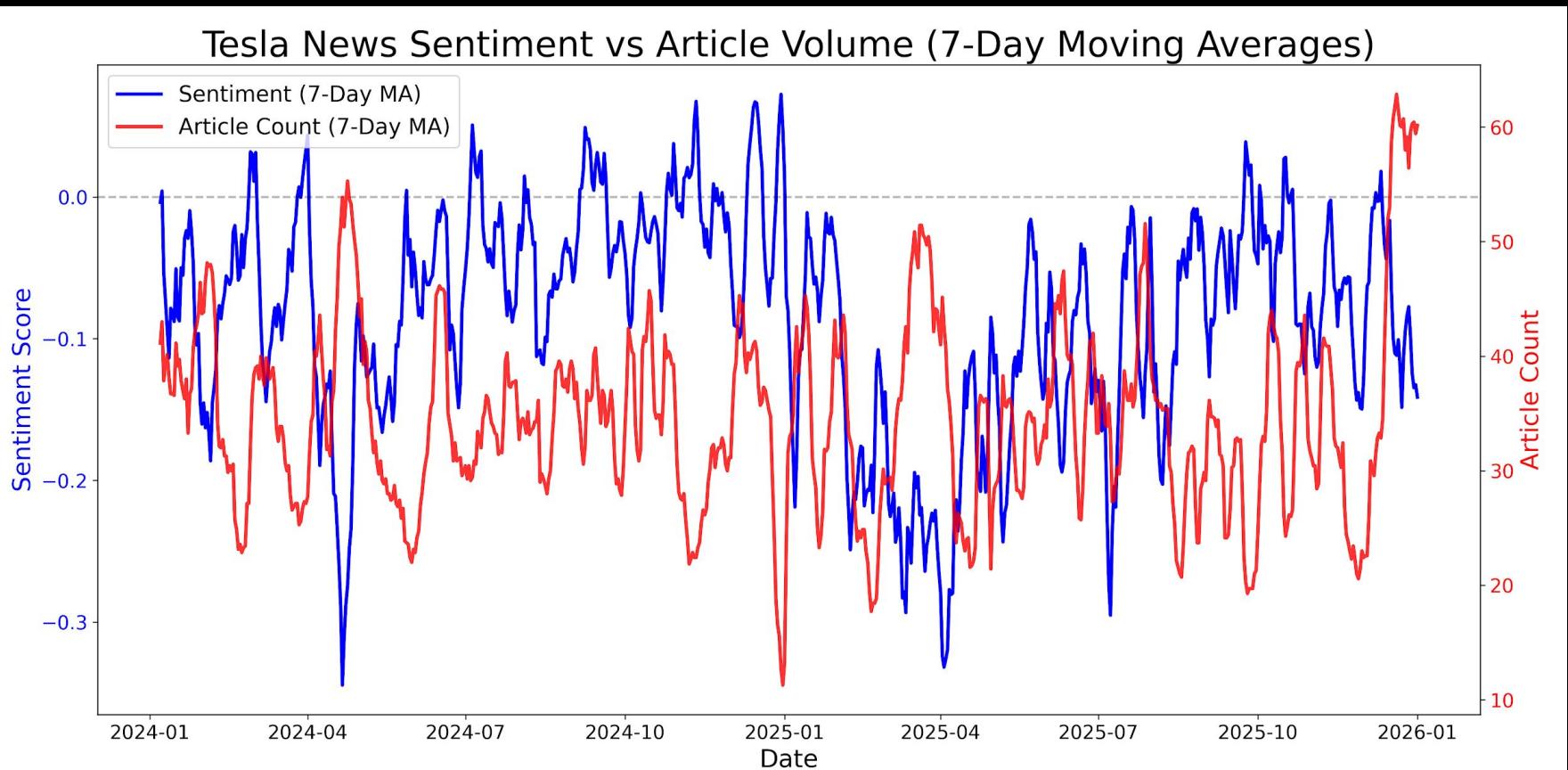
Outputs:

- Positive/negative/neutral sentiment and confidence score

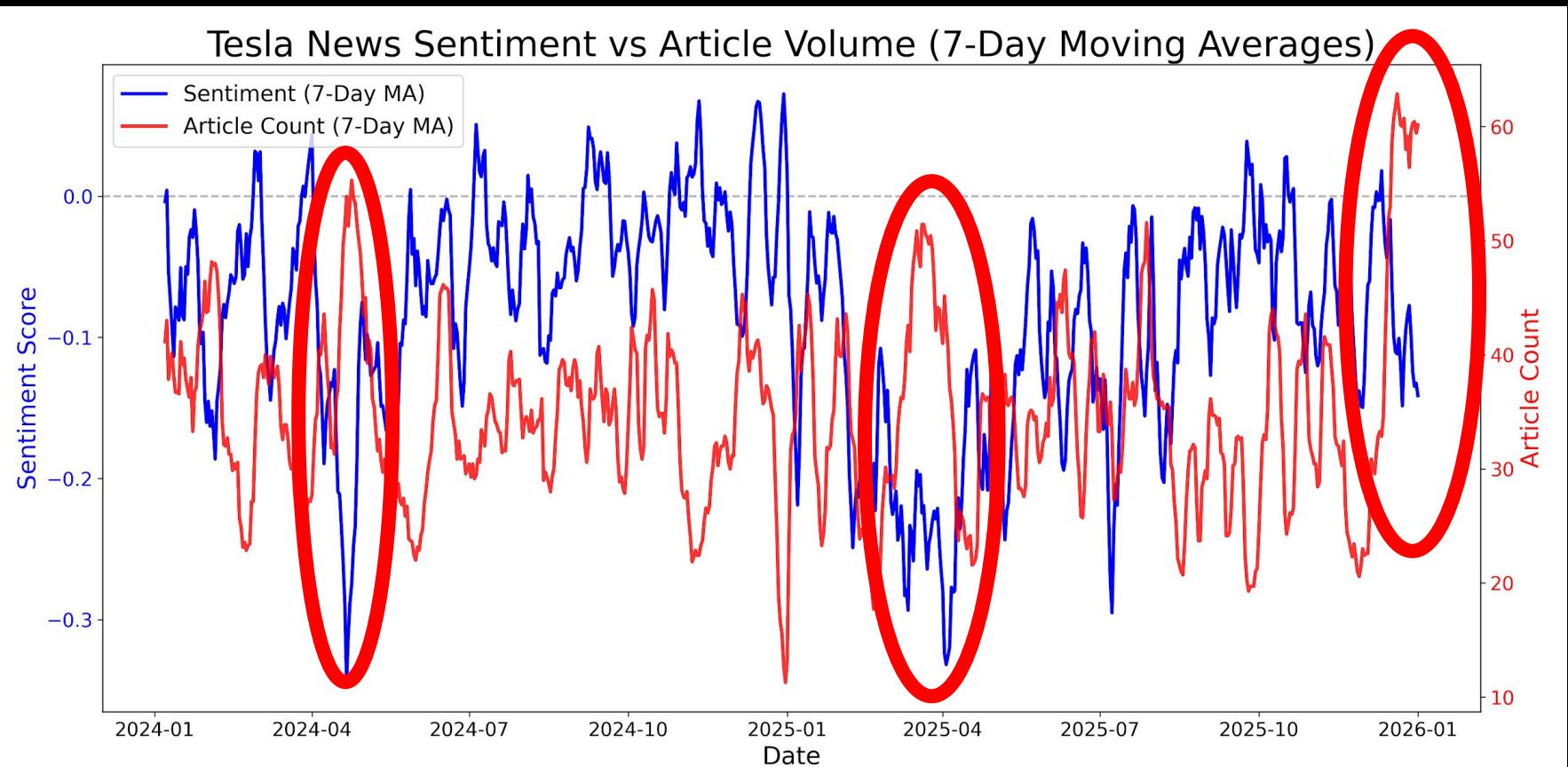
Applied to:

- News headlines (daily aggregation)
- Earnings transcripts (quarterly aggregation)

Key Discovery: Sentiment vs. Volume

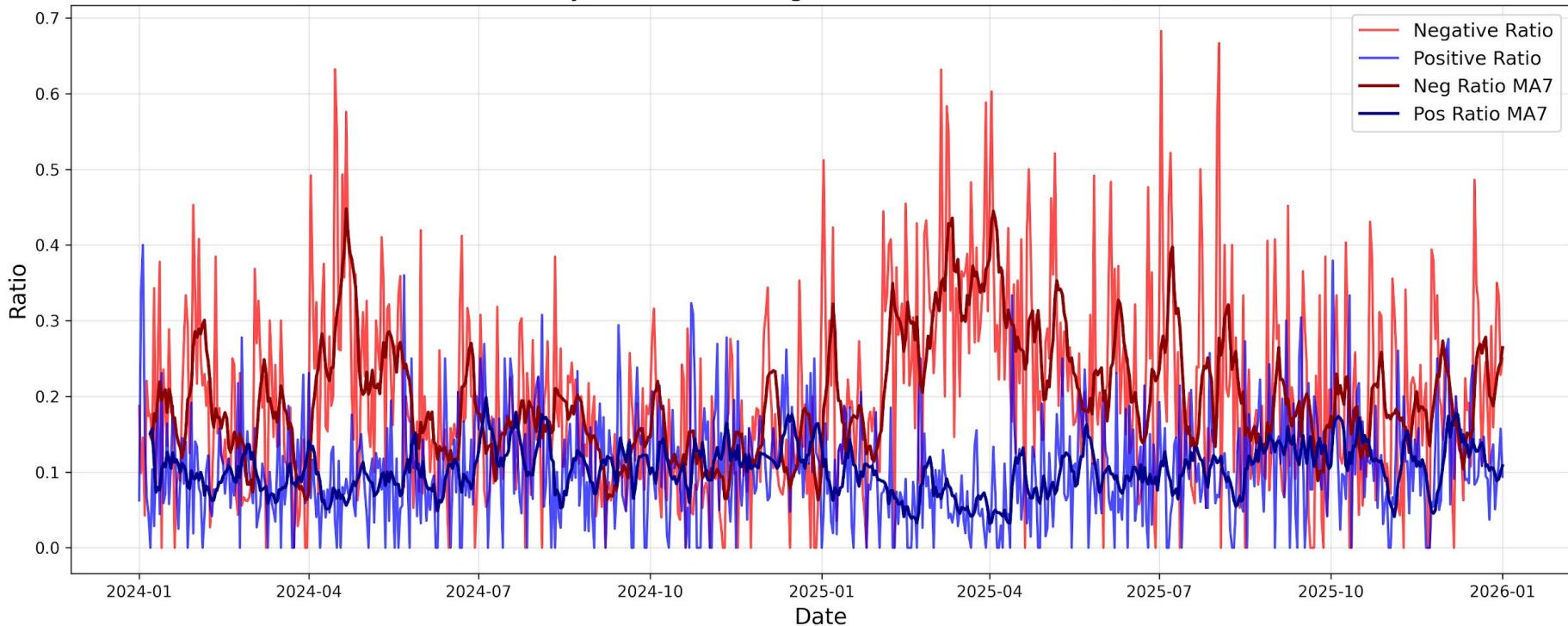


Article Volume Spikes Align with Sentiment Drop



Article Sentiment Distribution

Daily Positive and Negative News Ratios (Tesla)



Most Tesla news is negative: Explains high volume → negative sentiment

Statistical Validation

A/B Test: High vs. Low Article Volume Days (split by median)

- High-volume days:
 - Average sentiment = -0.097
- Low-volume days:
 - Average sentiment = -0.071

Difference: -0.026

Statistical significance: p-value < 0.00001

Conclusion: High-volume days are reliably more negative.

Feature Engineering

Sentiment Features:

- Average sentiment score: Daily, 7 day moving average, Previous day (lagged)

Volume Features:

- Article count: Daily, 7 day moving average, Previous day (lagged)

Interaction Term:

- Sentiment × Article count (captures effect: one is low/high vs the other)

Stock Features:

- Daily return, 5-day moving average return, volume change

Earnings:

- Latest quarterly earnings call sentiment score (forward-filled)

Earnings Call Analysis

Process:

1. Collected 9 earnings call transcripts (quarterly calls).
2. Chunked text into 512-character segments.
3. Analyzed sentiment per chunk (FinBERT).
4. Aggregated average sentiment to create quarterly score.

Application:

Forward-filled quarterly scores between calls to provide management tone signal for daily predictions.

Train/Test Split (Avoiding Data Leakage)

- Training:
 - Jan 1, 2024 - Jan 31, 2025 (257 samples)
- Buffer (prevents overlap from 7-day moving averages):
 - Feb 1-9, 2025 (10 days, discarded)
- Test:
 - Feb 10, 2025 - End of 2025 (225 samples)

Ensures no look-ahead or look-behind bias.

Class Balance:

- Train: 50% up / 50% down
- Test: 52% up / 48% down

Model Selection

Logistic Regression (L2): AUC 0.57 ✓

XGBoost: AUC 0.52

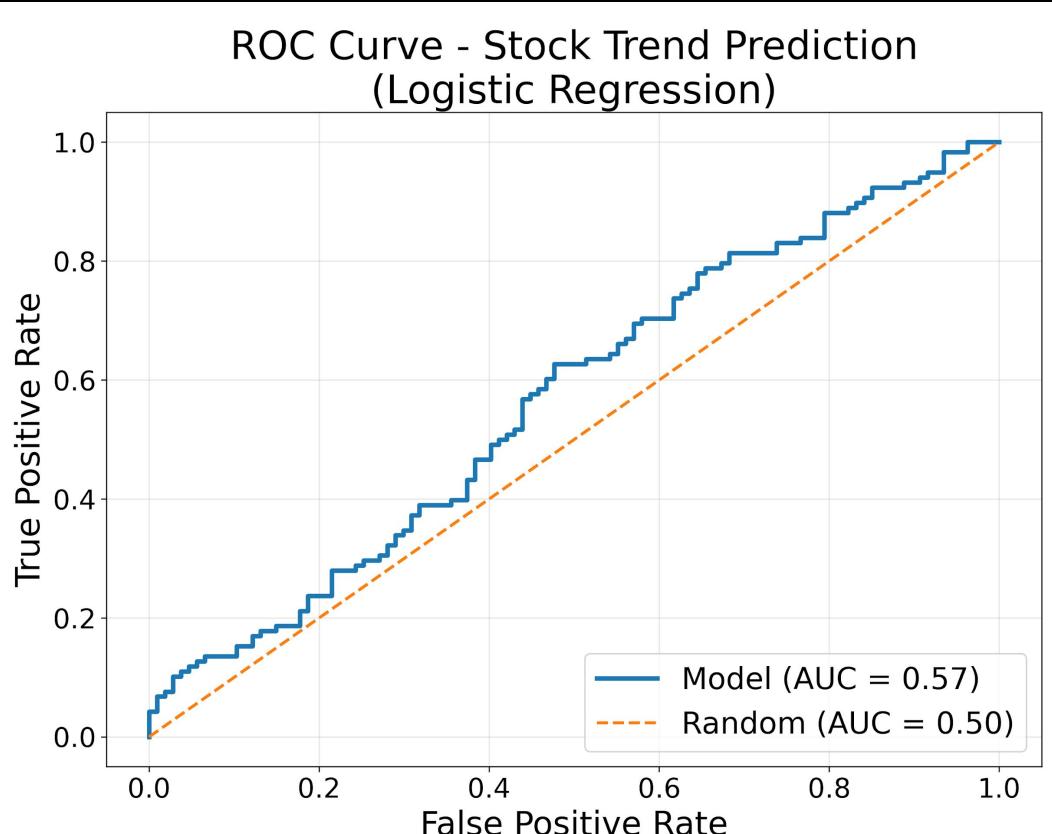
Why Logistic Regression?

- Better performance on small dataset.
- Interpretable coefficients.
- Well-suited for linear relationships.

XGBoost underperformed due to:

- Limited training data (257 samples).
- Linear feature relationships.

Model Performance - ROC Curve

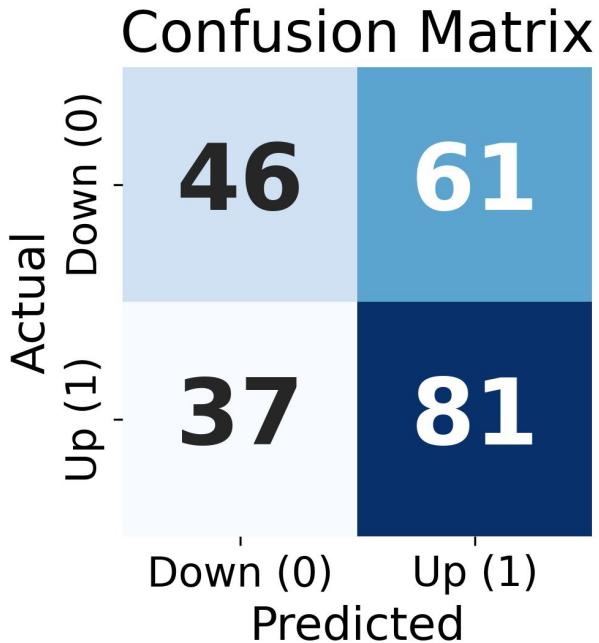


AUC: 0.57 → 14% improvement over random guessing (0.50)

Interpretation:

- Model successfully extracts signal from noisy data.
- Realistic performance for stock prediction.

Confusion Matrix



Prediction Accuracy Breakdown

Accuracy: 56%

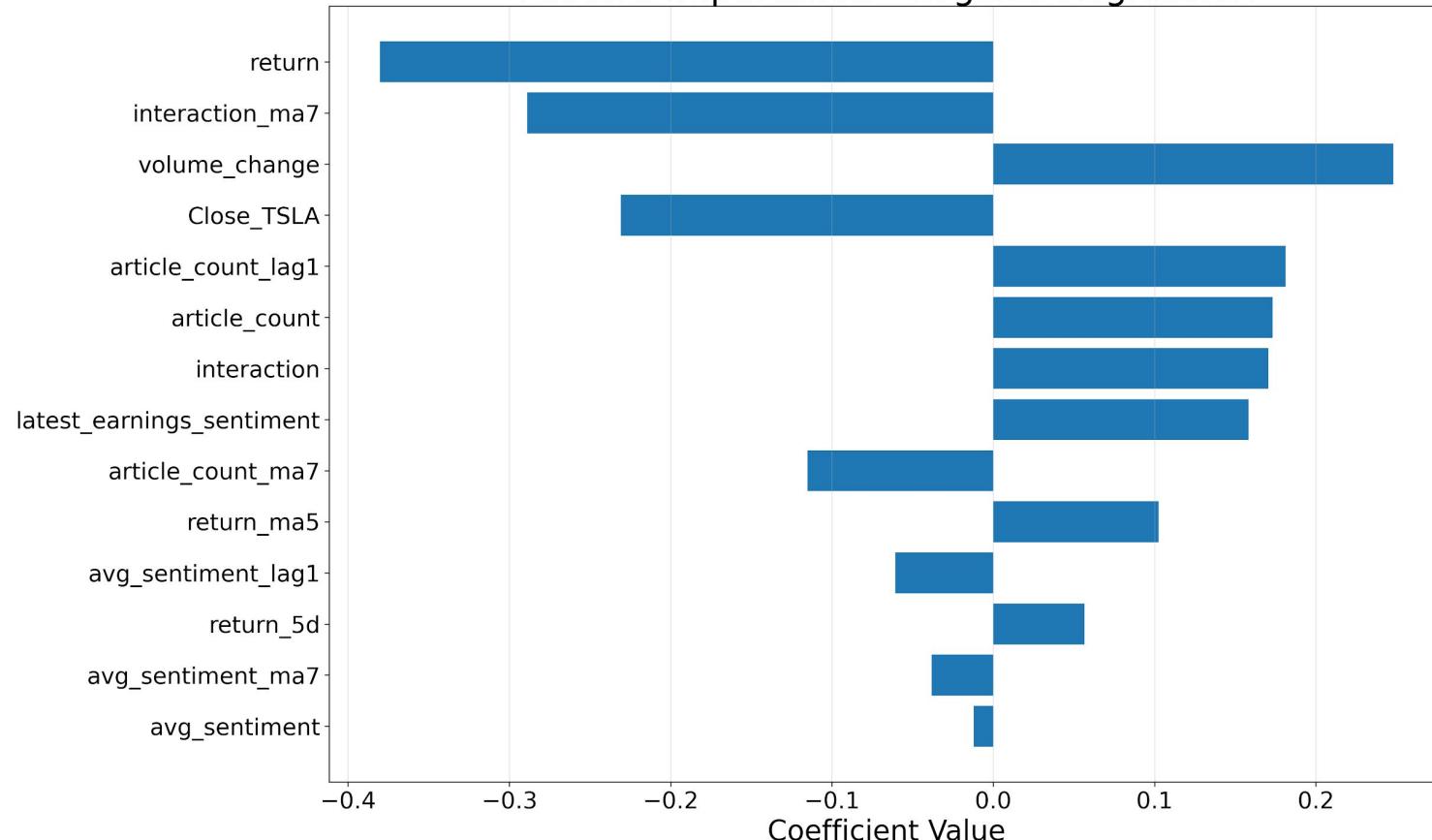
Precision (Up): 57%

Recall (Up): 69%

Model **slightly better** at predicting upward stock trends.

Feature Importance

Feature Importance – Logistic Regression



Answering Research Question

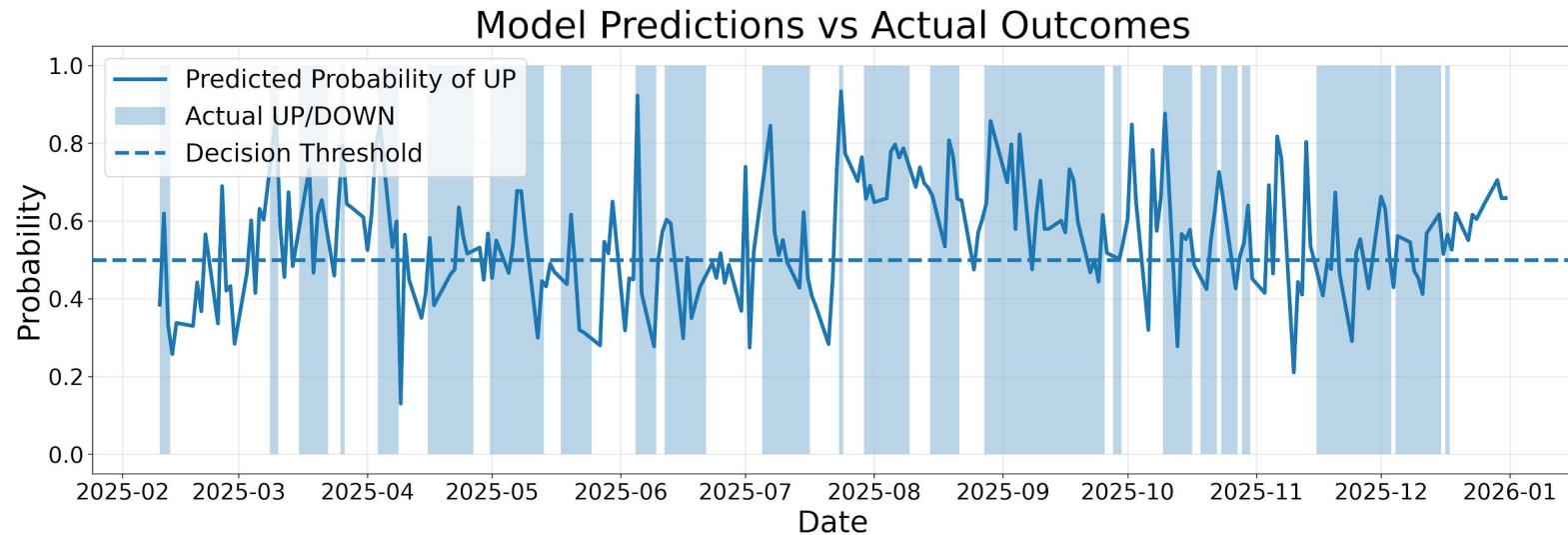
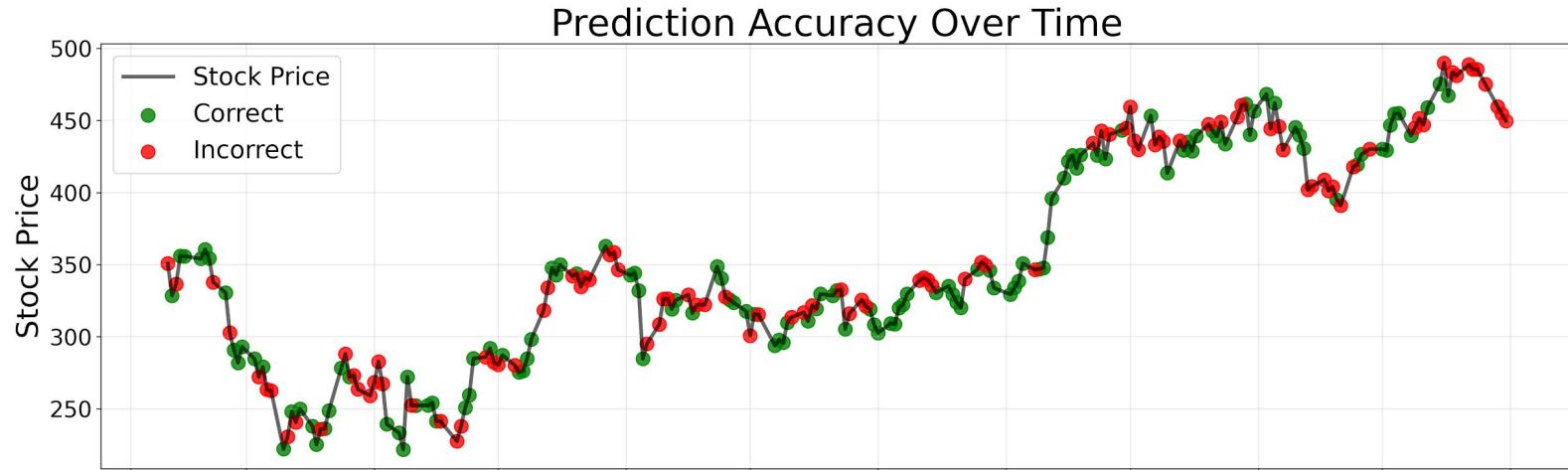
Does news sentiment contain predictive signals? YES.

Evidence:

- ✓ Sentiment × article count interaction moving average: Top 5 feature
- ✓ Article count (lagged day): Top 5 feature
- ✓ Article count (current day): Top 5 feature
- ✓ Model AUC 0.57 vs 0.50 random (14% improvement)

Implication:

News sentiment is a valuable supplementary signal for short-term Tesla stock prediction.



Applications for OTPP

- Early Warning System:
 - Detect unusual negative sentiment spikes.
- Investment Timing:
 - Use sentiment as additional signal for investment.
- Portfolio Monitoring:
 - Track sentiment across multiple holdings.
 - Identify companies facing negative media attention.
- Scalable Framework:
 - Same approach works for any publicly traded company.
 - Can analyze simultaneously.
- Automated System:
 - Daily updates: new articles → sentiment scores → predictions

Limitations & Next Steps

Limitations:

- Small dataset (257 training samples).
- Single company (Tesla-specific patterns).
- News sentiment only (excludes social media).
- Correlation, not causation.

Next Steps:

- Expand to multiple companies/sectors.
- Add social media sentiment (Twitter/Reddit).
- Longer time horizons (weekly/monthly trends).