

Cryptocurrency Price Prediction with Sentiment Analysis

Experiment Report: Data Collection, Methodology, and Results

Crypto Sentiment Research Team

December 16, 2025

Abstract

This report presents a comprehensive analysis of cryptocurrency price prediction using news sentiment analysis. We collected 10,805 hourly price observations across five major cryptocurrencies and 301 news articles from nine RSS sources. Our methodology combines traditional machine learning models (Linear Regression, Ridge, Random Forest, XGBoost, LightGBM) with deep learning architectures (LSTM, BiLSTM, GRU). Results demonstrate statistically significant correlations between sentiment features and price returns (Pearson $r = 0.12, p < 0.001$). BiLSTM achieved the highest F1-score (66.0%) for direction classification, while XGBoost demonstrated the best balanced performance (55.4% accuracy, 57.4% ROC-AUC). Ridge regression showed optimal generalization for regression tasks ($R^2 = 0.007$).

Contents

1 Executive Summary	3
1.1 Key Findings	3
2 Data Collection	3
2.1 Price Data	3
2.1.1 Price Statistics	3
2.2 News Data	4
2.2.1 News Sources Distribution	4
2.3 Data Pipeline	4
3 Methodology	5
3.1 Data Preprocessing	5
3.1.1 Text Cleaning	5
3.1.2 Sentiment Analysis	5
3.1.3 Feature Engineering	5
3.2 Data Split	6
3.3 Models Evaluated	6
3.3.1 Traditional Machine Learning	6
3.3.2 Deep Learning Models	6
3.3.3 LSTM Architecture	6

3.4	Evaluation Metrics	7
3.4.1	Regression Metrics	7
3.4.2	Classification Metrics	7
4	Results	7
4.1	Correlation Analysis	7
4.2	Hypothesis Test Results	8
4.3	Regression Results	8
4.3.1	Validation Set Performance	8
4.3.2	Test Set Performance	9
4.4	Classification Results	9
4.4.1	Validation Set Performance	9
4.4.2	Test Set Performance	9
4.5	Residual Analysis	9
5	Discussion	10
5.1	Sentiment-Price Relationship	10
5.2	Model Performance Analysis	10
5.2.1	Regression Tasks	10
5.2.2	Classification Tasks	10
5.3	Deep Learning Observations	11
6	Conclusions	11
6.1	Key Findings	11
6.2	Limitations	11
6.3	Recommendations	12
A	Configuration Parameters	12
B	LSTM Hyperparameters	12
C	Feature List	12

1 Executive Summary

This research investigates the predictive relationship between news sentiment and cryptocurrency price movements. We collected price data for five major cryptocurrencies and news articles from nine RSS sources, applied multiple sentiment analysis techniques, and evaluated both traditional machine learning and deep learning models for price movement prediction.

1.1 Key Findings

- Sentiment features show statistically significant correlations with price returns ($p < 0.001$)
- BiLSTM achieved the highest F1-score (66.0%) for direction classification
- XGBoost demonstrated the best balanced classification performance (55.4% accuracy)
- Ridge regression showed the best generalization for regression tasks ($R^2 = 0.007$)
- Statistical hypothesis tests confirm significant impact of sentiment on returns (Welch's $t = 9.53$, $p < 10^{-21}$)

2 Data Collection

2.1 Price Data

We collected hourly cryptocurrency price data for five major coins using multiple API sources.

Table 1: Price Data Overview

Metric	Value
Total Observations	10,805
Cryptocurrencies	BTC, ETH, SOL, ADA, DOT
Granularity	Hourly
Features	Price, Volume, Market Cap

2.1.1 Price Statistics

Table 2 presents comprehensive descriptive statistics for the collected price data.

Table 2: Descriptive Statistics for Price Data

Statistic	Price (USD)	Volume (USD)	Market Cap (USD)
Mean	\$21,755.85	\$21.45B	\$532.91B
Median	\$184.78	\$6.64B	\$101.30B
Std Dev	\$41,904.14	\$28.09B	\$802.60B
Min	\$0.37	\$79.14M	\$3.22B
Max	\$126,079.89	\$200.97B	\$2,507.87B
25th Percentile	\$2.42	\$879.93M	\$16.89B
75th Percentile	\$4,147.92	\$36.64B	\$500.72B
Skewness	1.549	1.614	1.415
Kurtosis	0.491	2.954	0.284

2.2 News Data

News articles were collected from nine cryptocurrency-focused RSS feeds.

Table 3: News Data Overview

Metric	Value
Total Articles	301
Sources	9 RSS Feeds
Keywords Tracked	bitcoin, ethereum, cryptocurrency, crypto, blockchain

2.2.1 News Sources Distribution

Table 4: Distribution of News Articles by Source

Source	Articles	Percentage
U.Today	94	31.2%
Decrypt	58	19.3%
CryptoPotato	36	12.0%
Cointelegraph	30	10.0%
CoinDesk	25	8.3%
CryptoNews	20	6.6%
AMBCrypto	16	5.3%
BeInCrypto	12	4.0%
Bitcoin Magazine	10	3.3%
Total	301	100%

2.3 Data Pipeline

The data processing pipeline follows a structured workflow:

1. **News Collection:** RSS feeds parsed for cryptocurrency-related articles
2. **Text Preprocessing:** HTML removal, tokenization, stopword filtering

3. **Sentiment Analysis:** VADER and TextBlob sentiment scoring
4. **Hourly Aggregation:** Sentiment metrics aggregated to hourly intervals
5. **Feature Engineering:** Technical indicators and derived features
6. **Data Merging:** Price and sentiment data aligned by timestamp

3 Methodology

3.1 Data Preprocessing

3.1.1 Text Cleaning

- HTML tag removal using BeautifulSoup
- URL and special character filtering via regex
- Lowercase conversion for normalization
- Tokenization using NLTK
- Stopword removal (English stopwords)

3.1.2 Sentiment Analysis

Two complementary approaches were employed:

- **VADER:** Valence Aware Dictionary and sEntiment Reasoner for social media-optimized sentiment
- **TextBlob:** Polarity and subjectivity scoring

3.1.3 Feature Engineering

The following features were engineered from raw data:

Price Features:

- Price returns: $r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$
- Moving averages: MA_{24h}, MA_{168h}
- Volatility: σ_{24h} , σ_{168h}
- Momentum: $M_{24h} = P_t - P_{t-24}$

Sentiment Features:

- Hourly aggregations: mean, min, max, std
- Polarity and subjectivity means
- Positive, negative, neutral proportions
- Article count per hour

3.2 Data Split

Data was split chronologically to prevent look-ahead bias:

Table 5: Data Split Configuration

Set	Ratio	Samples
Training	60%	6,477
Validation	20%	2,159
Test	20%	2,159
Total	100%	10,795

3.3 Models Evaluated

3.3.1 Traditional Machine Learning

Table 6: Traditional ML Models

Model	Description
Linear Regression	Baseline OLS regression
Ridge Regression	L2 regularized linear model
Random Forest	Ensemble of 100 decision trees
XGBoost	Gradient boosting with regularization
LightGBM	Histogram-based gradient boosting

3.3.2 Deep Learning Models

Table 7: Deep Learning Architectures

Model	Description
LSTM	Long Short-Term Memory network
BiLSTM	Bidirectional LSTM for forward/backward context
GRU	Gated Recurrent Unit

3.3.3 LSTM Architecture

The LSTM models follow a standardized architecture:

- **Sequence Length:** 24 hours (lookback window)
- **Hidden Units:** [128, 64] (two-layer stack)
- **Dropout Rate:** 0.2
- **Regularization:** L2 ($\lambda = 0.01$)
- **Batch Normalization:** After each recurrent layer
- **Optimizer:** Adam (learning rate = 0.001)

- **Early Stopping:** Patience = 15 epochs

The LSTM gate equations are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

3.4 Evaluation Metrics

3.4.1 Regression Metrics

- **MAE:** Mean Absolute Error = $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- **RMSE:** Root Mean Squared Error = $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
- **R^2 :** Coefficient of Determination = $1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$
- **MAPE:** Mean Absolute Percentage Error = $\frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$

3.4.2 Classification Metrics

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** $\frac{TP}{TP+FP}$
- **Recall:** $\frac{TP}{TP+FN}$
- **F1-Score:** $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- **ROC-AUC:** Area under Receiver Operating Characteristic curve

4 Results

4.1 Correlation Analysis

Table 8 presents the correlation analysis between sentiment features and price returns.

Table 8: Sentiment Features vs. Price Returns Correlation

Feature	Pearson r	p -value	Spearman ρ	p -value
sentiment_mean	0.120	$3.60 \times 10^{-36}***$	0.081	$2.81 \times 10^{-17}***$
polarity_mean	0.102	$3.82 \times 10^{-26}***$	0.065	$1.03 \times 10^{-11}***$
negative_mean	-0.086	$2.71 \times 10^{-19}***$	-0.074	$9.43 \times 10^{-15}***$
sentiment_min	0.084	$2.09 \times 10^{-18}***$	0.032	$8.92 \times 10^{-4}**$
positive_mean	0.077	$1.12 \times 10^{-15}***$	0.065	$1.60 \times 10^{-11}***$
price_return	0.069	$5.03 \times 10^{-13}***$	-0.013	0.167
sentiment_std	-0.037	$1.40 \times 10^{-4}***$	-0.020	0.039*

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.2 Hypothesis Test Results

We conducted statistical hypothesis testing to evaluate the impact of sentiment on price returns.

Table 9: Descriptive Statistics by Sentiment Condition

Condition	Mean Return	Std Dev	Sample Size
Positive Sentiment	-0.00027	0.135	6,907,976
Negative Sentiment	-0.00097	0.140	7,143,716

Table 10: Hypothesis Test Results

Test	Statistic	p -value	Significant
Welch's t -test	9.530	$1.57 \times 10^{-21}***$	Yes
Mann-Whitney U	2.47×10^{13}	$9.15 \times 10^{-11}***$	Yes
Cohen's d	0.0051	—	Small effect

4.3 Regression Results

4.3.1 Validation Set Performance

Table 11: Regression Model Performance on Validation Set

Model	MAE	RMSE	R^2
XGBoost	0.543	0.791	0.048
LightGBM	0.545	0.800	0.026
Ridge	0.544	0.803	0.017
Random Forest	0.543	0.805	0.013
Linear	0.543	0.811	-0.002

4.3.2 Test Set Performance

Table 12: Regression Model Performance on Test Set

Model	MAE	RMSE	R^2
BiLSTM	0.511	0.793	-0.0001
LSTM	0.511	0.794	-0.0003
GRU	0.513	0.793	-0.0002
Linear	0.511	0.794	-0.002
Random Forest	0.511	0.800	-0.018
LightGBM	0.516	0.793	0.0004
Ridge	0.519	0.790	0.007
XGBoost	0.521	0.854	-0.160

4.4 Classification Results

4.4.1 Validation Set Performance

Table 13: Classification Model Performance on Validation Set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
XGBoost	53.5%	52.8%	50.1%	51.4%	54.6%
LightGBM	51.8%	51.1%	46.7%	48.8%	52.6%
Random Forest	51.8%	51.0%	48.4%	49.6%	51.7%
Logistic	50.5%	49.7%	53.3%	51.4%	51.6%

4.4.2 Test Set Performance

Table 14: Classification Model Performance on Test Set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
BiLSTM	49.2%	49.2%	100%	66.0%	51.9%
XGBoost	55.4%	54.7%	54.6%	54.7%	57.4%
LightGBM	52.8%	52.3%	47.6%	49.8%	55.3%
LSTM	50.8%	0.0%	0.0%	0.0%	50.0%
Logistic	49.3%	48.6%	54.5%	51.4%	49.3%
Random Forest	50.2%	49.4%	49.2%	49.3%	51.1%

4.5 Residual Analysis

Residual analysis was performed on the best-performing regression model (Ridge).

Table 15: Residual Analysis for Ridge Regression

Metric	Value
Mean Residual	-0.010
Std Residual	0.790
Min Residual	-6.417
Max Residual	7.463
Durbin-Watson	1.967 (no autocorrelation)

Table 16: Normality Tests for Residuals

Test	Statistic	p-value	Normal?
Shapiro-Wilk	0.886	3.14×10^{-37}	No
D'Agostino K^2	446.66	1.02×10^{-97}	No

5 Discussion

5.1 Sentiment-Price Relationship

Our analysis reveals statistically significant correlations between news sentiment and cryptocurrency price returns. The sentiment mean exhibits the strongest correlation ($r = 0.120$), followed by polarity mean ($r = 0.102$). Notably, negative sentiment shows an inverse relationship ($r = -0.086$), confirming the intuitive expectation that negative news correlates with price declines.

The hypothesis tests strongly support the existence of differential returns based on sentiment conditions. Welch's t -test ($t = 9.53, p < 10^{-21}$) and Mann-Whitney U test confirm statistical significance. However, the small effect size (Cohen's $d = 0.005$) suggests limited practical predictability, consistent with market efficiency hypotheses.

5.2 Model Performance Analysis

5.2.1 Regression Tasks

For regression tasks, Ridge regression demonstrated the best generalization with $R^2 = 0.007$ on the test set, despite XGBoost achieving higher validation performance ($R^2 = 0.048$). This suggests XGBoost may have overfit to training data patterns. Deep learning models (LSTM, BiLSTM, GRU) achieved competitive MAE scores (≈ 0.511) but near-zero R^2 values, indicating difficulty capturing the variance in returns.

5.2.2 Classification Tasks

For direction prediction, BiLSTM achieved the highest F1-score (66.0%) by predicting positive movement for nearly all samples (100% recall). While this exploits class imbalance, it lacks practical utility. XGBoost provided the most balanced performance with 55.4% accuracy and 57.4% ROC-AUC, representing a meaningful improvement over random chance (50%).

5.3 Deep Learning Observations

The LSTM variants showed mixed results:

- **BiLSTM:** Achieved best MAE and F1-score but exhibited extreme prediction behavior
- **Standard LSTM:** Failed to learn meaningful classification patterns (0% F1)
- **GRU:** Comparable performance to LSTM with simpler architecture

These findings suggest that recurrent architectures may require longer sequences or additional regularization for cryptocurrency prediction tasks.

6 Conclusions

6.1 Key Findings

1. **Sentiment-Price Relationship:** News sentiment shows statistically significant correlation with price returns. Sentiment mean ($r = 0.12$) and polarity mean ($r = 0.10$) are the strongest predictors. Negative sentiment correlates inversely with returns ($r = -0.086$).
2. **Model Performance:**
 - *Regression:* Ridge regression provides the best generalization ($R^2 = 0.007$)
 - *Classification:* XGBoost achieves the best balanced performance (55.4% accuracy, 57.4% AUC)
 - *Deep Learning:* BiLSTM excels at recall (100%) but lacks precision, yielding high F1 (66.0%)
3. **Statistical Significance:** Hypothesis tests confirm significant difference between positive and negative sentiment conditions. Effect size is small (Cohen's $d = 0.005$), suggesting limited practical predictability.

6.2 Limitations

- Short-term data collection period limits long-term pattern analysis
- Limited news article volume (301 articles) may not capture full sentiment spectrum
- LSTM models show signs of overfitting/underfitting on test data
- Cryptocurrency market efficiency may inherently limit predictability
- Single sentiment analysis approach (VADER/TextBlob) may miss nuanced sentiment

6.3 Recommendations

1. Expand data collection to longer time periods (6-12 months minimum)
2. Incorporate additional sentiment sources (Twitter, Reddit, Telegram)
3. Explore ensemble methods combining ML and DL approaches
4. Consider market regime-dependent models (bull/bear markets)
5. Implement real-time prediction pipeline for live validation
6. Investigate attention mechanisms and transformer architectures

A Configuration Parameters

Table 17: Experiment Configuration

Parameter	Value
Cryptocurrencies	BTC, ETH, SOL, ADA, DOT
Sentiment Threshold (Positive)	0.05
Sentiment Threshold (Negative)	-0.05
Prediction Window	7 days
Lookback Window	30 days
Train Ratio	0.6
Validation Ratio	0.2
Test Ratio	0.2
Random State	42

B LSTM Hyperparameters

Table 18: LSTM Model Hyperparameters

Parameter	Value
Sequence Length	24 hours
Hidden Units	[128, 64]
Dropout Rate	0.2
Learning Rate	0.001
Batch Size	32
Max Epochs	100
Early Stopping Patience	15
L2 Regularization	0.01
Optimizer	Adam

C Feature List

Price Features:

- price, volume, market_cap
- price_return, volume_change
- price_ma_24h, price_ma_168h
- volatility_24h, volatility_168h
- momentum_24h

Sentiment Features:

- sentiment_mean, sentiment_min, sentiment_max, sentiment_std
- polarity_mean, subjectivity_mean
- positive_mean, negative_mean, neutral_mean
- article_count

Report generated from experiment: crypto_sentiment_research_2025

Timestamp: 20251216_110817