ITM Web of Conferences **84**, 02008 (2026) https://doi.org/10.1051/itmconf/20268402008
ATCIDS 2026

Dynamic System for Multi-Factor Cryptocurrency Forecasting: Integrating Emotion Weighting and Herding Effects on Binance Top 10

Qian Chong[1*], *Boyuan* Zheng[2], *Baicheng* Zhou[3]

[1]Shanghai Mingsui Creative School, 200120 Shanghai, China
[2]Albert College, Belleville City, Canada
[3]School of Ethnology and Sociology, Minzu University of China,100000 Beijing, China

Abstract. Cryptocurrency markets are notorious for their extreme volatility and social sentiment susceptibility, making them challenging for accurate price forecasting. To tackle these problems, this paper proposes Dynamic Sentiment Engine– a predictive model for the top ten cryptocurrencies on Binance including Bitcoin and Ethereum. The system proposes a multi-dimensional solution which incorporates real-time multi-source social media sentiment analysis results and extends herd behavior tracking to financial markets. The innovation lies in the adaptive weighting scheme which adjusts the relative influence between collective investor sentiment and emerging herd dynamics in a market-regime calibrated way. By continuously adapting the relative influence between sentiment and herd effects in stable or turbulent market regimes, the prediction robustness is greatly improved over traditional univariate or static models. The system provides traders with a novel decision-support tool to operate in the highly volatile digital asset market, and contributes to behavioral finance by providing a measurable model to explore sentiment-herd interdependencies.

1 Introduction

Since the launch of Bitcoin (BTC) in 2009, cryptocurrencies have evolved into an integral component of the global financial market. Bitcoin is performing well in the first half of 2025 according to the research department of Binance. Bitcoin is showing maturity as a key macro asset. The market cap of Bitcoin is expected to hover around $2 trillion, and the cryptocurrency dominance may touch 65.1% – the highest in over 4 years – which means that BTC will be the world’s best-performing asset. Even in a highly volatile macro environment, BTC is likely to deliver a double-digit return [1].

However, Cryptocurrencies are still viewed as highly speculative assets. One of the reasons is that the cryptocurrency market operates 24/7, which leads to high price volatility.

* Corresponding author’s email: simelrick4444@gmail.com

High volatility requires robust forecasting system, especially for short-trading strategy investors [2].

Technologically, the crypto price prediction has evolved from traditional linear statistical models to machine/deep learning nonlinear models. Previous studies have employed ARIMA to model the temporal information but failed to consider the non-stationary nature of crypto [3]. In recent years, deep learning models such as LSTM and GRU have become mainstream for modeling long-term temporal dependencies, and hybrid models combining multi-technical advantages have emerged to improve the prediction accuracy [4].

The core strength of the hybrid prediction model lies in the ability to jointly capture two types of market dynamics. Market sentiment transmission effect – social media public opinion has asymmetric diffusion characteristics, and the information of panic will more likely trigger excessive behavior of investors and cause deviations in pricing [5]; Herding behavior – among the top 10 most liquid cryptocurrencies on Binance, retail investors' clustered trading behavior can easily trigger a self-reinforcing price spiral and enlarge short-term market volatility. This ability to model non-linear factors provides the mechanistic basis for constructing a prediction system that can adapt to the real market.

As for the selection and comparison of algorithm models, the systematic research results of Salehi shows that: in the prediction based on closing prices of Ethereum (ETH) and Binance Coin (BNB), the hybrid model ANFIS achieves significantly better performance compared with mainstream machine learning algorithms. The design of fuzzy logic and neural networks in the hybrid model (compared with SVR linear kernel/LSBoost/ANN) shows the best performance in the RMSE and R^2 metrics, which further verifies that the hybrid architecture can better capture the non - linear dynamics of cryptocurrencies [6].

Tiwari et al. proposed a stacked LSTM model based on Particle Swarm Optimization (PSO). Not only can it capture deep contextual semantic features in a multi - layer LSTM architecture, but it also uses the PSO algorithm for global optimization of the hyperparameters. It significantly improves the generalization ability and prediction accuracy of the model. Compared with traditional ensemble models, it achieved better performance in Bitcoin tweet sentiment analysis and price prediction tasks [7].

New research trends have started to transcend simply optimizing technical models and instead have delved deep into the internal financial and behavioral factors that cause price fluctuations in the market.

Duygun et al. discovered that during the global financial crisis, both real herding and pseudo herding were present in the highly volatile market environment of both US and European stock markets, and this showed a significant asymmetry [8].

The latest research on Cao et al. first introduced the important market behavior theory of the "herd effect" into the field of cryptocurrency price prediction. In addition to proving that during the market, not only are real (driven by non - fundamental information) herd effects exist but pseudo (driven by fundamental information) herd effects also exist, they also constructed corresponding quantitative measurement indices [9].

For example, the review article by John et al. titled "Overview of Research on Cryptocurrency Price Prediction from 2014 to 2024" makes a comprehensive summary of the research process in the field of cryptocurrency price prediction from 2014 to 2024. In this study, it is clearly stated that the main parameters that affect prices are historical prices and trading volumes, technical indicators (moving averages, RSI), on - chain blockchain data (miner revenues, transaction fees), and social media sentiment (mainly from Twitter and Reddit) [10].

This study explores the limitations of existing cryptocurrency prediction models' treatment of market sentiment noise and herd effect as disturbances but not as their driving forces, which have been regarded as interfering factors by traditional cryptocurrency prediction models for a long time. Based on the dynamic analysis framework of coupling mechanism of sentiment weighting and herd effect, this study attempts to enhance the accuracy of identification for market vulnerability, extend the duration of price inflection point warning window, and help monitoring the transmission path of systemic risks by quantifying the resonance effect of spread gradient of social media public opinion on retail cluster trading. It provides new theoretical support for the prediction method transition of top - tier crypto - assets (Binance Top 10) from linear analysis to ecological simulation.

2 Research methods

2.1 Prediction Framework: Dynamic Sentiment Engine (DSE)

2.1.1 Core Predictor: Bi-LSTM Model

Bi-LSTM Model. As shown in Figure 1, a bidirectional long short-term memory network (Bi-LSTM) is chosen as core predictor. The bidirectional dependencies in time-series data ("past → present" and "present → future") can be learned by the forward LSTM and backward LSTM, respectively. Compared with traditional unidirectional LSTM processing method, the bidirectional long short-term memory network can capture bidirectional dependencies [9]. The improvement method of LSTM is referred to Hao et al. The LSTM is further optimized based on the characteristics of cryptocurrency data as follows: Bi-LSTM (L=24, h) → Dropout (p=0.2) → Dense (1, Linear/Sigmoid). Input time step L=24: Based on experimental validation, 24 hours (with 1-hour granularity) of data can balance the accuracy and efficiency while avoiding overfitting [6]. Number of hidden layer neurons h: 64 neurons are adopted for large-market-cap coins (BTCUSDT/ETHUSDT) and 32 neurons are adopted for small-market-cap coins (SOLUSDT/ADAUSDT) according to the volatility characteristics of different coins (smaller-market-cap coins are more volatile and reducing neurons can avoid overfitting) [2]. Output layer: The Linear activation function outputs price value and the Sigmoid function outputs the probability of price movement (>0.5 represents upward movement and <0.5 represents downward movement) to meet the requirement of two-dimensional prediction [7].

2.1.2 DSE Adaptive Weighting Mechanism

DSE is composed of "multi-factor input, core predictor, dynamic weighting module, and output layer." The innovation of dynamic weighting module lies in adjusting the weights according to the real market situation. The "herding effect quantification logic" is referred to Cao et al. and "sentiment-volatility correlation" is referred to Biswas et al.

Market state classification: According to the volatility rate of the last 24 hours: volatility rate >2%: volatile period (e.g., release of policies, occurrence of major events); volatility rate <1%: stable period (no major events, low trading volume) [9].

Weight allocation: Volatile period: ωt = [0.1 (ARIMA), 0.3 (XGBoost), 0.2 (ANFIS), 0.4 (Bi-LSTM)] ↑, increase the weight of Bi-LSTM to improve the time-series capture capability [7]; Stable period: ωt = [0.1 (ARIMA), 0.25 (XGBoost), 0.35 (ANFIS), 0.3 (Bi-LSTM)] ↑, increase the weight of ANFIS to improve the nonlinear fitting [6].

Extreme sentiment adjustment: When the sentiment score < -50 (very panic) or > 50 (very greedy), the weight contribution of Ht is increased by 10% more to deal with “emotion-driven irrational trading” [5].

2.2 Dataset and Feature Engineering

2.2.1 Data Sources Trading data

Through the Python connection to the Binance API, we collected 1-hour K-line data (Jan 2023–Jun 2024, 12,960 data per coin), and 1-day K-line data (540 data per coin) for the top 10 coins (top 10 coins by 24-hour trading volume). Fields: open price, close price, trade amount. Data completeness >99.5% [1].

Sentiment data: Based on the social media data collection method proposed by Tiwari et al., we scraped the coin-related tweets (50 tweets per coin per day) via Twitter API and calculated the sentiment score (St) between -100 and 100 using the VADER model. The higher the value, the stronger the bullish sentiment [7].

Herding data: We calculated the herding effect (Ht) based on the CSAD index formula. CSADt = (1/N) Σ |ri,t−rm,t|(here ri,t is the return rate of coin i at time t, and rm,t is the market average return rate). The formula results were normalized to between 0 and 100 (smaller values represent a stronger herding effect) [9].

2.2.2 Feature Preprocessing

Missing value imputation: We filled K-line data using "forward filling" (since cryptocurrency trading is continuous, missing values usually appeared to be temporary interface problems). We filled sentiment data using a "7-day rolling average" to avoid the influence of daily outliers [6].

Normalization: We Min-Max normalized all features (price, technical indicators, St, Ht) into [0, 1] using xscaled=(x−xmin)/(xmax−xmin) to eliminate the influence of scales that might impact model training [2].

Data splitting: We divided the final dataset into training (Jan 2023 – Feb 2024) and testing (Mar 2024 – Jun 2024) sets based on time series to avoid data leakage that might result from shuffling the time sequence [3]

3 Research results

3.1 Statistical Characteristics of the Dataset

The core statistical information for the top 10 Binance coins is as follows, validating Biswas and Sharma's conclusion that "smaller-market-cap coins exhibit higher volatility" while also reflecting the market capitalization differences in herding effects, as detailed in table 1:

Table 1. Statistical Summary of 1-Hour Data for Top 10 Binance Coins

Coin	Data value (1h)	Avg. 1h Volatility	Avg. Sentiment Score (St)	Avg. Herding Index (Ht)
BTCUSDT	12960	1.2%	8.7	15.3
ETHUSDT	12960	1.8%	5.2	18.6
BNBUSDT	12960	2.5%	3.1	22.1
SOLUSDT	12960	3.2%	-1.2	25.4
XRPUSDT	12960	2.8%	0.5	23.7

ADAUSDT	12960	3.0%	-0.8	24.2
DOGEUSDT	12960	3.5%	2.1	26.8
DOTUSDT	12960	2.7%	1.3	22.9
AVAXUSDT	12960	2.9%	-1.5	23.5
LINKUSDT	12960	2.6%	0.9	21.8

3.2 Model Performance Comparison

3.2.1 Single-Coin Performance (Using BTCUSDT as an Example)

As the coin with the largest market capitalization and the lowest volatility, BTCUSDT's performance is representative. The results show that DSE significantly outperforms single models (table 2):

RMSE (USD)：Root Mean Square Error, measuring the average deviation between predicted and actual prices (lower values indicate better accuracy).

MAE (USD)：Mean Absolute Error, reflecting the average absolute deviation (lower values indicate better accuracy).

R^2：R-squared, a goodness-of-fit indicator (values closer to 1 indicate better alignment with real data).

Weighted F1-Score： A balanced metric for classification tasks (predicting price direction), weighted by coin trading volume to avoid bias (higher values indicate better performance).

Table 2. Model Performance Comparison for BTCUSDT (Test Set)

Model / System	**RMSE（USD）**	**MAE（USD）**	**R^2**	**Weighted F1-Score**
ARIMA（Price Only）	380.5	298.7	0.78	0.72
XGBoost（Multi-Feature）	287.2	215.3	0.85	0.81
ANFIS（Multi-Feature）	263.4	192.1	0.87	0.83
Bi-LSTM（Multi-Feature）	245.3	189.6	0.92	0.88
DSE (Dynamic Weighting)	230.5	178.2	0.94	0.91

3.2.2 Multi-Coin Average Performance

DSE demonstrates stable performance across coins of different market capitalizations, with average performance significantly surpassing that of single models, validating its adaptability to multiple coins (table 3):

Table 3. Average Performance Comparison for Top 10 Binance Coins

Model / System	**RMSE (USD)**	**R^2**	**Weighted F1-Score**	**Accuracy Loss in High Volatility Periods**
ARIMA	156.8	0.75	0.69	25%
XGBoost	102.3	0.83	0.79	18%

ANFIS	89.7	0.86	0.82	16%
Bi-LSTM	78.5	0.89	0.87	12%
DSE	72.1	0.91	0.90	8%

3.3 Stratified Stability Testing

3.3.1 Performance Across Different Sentiment Ranges

DSE significantly outperforms the single Bi-LSTM model under extreme sentiment conditions, aligning with Biswas and Sharma's conclusion that "sentiment influences volatility." Its dynamic weighting mechanism effectively mitigates sentiment-related interference (table 4):

Table 4. Comparison of Weighted F1-Scores Across Different Sentiment Ranges (Average)

Model / System	High Panic ($St < -50$)	Neutral ($-20 \leq St \leq 20$)	High Greed ($St > 50$)	Range Volatility Difference
Bi-LSTM	0.82	0.88	0.84	6%
DSE	0.86	0.90	0.88	4%

3.3.2 Performance Across Different Volatility Ranges

DSE demonstrates strong risk resistance during high-volatility periods, validating the effectiveness of Cao et al.'s "dynamic adaptation to herding effects" (table 5):

Table 5. Comparison of R^2 Across Different Volatility Ranges (Average)

Model / System	High volatility (>2%)	Medium volatility (1%-2%)	Low volatility (<1%)	Range Volatility Difference
Bi-LSTM	0.84	0.89	0.92	8%
DSE	0.88	0.91	0.93	5%

4 Conclusion

Our proposed "Dynamic Sentiment Engine" has demonstrated notable potential in bridging social media sentiment analysis and market herd behavior tracking and made a breakthrough in the methodology of predicting price fluctuations of major Binance-listed cryptocurrencies. The adaptive weighting scheme of "Dynamic Sentiment Engine" could adjust the influence of sentiment and herd effect adaptively according to different market conditions and improved model robustness when the market was highly volatile. However, there are still many theoretical and practical challenges existing in this work which would be further explored in the future.

Firstly, there were still large amounts of noise and bias existing in social media data (such as Twitter, Reddit). Spammers, hype accounts, manipulative bots and other suspicious accounts would bring interference to the extraction of correct sentiment information. Although multi-source fusion could alleviate this problem to some extent, when the market was extreme, the overall emotional bias would still be extremely intense and would cause mistaken trend prediction. Another big challenge was that herd behavior was also hard to measure objectively. Most indicators used in this work were proxies for trading concentration and liquidity shifts (such as trading volume, trading hash rate, market maker ratio, and order book liquidity). These indicators could not reflect investors' real intention directly. When the price was manipulated by a large number of trades from the same group of investors, it was still challenging to model the true herding behavior separately from coincidental mass action at such high frequencies. Besides, the computational efficiency also limited the practical application. It would take a large number of computational resources to process unstructured text data, on-chain data and real-time market data on a large scale. The latency problem might appear and affect the practical application of the model. At last, the model could not fully incorporate external macro shocks (such as regulatory shocks, exchange failures, economic crises, etc.) which would cause structural breaks in the model and limit the generalization of the model in black-swan events.

In the future, we will further explore the following research directions: Firstly, we will explore cross-platform sentiment consensus mechanisms with more sources (such as Telegram, Discord, Weibo, etc.) to better represent the global investor sentiment with graph neural networks (GNNs). Secondly, we will explore interpretable hybrid architectures with attention mechanisms, causal inference, and temporal models to improve model transparency. Thirdly, we will embed reinforcement learning agents to optimize the influence of adaptive weighting. Finally, we will further explore applications on DeFi and NFT markets through domain adaptation.

In summary, this work proposed a new paradigm for cryptocurrency forecasting and opened new research directions for behavioral finance research in digital economies. With the continuous improvement of data quality and modeling methods, the participation of sentiment-aware prediction systems in the future financial market will become more and more important.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Binance Research Institute, Binance Research. 2025 Mid-Year Research Report. (2025)
2. T. Shrivastava, B. Suleiman, A. J. Sachit et al. Benchmarking modeling architectures for cryptocurrency price prediction using financial and social media data.Social Network Analysis and Mining,15(1), 93-93, (2025)
3. Y. Hua. Bitcoin price prediction using arima and lstm, E3S Web of Conf, (2020)
4. W. Hao, X. Sun, C. Wang, H. Chen, and L. Huang, A hybrid EMD-LSTM model for nonstationary wave prediction in offshore China, Ocean Engineering, vol. 246, p. 110566, (2022)
5. P. Biswas & C. Sharma. Does fear sentiment drive cryptocurrency volatility? Evidence from Google trends data. Finance Research Letters,86(PB),108510-108510.https://doi.org/10.1016/J.FRL.2025.108510, (2025)
6. S. Salehi. Comparative Analysis of Machine Learning Techniques for Cryptocurrency Price Prediction. Journal of Internet of Systems, 48(2), 7. DOI: 10.31341/jios.48.2. (2024)

7. D. Tiwari, B. Singh Bhati, B. Nagpal. A swarm-optimization based fusion model of sentiment analysis for cryptocurrency price prediction.Scientific Reports,15(1),8119-8119.https://doi.org/10.1038/S41598-025-92563-Y (2025)
8. M. Duygun, R. Tunaru, & D. Vioto. Herding by corporates in the US and the Eurozone through different market conditions. Journal of International Money and Finance, 110, 102311, (2021)
9. G. X. Cao, M. J. Ling, J. W. Wei& C. Chen. Dynamic Market Behavior and Price Prediction in Cryptocurrency: An Analysis Based on Asymmetric Herding Effects and LSTM. Computational Economics,65(6),1-36. https://doi.org/10.1007/S10614-024-10676-4, (2024)
10. L. David. J. S. Binnewies & B. Stantic. Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions. Forecasting, 6(3), 637-671.https://doi.org/10.3390/FORECAST6030034, (2024)