Comprehensive EDA Analysis Report

AI-Powered Cryptocurrency Trading Signal Analysis

Dataset Information

File: processed_crypto_data_20250918_214215.csv

Records: 13,507

Unique Coins: 93

Date Range: 2025-08-21 12:00:00 to 2025-09-18 16:00:00

Analysis Date: 2025-09-19 12:33:12

1. Executive Summary

The dataset comprises of 13,507 records across 93 meme coins from the period of 21st August to 18th September 2025. The target variable is imbalanced with a ratio of 2.67, favoring HOLD signals. The average price volatility is 5.78%, with a maximum of 93.56% and 8337 outliers. The technical indicators show a mean RSI of 47.7, with 5.8% oversold and 4.9% overbought. The data quality is good with strong correlations between several features, but the class imbalance and high volatility present challenges.

2. Market Behavior Analysis

The meme coin market shows high volatility with an average range of 5.78% and a maximum range of 93.56%. The top coin has a dominance of 2.5%, and the top 5 coins account for 8.3% of the data. This suggests a highly fragmented and volatile market. We recommend using volatility-adjusted position sizing, such as the ATR-based method, to manage risk.

3. Technical Indicator Effectiveness

The RSI has a mean of 47.7, with 5.8% of observations in the oversold region and 4.9% in the overbought region. This suggests that the market is mostly in a neutral state. The EMA and volume indicators show that 39.4% of prices are above the EMA and 37.9% of volumes are above average. These indicators can be used to generate trading signals when combined with other features. For example, a BUY signal could be generated when the price is above EMA and RSI is oversold.

4. Target Variable Analysis

The target variable is imbalanced with a ratio of 2.67, favoring HOLD signals. This imbalance can be addressed using techniques such as SMOTE or ADASYN for oversampling the minority classes, or using class weights in the machine learning model.

5. Feature Engineering

New features can be created from the existing ones, such as the ratio of close price to EMA, or the difference between the current volume and the 20-period average volume. Feature selection can be done using methods like recursive feature elimination or feature importance from a random forest model.

6. Model Development Strategy

A multi-class classification model can be developed using algorithms like XGBoost or LightGBM. The model should be validated using a time-based split to avoid lookahead bias. Evaluation metrics should include accuracy, precision, recall, and F1-score for each class.

7. Risk Assessment

The high volatility of meme coins presents a significant market risk. Overfitting is another concern due to the large number of features and class imbalance. Regularization techniques and early stopping can be used to mitigate overfitting.

8. Action Plan

The preprocessing steps include handling outliers, balancing the classes, and scaling the features. The development roadmap includes feature engineering, model development, validation, and tuning. Benchmarks should be set based on the performance of a simple baseline model, such as a logistic regression or a decision tree. Python libraries that can be used include pandas for data manipulation, matplotlib and seaborn for visualization, sklearn for preprocessing and model development, and keras for deep learning models. Code examples can be provided upon request.

Performance Expectations

Given the high volatility and market risk, a realistic expectation would be an accuracy of around 60-70% on the validation set. The model's performance should be continuously monitored and adjusted as necessary.

Key Statistics Summary

Label Distribution:
• SELL (0): 3,347 (24.8%)
• HOLD (1): 7,391 (54.7%)
• BUY (2): 2,769 (20.5%)

Price Volatility:
• Average Price Range: 5.78%
• Maximum Price Range: 93.56%

Technical Indicators:
• RSI Oversold Signals: 5.8%
• RSI Overbought Signals: 4.9%
• Price Above EMA: 39.4%

Market Structure:
• Top Coin Dominance: 2.5%
• Top 5 Coins Represent: 8.3% of data