Bitcoin Comprehensive Analysis Report

Machine Learning Price Prediction Framework

(FIXED VERSION - Overfitting Checks Applied)

Generated: 2025-11-15 21:49:37

⚠ WARNING: Current ML model failed overfitting checks

Only validated analyses are included in this report

Executive Summary

EXECUTIVE SUMMARY This comprehensive analysis presents a machine learning framework for predicting
Bitcoin price movements. IMPORTANT: This report only includes models that have passed rigorous
overfitting and data leakage checks. Key Findings: 1. Data Leakage Prevention - All features use
only past data (proper time series) - No same-day information leakage - Features are properly lagged
2. Overfitting Detection - Comprehensive checks for train/val/test gaps - Detection of suspiciously
high scores - Feature redundancy analysis 3. Model Validation - Only models passing all checks are
included - Realistic performance metrics - Proper confidence intervals Note: If the current ML model
shows suspiciously high R^2 scores (>0.95), it indicates either: - Data leakage (using future
information) - Trivial prediction task (e.g., predicting price from previous price) - Overfitting

Such models are excluded from this report.

Overfitting Analysis

OVERFITTING AND DATA LEAKAGE ANALYSIS Comprehensive Overfitting Check Results: Overall Status: FAIL
- Severe Overfitting No Data Leakage Detected Overfitting Detected (Severity: severe) -
Train-Val gap (100.00%) exceeds threshold (15.00%) - Train-Test gap (100.00%) exceeds threshold
(15.00%) Recommendations: - Increase regularization - Reduce model complexity - Add more training
data - Use feature selection to remove redundant features - Remove or combine highly correlated

features

Feature Selection

FEATURE SELECTION ANALYSIS Initial Features: 31 Final Features: 10 Reduction: 21 features removed
Selected Features by Category: FRED: 7 - CPIAUCSL, CPI_YoY, M2_Growth_Rate, DGS10, Net_Liquidity,
FEDFUNDS, UNRATE Technical: 3 - sma_50, price_sma200_ratio, volatility_20d
Feature Selection Steps:
Remove constant features: Removed 3 features Remove highly correlated features: Removed 13 features
Remove cross-category redundancy (FRED vs Technical): Removed 0 features
Mutual information

selection: Removed 0 features

ML Model Selection

MACHINE LEARNING MODEL SELECTION Current ML model failed overfitting/data leakage checks. Status:
Model failed checks: FAIL - Severe Overfitting The model was excluded from this report because: -
Suspiciously high R^2 scores (>0.95) indicate overfitting or data leakage - Model may be using future
information or trivial features - Predictions would not be reliable
Recommendations: - Use only
lagged features (past data only) - Predict returns instead of absolute prices - Remove
redundant/highly correlated features - Increase regularization - Use proper time series

cross-validation