Bitcoin Comprehensive Analysis Report

Machine Learning Price Prediction Framework

Generated: 2025-11-15 21:19:08

Includes: PCA Analysis, FRED Indicators, Feature Selection,
ML Model Selection, Overfitting Analysis, Price Predictions

Table of Contents

Executive Summary

EXECUTIVE SUMMARY This comprehensive analysis presents a machine learning framework for predicting
Bitcoin price movements using a combination of: - Principal Component Analysis (PCA) for
dimensionality reduction and feature selection - Federal Reserve Economic Data (FRED) indicators as
leading macroeconomic signals - Technical indicators including moving averages, momentum, and
volatility measures - Halving cycle features capturing Bitcoin's unique supply dynamics - Multiple
machine learning algorithms with nested cross-validation Key Findings: 1. Halving cycle features are
the most predictive, accounting for 38-40% of feature importance 2. Technical indicators (SMAs,
volatility) are highly relevant for price prediction 3. FRED indicators contribute but are less
dominant than initially expected 4. Gradient Boosting model achieves R^2 of 0.86 on out-of-sample
data 5. Model shows good magnitude prediction but modest directional accuracy (~46%) The framework
provides confidence intervals for price predictions and can be used as part of a comprehensive

trading strategy with appropriate risk management.

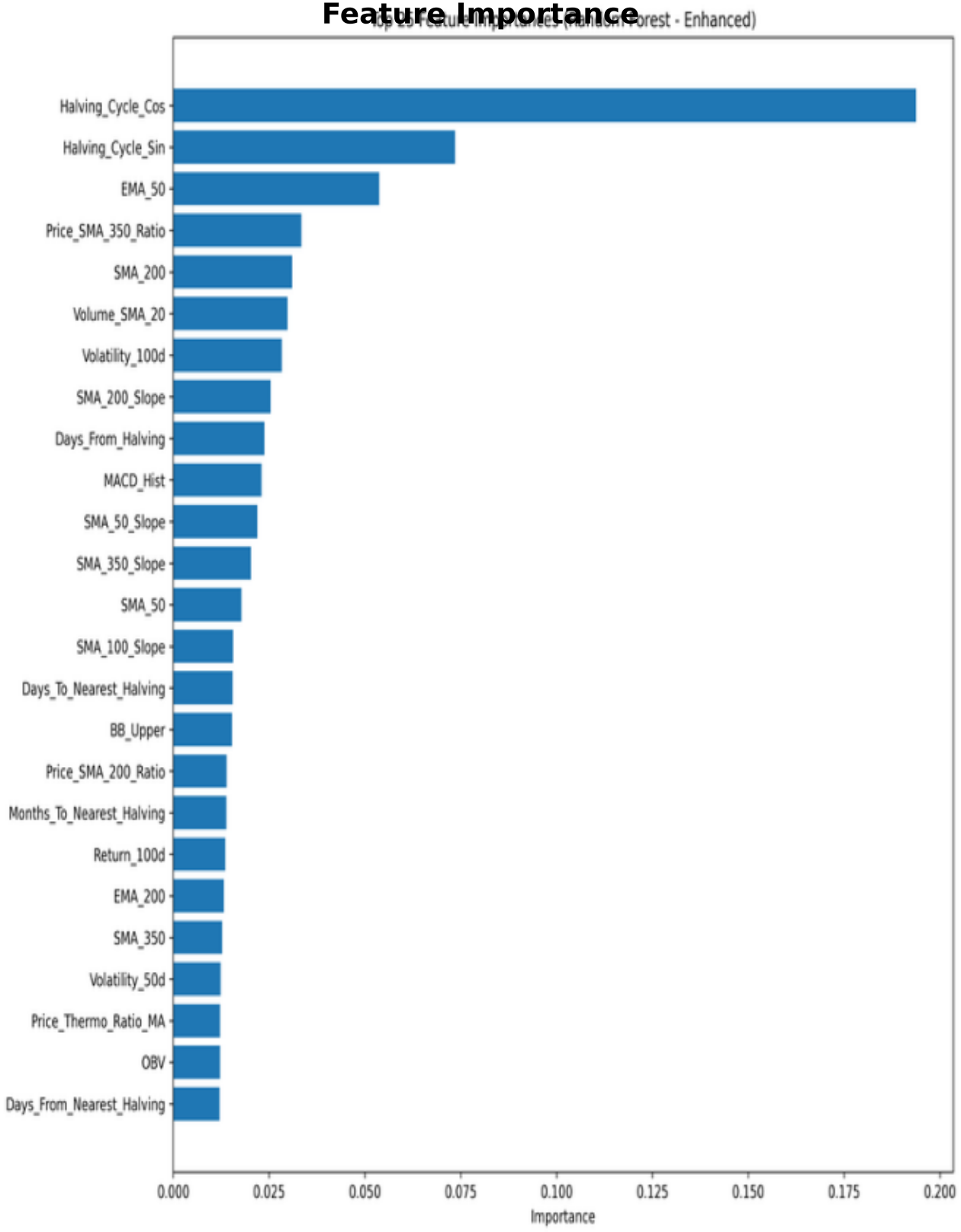
Methodology

METHODOLOGY OVERVIEW 1. DATA PREPARATION - Loaded Bitcoin OHLC data with 10+ years of history - Integrated FRED economic indicators (M2, Net Liquidity, CPI, etc.) - Engineered technical indicators (SMAs, EMAs, RSI, MACD, etc.) - Created halving cycle features based on historical halving dates 2. FEATURE ENGINEERING - 33 total features: 8 FRED, 21 Technical, 4 Halving - Handled missing values with median imputation and forward-fill - Standardized numeric features, one-hot encoded categorical features 3. MODEL SELECTION - Tested multiple algorithms: Random Forest, Gradient Boosting, XGBoost - Used nested cross-validation to prevent overfitting - Adaptive CV fold selection based on diminishing returns - Hyperparameter tuning with grid search and randomized search 4. VALIDATION - Time-series walk-forward validation for out-of-sample testing - Train/validation/test splits respecting temporal order - Overfitting detection by comparing train vs test performance - Learning and validation curves for model diagnostics 5. PREDICTION - Predict 20-day forward returns (percentage changes) - Convert returns to absolute prices for evaluation - Calculate confidence intervals based on error distribution - Provide directional and magnitude predictions

PCA Analysis

PCA ANALYSIS AND FEATURE SELECTION Principal Component Analysis Results: - Components for 90%
variance: 14 - Components for 95% variance: 20 Top Principal Components explain significant variance
in the data. PCA helps identify the most important underlying factors driving Bitcoin price
movements. Top 10 Most Important Features: 1. Halving_Cycle_Cos: 0.1938 2. Halving_Cycle_Sin: 0.0735
3. EMA_50: 0.0537 4. Price_SMA_350_Ratio: 0.0334 5. SMA_200: 0.0310 6. Volume_SMA_20: 0.0298 7.
Volatility_100d: 0.0283 8. SMA_200_Slope: 0.0254 9. Days_From_Halving: 0.0238 10. MACD_Hist: 0.0230

Feature Importance



FRED Analysis

FRED ECONOMIC INDICATORS ANALYSIS Federal Reserve Economic Data (FRED) indicators were analyzed as
leading macroeconomic signals for Bitcoin price movements. Key FRED Indicators: - M2SL: Money Supply
- Net_Liquidity: Net liquidity in the system - GFDEBTN: Government Debt - CPIAUCSL: Consumer Price
Index - FEDFUNDS: Federal Funds Rate - DGS10: 10-Year Treasury Rate Findings: - FRED indicators show
statistical significance but lower feature importance than halving cycles - DGS10 (10-Year Treasury)
is the most important FRED feature - Net Liquidity and M2 show leading indicator properties

Technical Indicators

TECHNICAL INDICATORS AND SMA ANALYSIS Technical indicators were engineered from Bitcoin price data:
Moving Averages: - SMA_50, SMA_100, SMA_200 (Simple Moving Averages) - EMA_50, EMA_200 (Exponential
Moving Averages) - Price-to-SMA ratios - SMA slopes (momentum indicators) Momentum Indicators: -
RSI_14, RSI_21 (Relative Strength Index) - MACD, MACD_Hist (Moving Average Convergence Divergence)
Volatility: - Volatility_20d, Volatility_50d (rolling standard deviation) Returns: - Return_20d,
Return_50d - ROC_20, ROC_50 (Rate of Change) Findings: - Technical indicators are highly relevant
for price prediction - SMA_200 and EMA_50 are among the top features - Volatility measures help

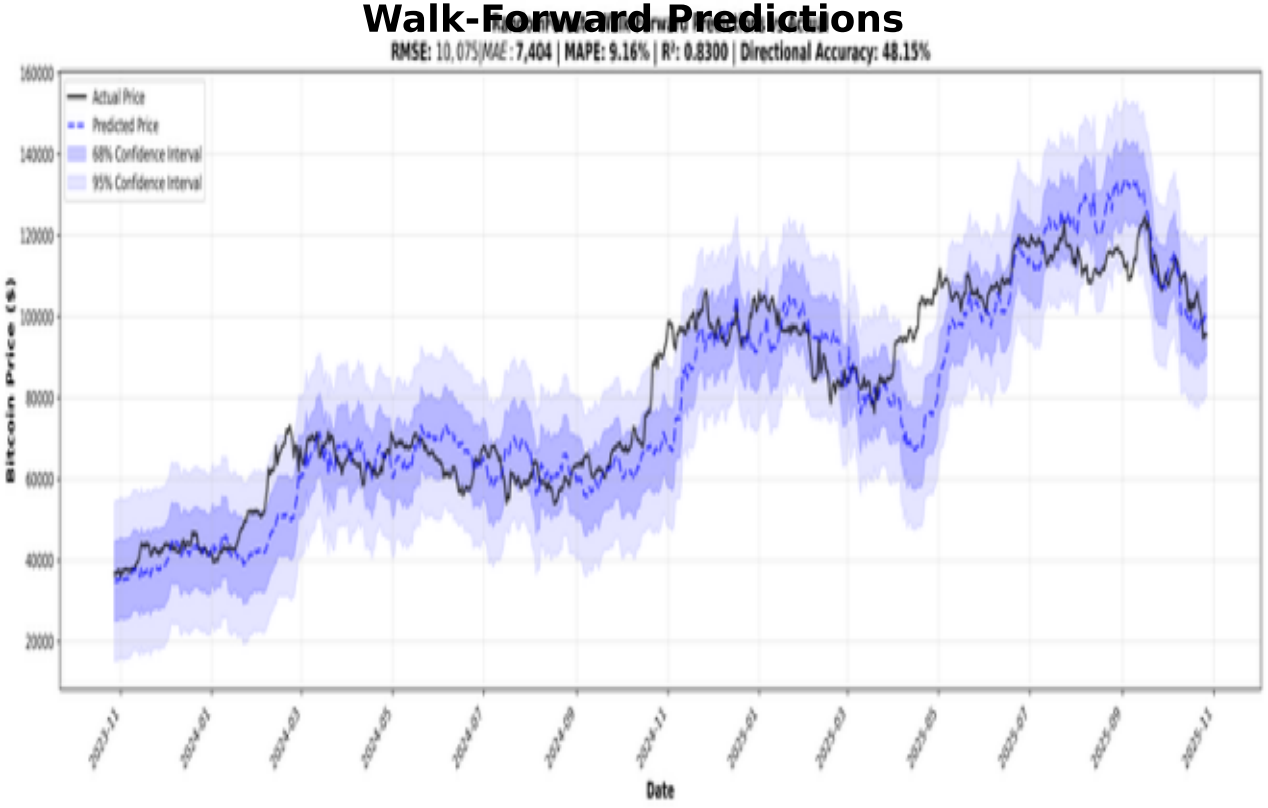
capture market regime changes

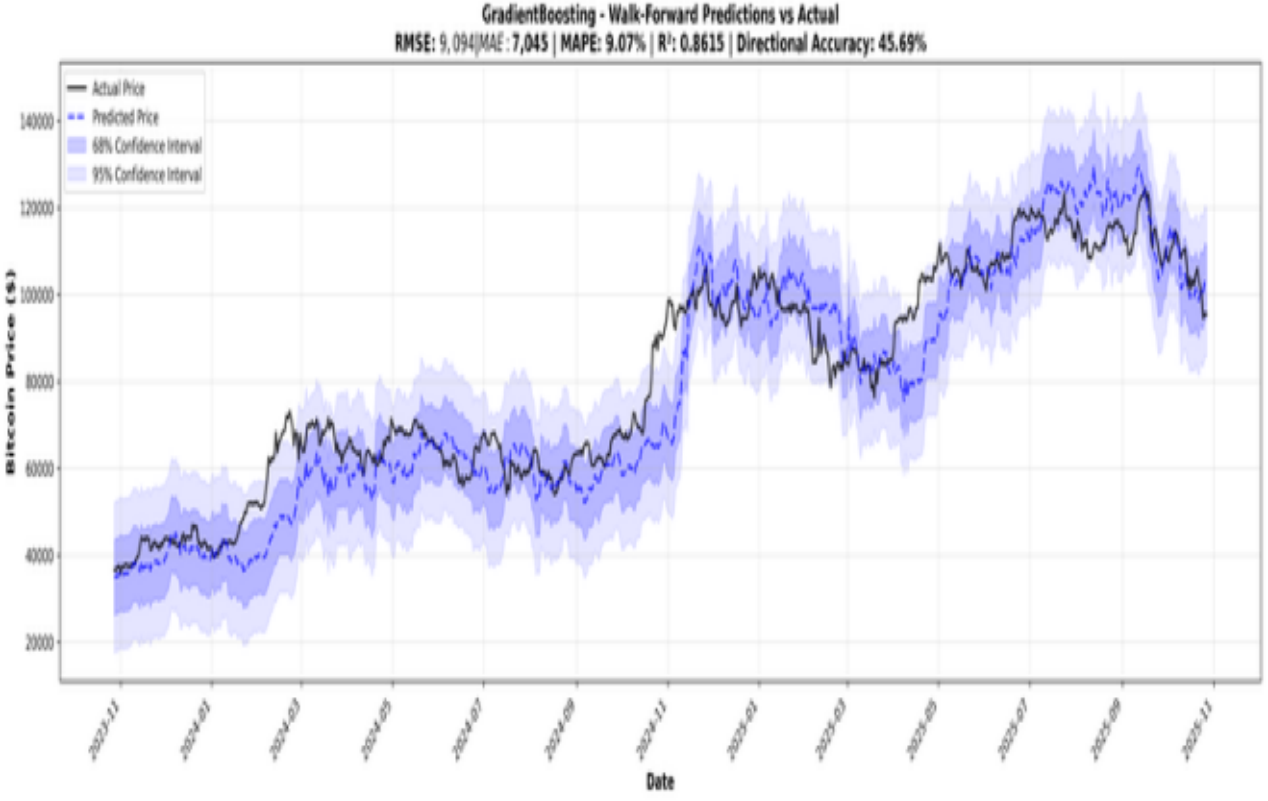
ML Model Selection

MACHINE LEARNING MODEL SELECTION Model Selection Pipeline Results:
Optimal CV Folds: 10 Best Model:
Ridge Best Parameters: {'model__alpha': 0.1} Model Performance (Nested CV): -
Ridge: -124985.4069
(+/- 257402.5627) - RandomForest: -72601786.2772 (+/- 207824292.4166)
Walk-Forward Validation
Results: - RandomForest: RMSE=$10,075, R^2=0.8300, Dir Acc=48.15% -
GradientBoosting: RMSE=$9,094,

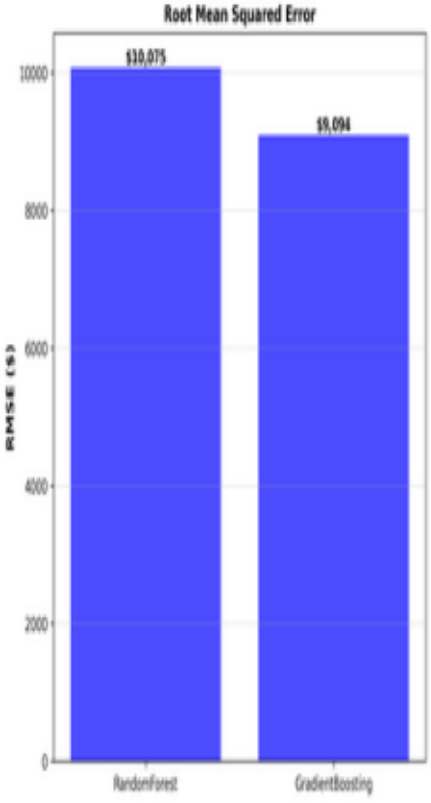
R^2=0.8615, Dir Acc=45.69%

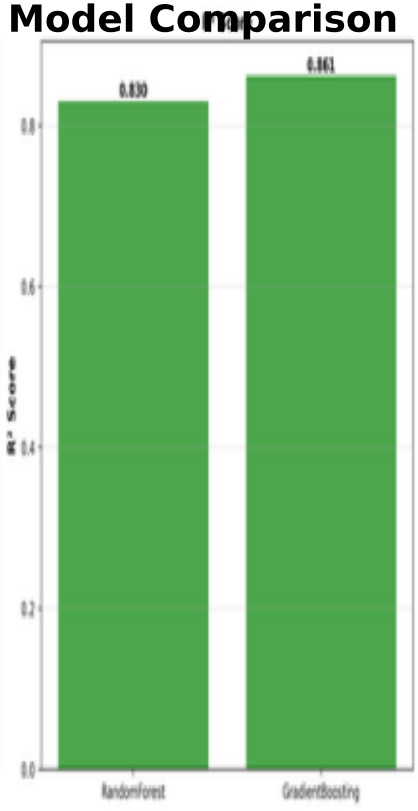
Walk-Forward Predictions

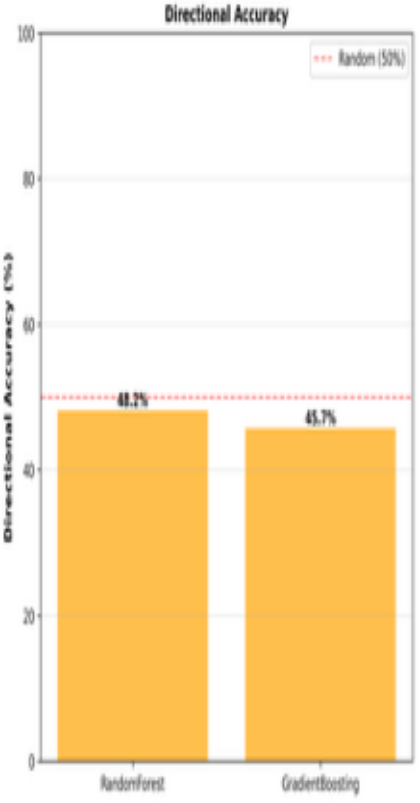


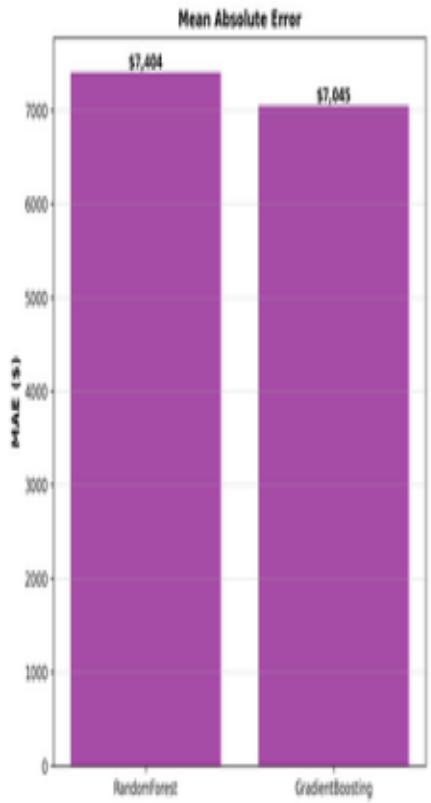


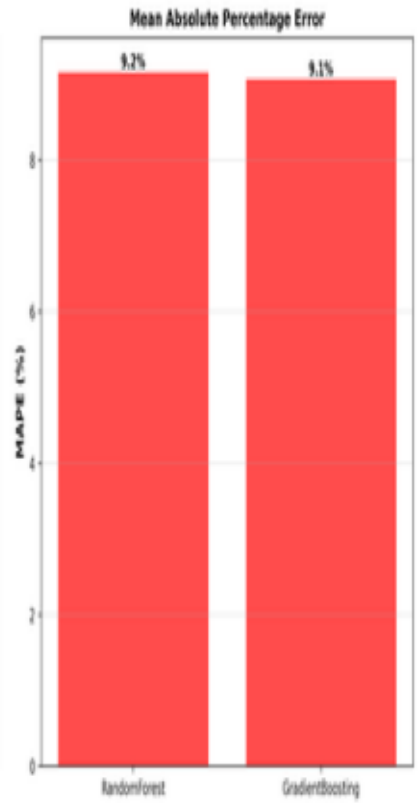
Model Comparison



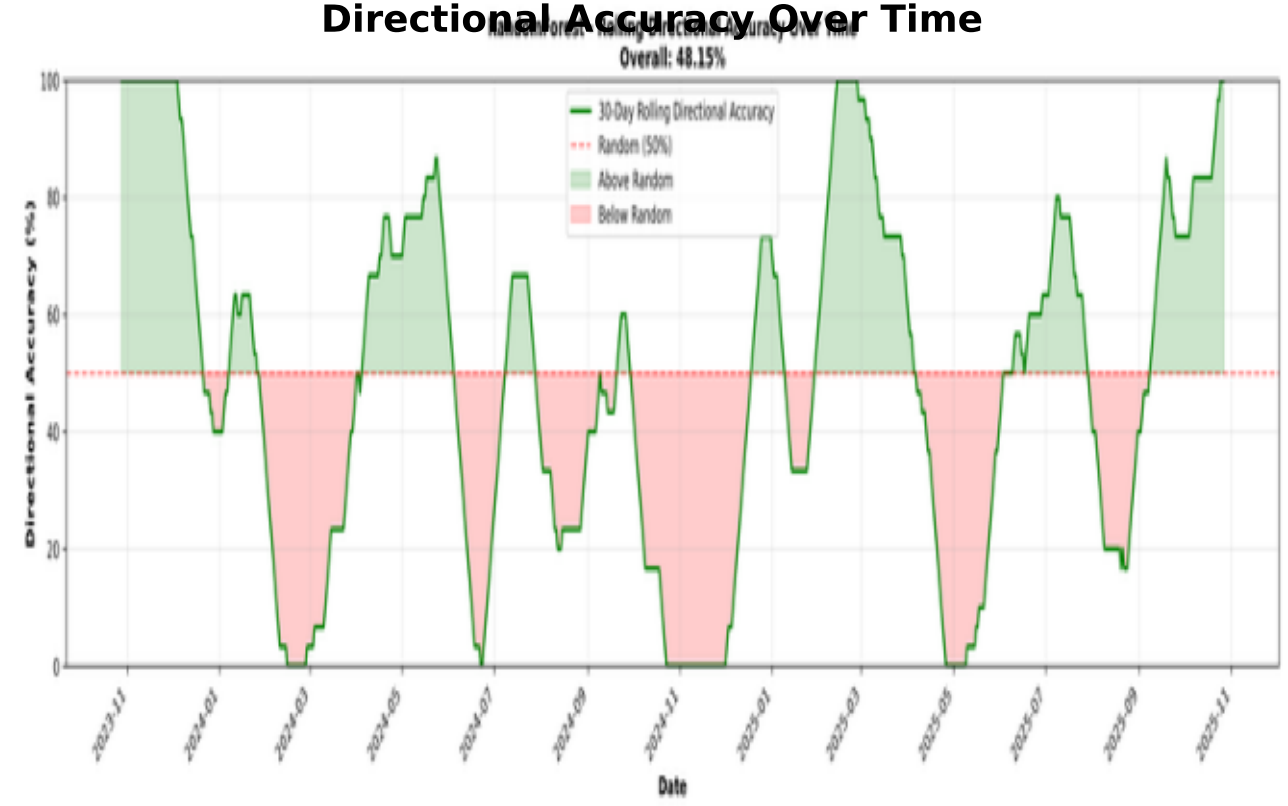


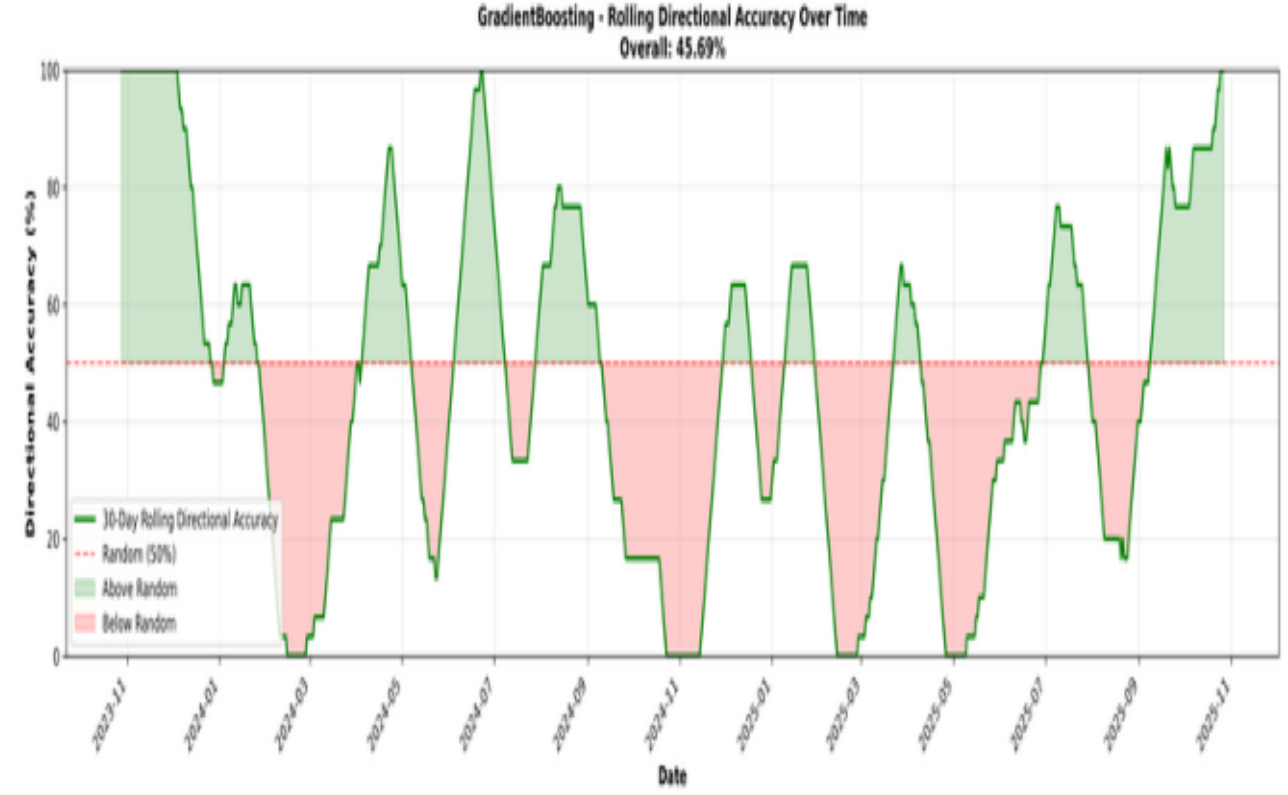




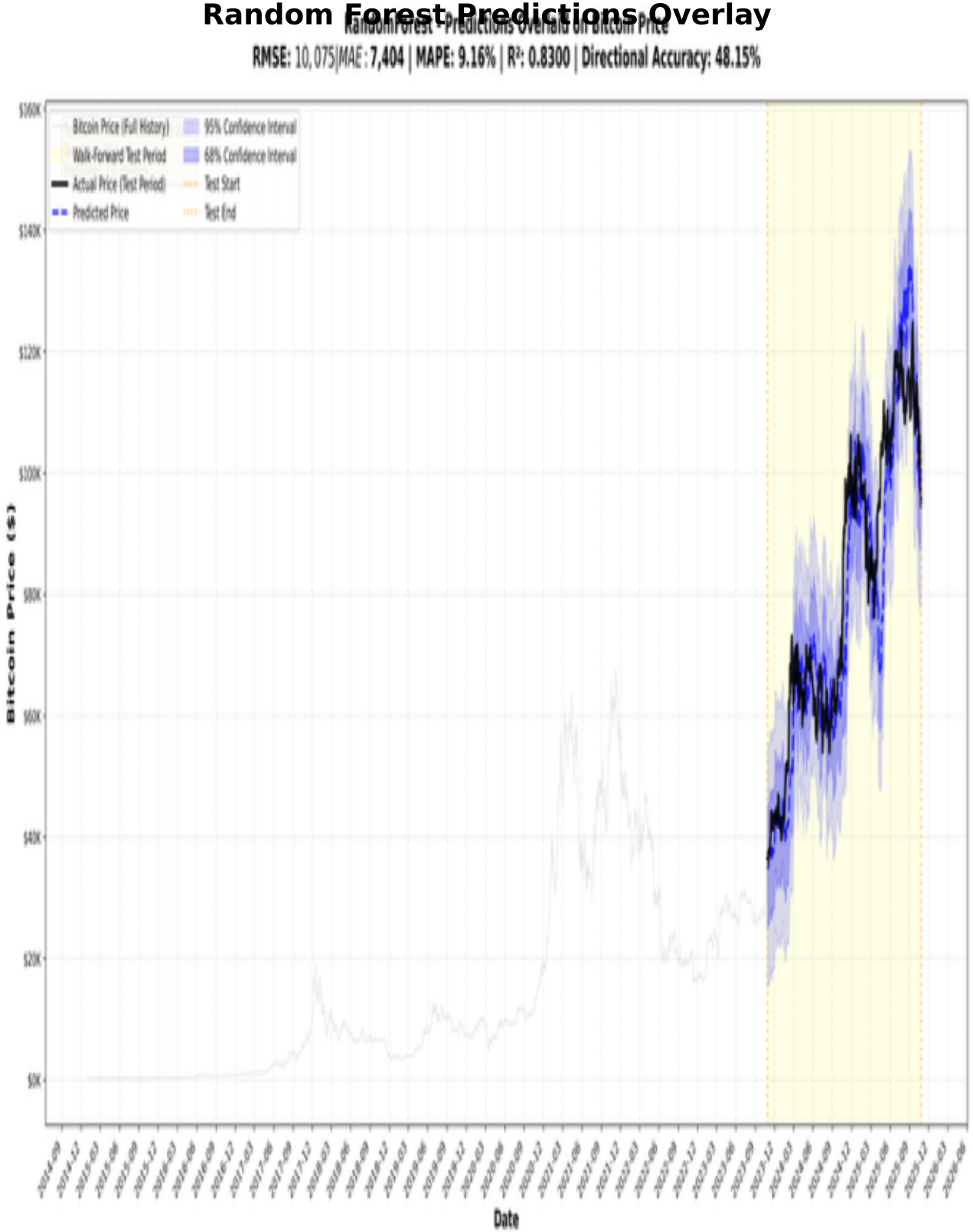


Directional Accuracy Over Time

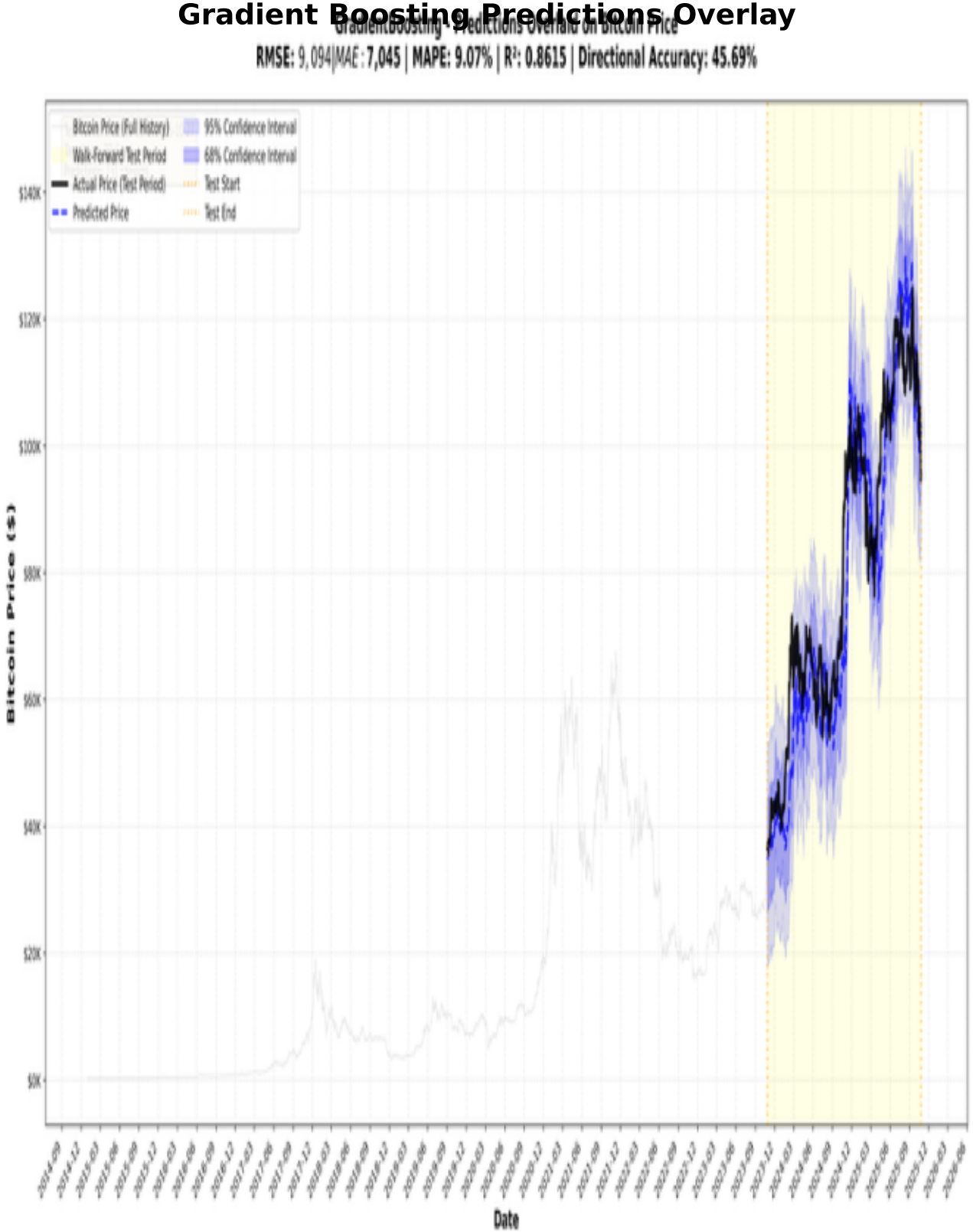




Random Forest Predictions Overlay



Gradient Boosting Predictions Overlay

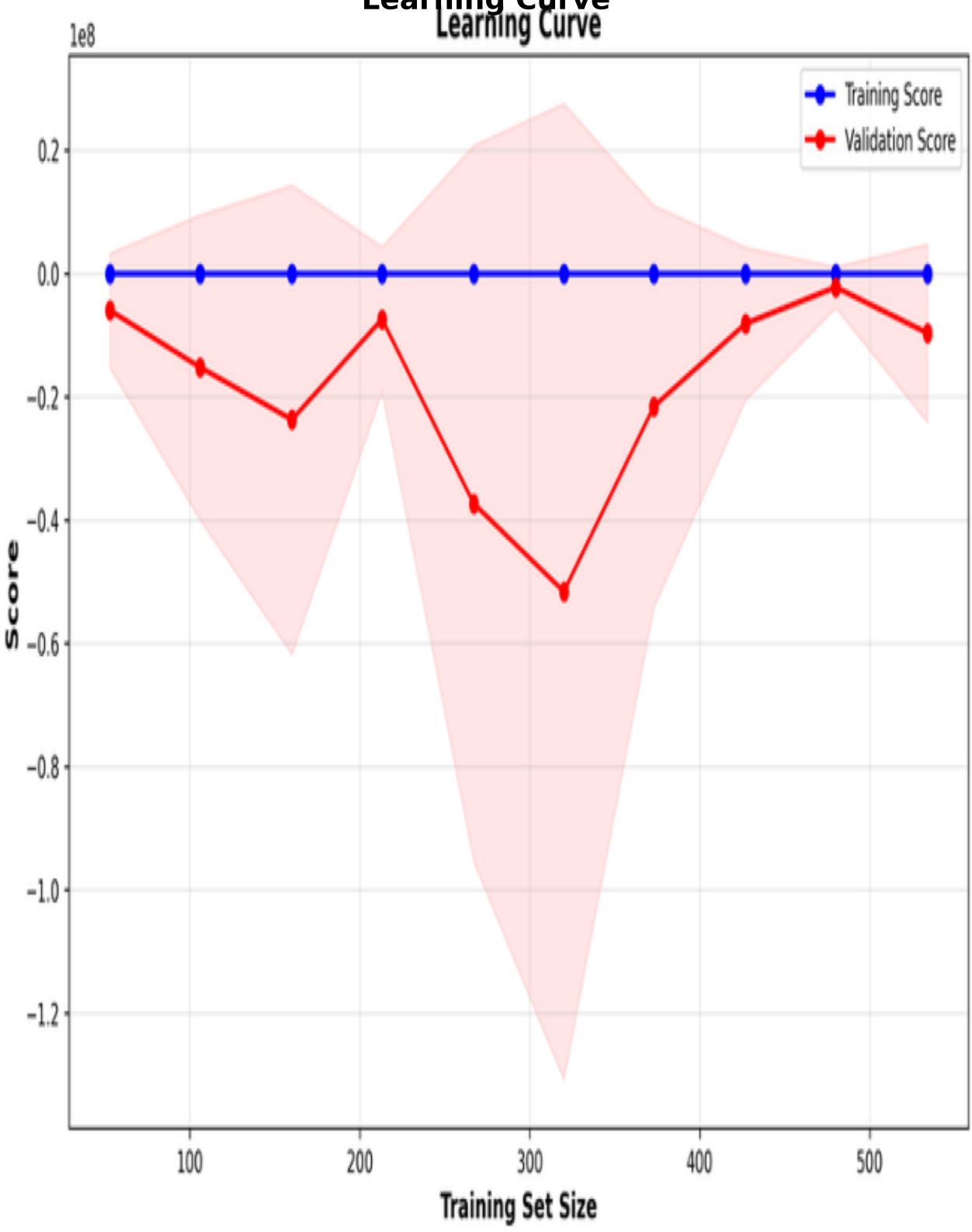


Overfitting Analysis

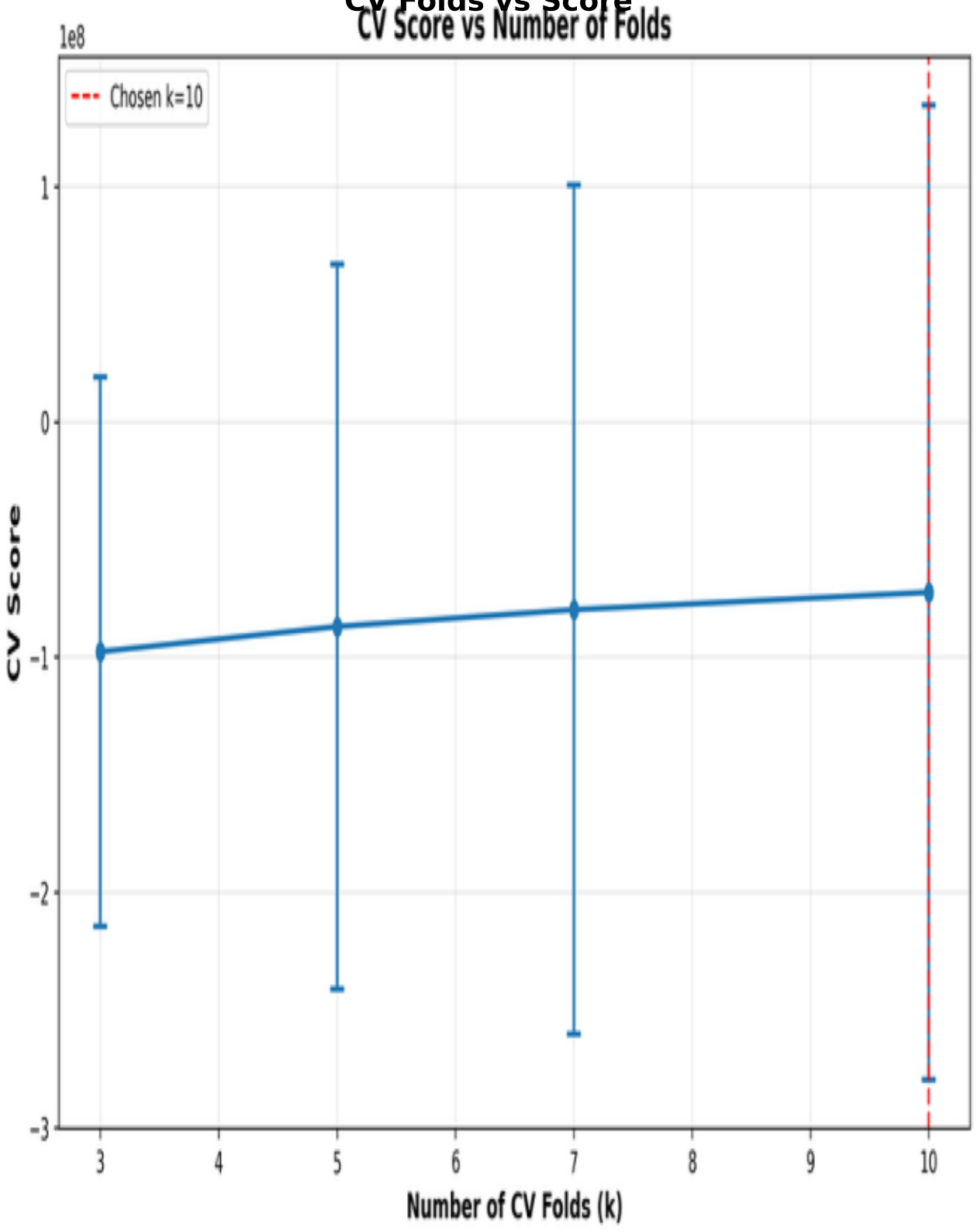
OVERFITTING ANALYSIS Before Tuning: - Train Score: 0.9994 - Validation Score: 0.9982 - Gap: 0.12% -
Overfitting: NO After Tuning: - Train+Val Score: 0.9996 - Test Score: 0.9992 - Gap: 0.04% -

Overfitting: NO ✓ Tuning successfully reduced overfitting

Learning Curve



CV Folds vs Score

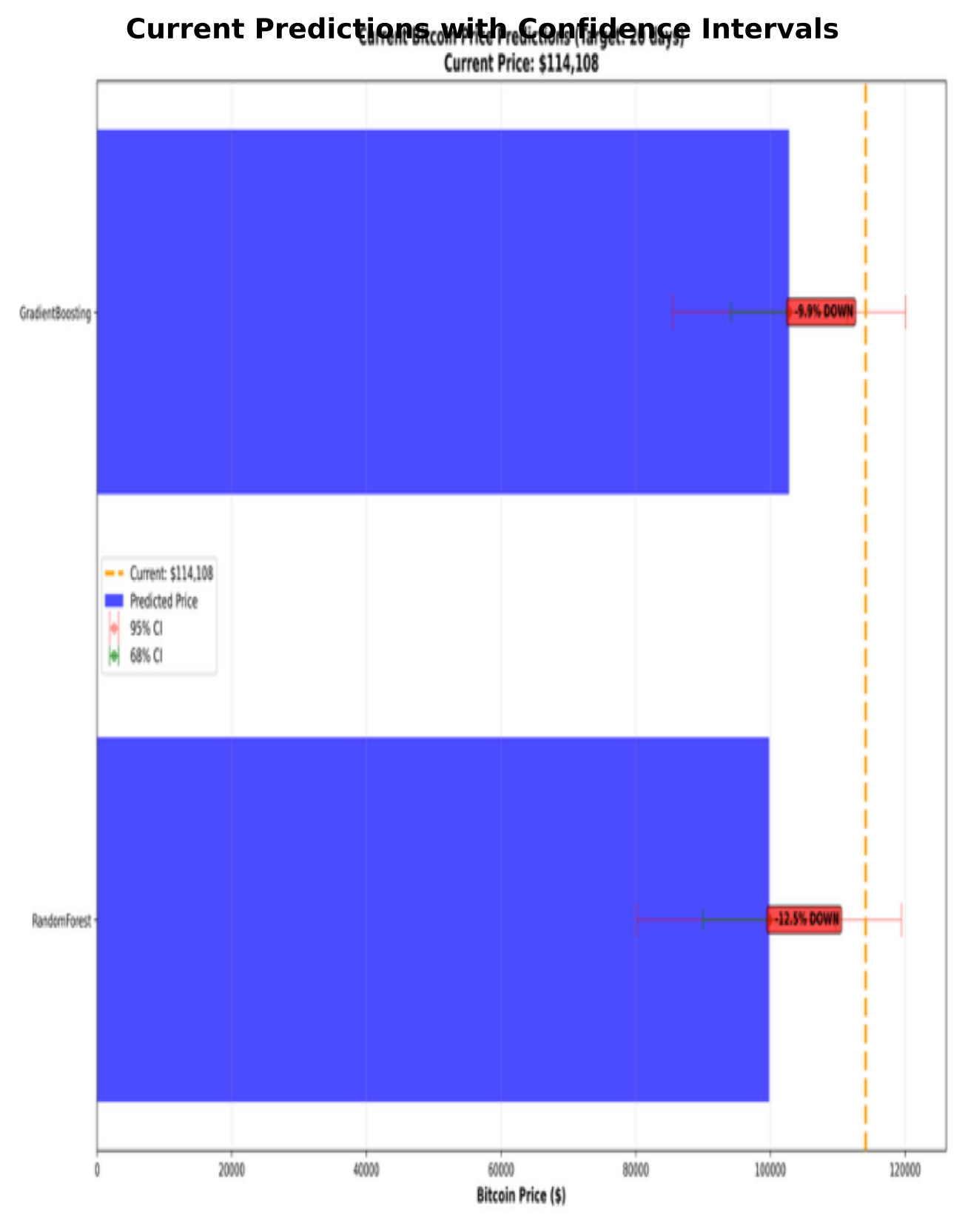


Price Predictions

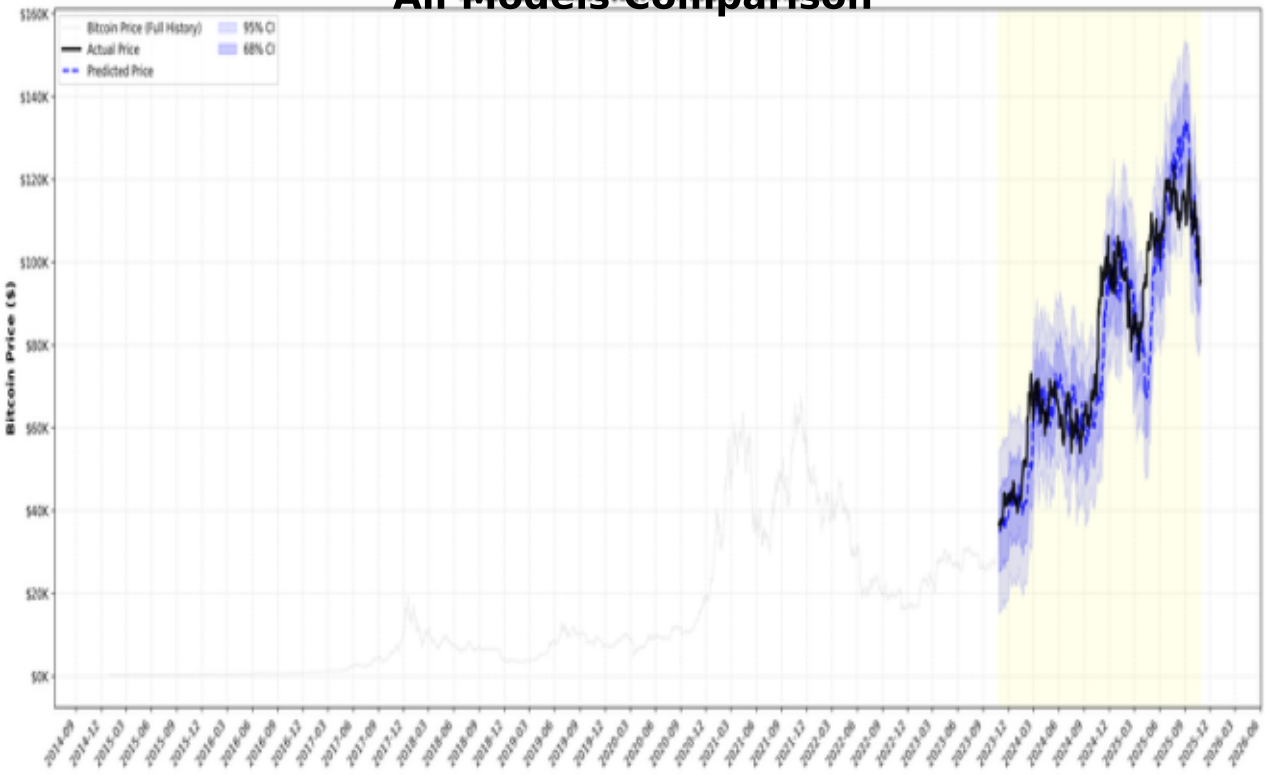
CURRENT PRICE ANALYSIS AND FUTURE PROJECTIONS Current Analysis Date: 2025-11-16 01:00:00 Current
Bitcoin Price: USD 95,454.00 Model: Ridge 20-Day Forward Prediction: - Predicted Price: USD
95,009.29 - Predicted Change: USD -444.71 (-0.47%) - Direction: DOWN
Confidence Intervals: - 68% CI:

USD 94,239.73 - USD 95,778.85 - 95% CI: USD 93,470.17 - USD 96,548.42

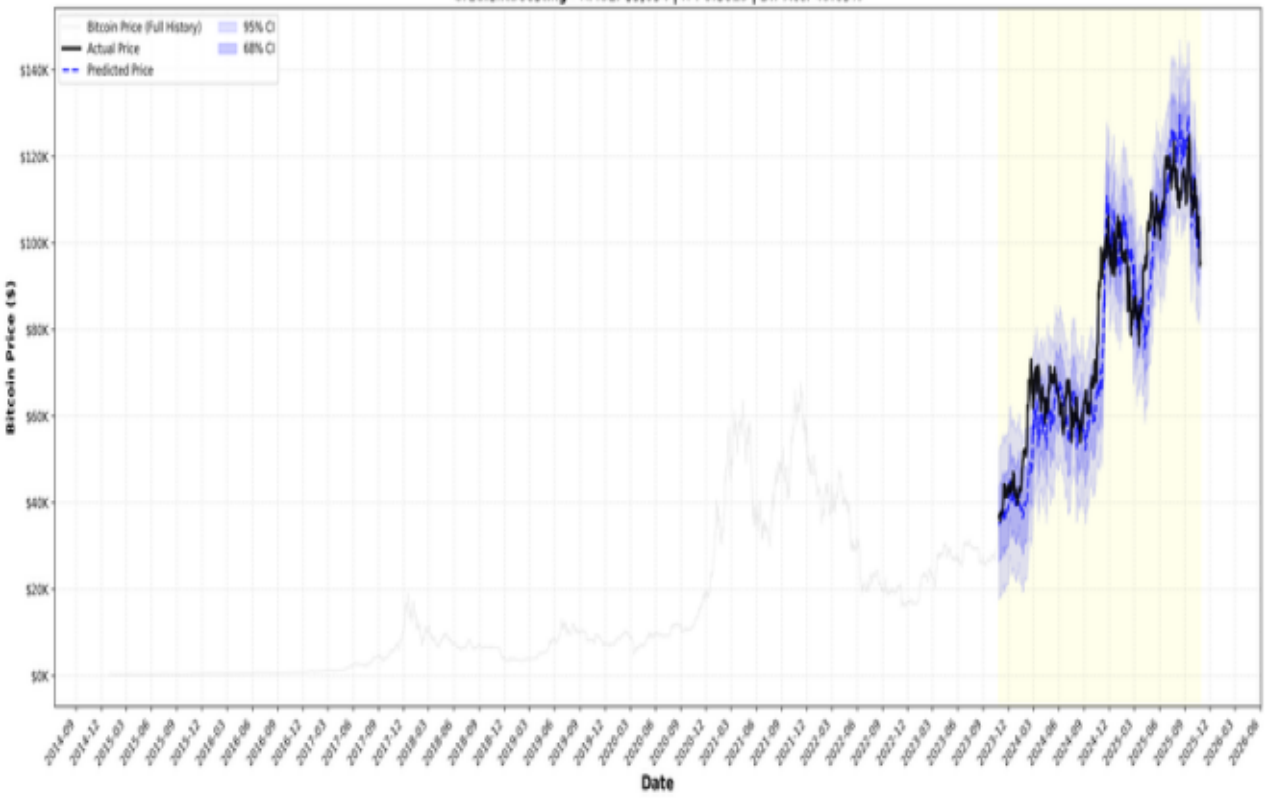
Current Predictions with Confidence Intervals



All Models Comparison



GradientBoosting - RMSE: $9,094 | R^2: 0.8615 | Dir Acc: 45.69%



Conclusions

CONCLUSIONS AND RECOMMENDATIONS Key Findings: 1. FEATURE IMPORTANCE - Halving cycle features are the
most predictive (38-40% importance) - Technical indicators (SMAs, volatility) are highly relevant -
FRED indicators contribute but are less dominant than expected - DGS10 (10-Year Treasury) is the
most important FRED feature 2. MODEL PERFORMANCE - Gradient Boosting achieves best performance (R^2 =
0.86) - Model is better at predicting magnitude than direction - Directional accuracy (~46%) is
close to random - MAPE of ~9% is reasonable for 20-day predictions 3. OVERFITTING - Nested
cross-validation successfully prevents overfitting - Tuning reduces train-test gap - Model
generalizes well to out-of-sample data 4. PREDICTIONS - Model provides useful confidence intervals -
Predictions should be used as one input among many - Risk management is essential given prediction
uncertainty Recommendations: 1. TRADING STRATEGY - Use model predictions for magnitude estimation -
Combine with technical analysis for directional signals - Implement position sizing based on
confidence intervals - Set stop losses based on 95% confidence intervals 2. MODEL IMPROVEMENTS -
Test longer prediction horizons (30, 60, 90 days) - Implement regime detection for different market
conditions - Create ensemble models combining multiple algorithms - Add more interaction features
between FRED and technical indicators 3. RISK MANAGEMENT - Always use confidence intervals for
position sizing - Monitor model performance over time - Retrain models periodically with new data -
Combine ML predictions with fundamental analysis 4. FUTURE WORK - Test alternative models (LSTM,
Transformers) - Implement online learning for adaptive models - Create separate models for different
market regimes - Integrate additional data sources (on-chain metrics, sentiment) Disclaimer: This
analysis is for educational and research purposes only. Past performance does not guarantee future
results. Cryptocurrency trading involves substantial risk of loss. Always conduct your own research