Bitcoin Price Prediction

Alejandra Sanchez **Aqeel Somani** **Jolie Blake** **Jose Hernandez**

Abstract

In today's world, where many aspects of life have become digitized, it's no surprise that currency has also followed this route. Cryptocurrencies started to become prevalent in the late 2000s, particularly with the growth of Bitcoin. But similar to stocks, cryptocurrencies are highly volatile, with price fluctuations occurring frequently. This can pose a problem to buyers of knowing when to invest or even withdrawing from the opportunity altogether. However, our team put several AI models to use in predicting the next seven days' Bitcoin price, over years of data, such as Random Forests, Linear Regression, Huber Regression, and LSTM. Each one was compared up against the other via the following metrics: mean absolute error (MAE), mean-squared error (MSE), root mean-squared error (RMSE), and the $R2$ score. Individually, the MAE, MSE, RMSE measure differences between predicted versus actual prices but on various levels, with the MSE being more sensitive to outliers than the other two. Meanwhile, the R^2 statistic measures how well the model can fit the observed data, so higher is better. We found that well-tuned linear models can match complex methods when supported by strong feature engineering, percentage-change transformations are essential for tree-based models on financial data, outlier removal significantly improves robustness, and LSTM's temporal modeling provides only marginal gains over statistical baselines for medium-term forecasts.

1 Introduction

Our experiment explores how accurately different machine learning models can predict the future price of Bitcoin using historical price action. Bitcoin is one of the most actively traded and heavily volatile cryptocurrencies, so being able to even slightly improve the performance in prediction accuracy can help traders manage risk, and find better entry or exit points. In this experiment our task is to use historical Bitcoin prices and related features to predict future prices over a near-term period. The following sections will cover the data used, the feature engineering process, the modeling approaches tested such as baseline statistical models and more advanced machine learning methods, and the evaluation metrics applied to measure performance on unseen data. The results will highlight how well the best model performed compared to simple benchmarks, what kinds of patterns it was able to capture, and where its predictions still struggle, especially during periods of extreme volatility in the Bitcoin market.

2 Background

Predicting cryptocurrency seems to best be suited with AI models that can handle time-series data decently, or extremely well. Linear models are often used due to their simplistic nature and accuracy

Involvement

Alejandra Sanchez:	*Random Forest, Model Evaluation, Conclusion*	25%
Aqeel Somani:	*Data Collection, Huber Regression, Background*	25%
Jolie Blake:	*LSTM, Outlier Detection, Introduction & Abstract*	25%
Jose Hernandez:	*Data Preprocessing, Linear Regression, EDA*	25%

when it comes to short-term forecasts, such as over the span of a week. Although, they struggle when faced with complex relationships and long-term trends. This is tackled by LSTM models, which often predict cryptocurrency prices relatively well - R^2 scores can reach up to the maximum of 1.00, meaning the variance in the data can be perfectly explained by the independent variable. In addition, both types of models are used in a plethora of research journals, while others have included Support Vector Machines, Random Forests, or more advanced forms of regression to determine prices. The approach taken in this report combines multiple models seen in various studies (regression models, LSTM, and Random Forests), along with several metrics (MAE, MSE, RSME, R^2) to measure how well the models compete when it comes to obtaining a Bitcoin value. Additionally, our approach was to predict prices over a span of seven days, rather than just the next day's price. In this way, it serves as more of a challenge as major price shifts can occur within a small amount of time. Therefore, the models need to be robust enough in learning when (e.g., on a specific month or time) and why prices change rapidly or stay the same.

3 Data

For this experiment, it was necessary to choose quantitative data, cryptocurrency values that have been measured and recorded, on specific dates. The dates ranged all the way from late 2014 up until late 2025, effectively totaling to 9 years worth of Bitcoin market prices. For every day listed, an `open`, `high`, `low`, `close`, `Volume BTC`, and `Volume USD` category, among others, was provided. This dataset is specifically coming from `CryptoDataDownload.com`, where they've obtained it from the Bitstamp exchange. The Bitstamp exchange is one of the initial platforms used in cryptocurrency's infancy and is well-known in the Bitcoin community today. Due to it being around since 2011, large amounts of data can be taken and compiled together to provide a more complete history of the market.

3.1 Preprocessing

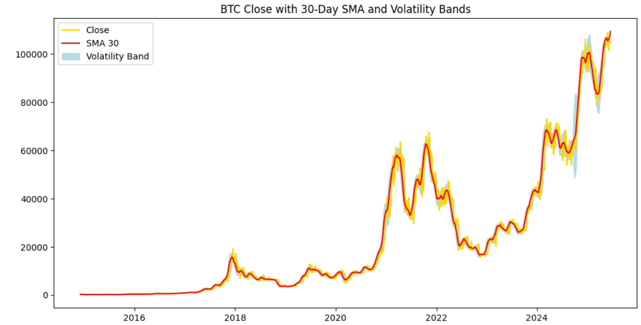
When it came time to work with this data, the first step was preprocessing. This involved removing unnecessary columns from the data (e.g., the `symbol` column, listing the type of cryptocurrency and what country it was from), resolving column inconsistencies, standardizing the data to fit between a certain range to minimize the impact of outliers, splitting the data into a train, validation, test set, and even adding new features. Such features included the exponential moving average (ema), for measuring short-term trends, the simple moving average (sma), for measuring long-term trends, and the relative strength index, which if greater than 70 pointed to overbought Bitcoin and if under 30 pointed to oversold Bitcoin.

4 Exploratory Data Analysis

Major Price Cycles

Bitcoin's price history consistently has followed a four-phase cycle driven by its halving schedule, which reduces block rewards roughly every four years and creates recurring periods of expansion, correction, and consolidation. The four phases are:

Figure 1: Bitcoin Price Over Time



1. The Halving: These programmed supply shocks (occurring in 2016, 2020, and 2024) serve as catalysts for new market cycles by reducing the rate of new Bitcoin entering circulation.

2. The Bull Run: Following each halving, Bitcoin enters an explosive growth phase. The graph shows three distinct bull runs: 2017 (reaching $15,000–16,000), 2021 (peaking near $60,000), and the current 2024–2025 rally (surpassing $100,000). Each bull run reaches progressively higher all-time highs.

3. The Crash: After reaching euphoric peaks, Bitcoin experiences severe corrections. The 2018 crash saw prices fall approximately 80% from peak to trough. The 2022 crash resulted in a roughly 65–70% decline from $60,000 to the $15,000–20,000 range.

4. Crypto Winter: Following each crash, Bitcoin enters an extended consolidation phase lasting 1–2 years. These periods are visible as the relatively flat sections: 2018–2020 ($3,000–12,000 range) and 2022–2023 ($15,000–30,000 range). During Crypto Winter, prices stabilize as the market digests gains, weak hands exit, and accumulation occurs before the next halving.

Figure 1 reinforces that, despite severe corrections, the overall trajectory remains strongly positive, with each cycle reaching new all-time highs.

Figure 2: Bitcoin Monthly Percentage Price Change

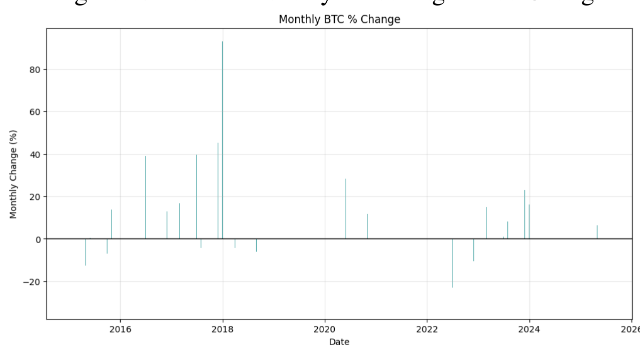
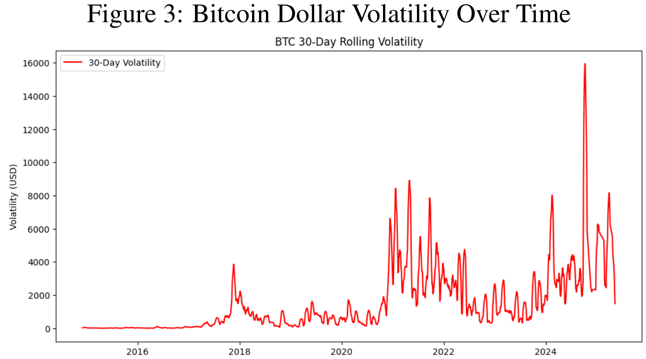


Figure 2 reveals the dramatic month-to-month price swings characteristic of Bitcoin. Several months show changes exceeding 40%, with one notable spike near 90% (likely early 2018). Monthly swings of 20-40% appear regularly throughout the timeline, demonstrating Bitcoin's highly speculative nature. Large positive and negative moves occur throughout all market phases, but the most extreme movements tend to cluster around cycle transitions—particularly during the 2017-2018 period and around 2020-2021. While still volatile, the most recent years (2023-2025) show somewhat smaller percentage swings compared to earlier periods, suggesting potential market maturation as institutional participation increases and market cap grows. The periods between major cycles show reduced volatility, with smaller monthly percentage changes during consolidation phases, supporting the four-phase cycle theory. Also, note that there is some clear assymetry where the largest upward moves tend to be bigger than the largest downward moves, reflecting Bitcoin's overall long-term appreciation despite short-term volatility.

Figure 3: Bitcoin Dollar Volatility Over Time



In Figure 3, we see that the dollar-denominated volatility of Bitcoin has been increasing dramatically over time. The massive percentage change in 2018 [see Figure 2] was enough to create the first major dollar volatility peak, despite the price being lower than today. Moreover, a 20-30% move in 2024, when the price is $\$100,000$, is a dollar move of $\$20,000 - \$30,000$. This is why the recent percentage change bars are smaller than 2018, but the dollar volatility is four times higher (from $\$4,000$ to $\$16,000$).

Political Influence and Institutional Investment

Bitcoin's price movements are not driven solely by market cycles but are also influenced by macroeconomic events and statements from high-profile political figures. U.S. President Donald Trump has made several public comments about cryptocurrency over the years—ranging from skepticism in 2019 to more supportive remarks this year, stating that he would make the United States the "crypto capital of the world." These statements have occasionally aligned with short-term increases in market volatility, reflecting how political attention can affect investor sentiment.

At the same time, institutional participation has played an increasingly important role in Bitcoin's behavior. The expansion of Bitcoin ETFs, corporate treasury adoption, and involvement from major financial institutions have contributed to greater market stability and liquidity. Political commentary, especially when tied to regulatory expectations, often intersects with these institutional trends, amplifying their impact on trading activity.

Overall, the interaction between political signaling and institutional investment highlights the broader external forces that shape Bitcoin's short-term price dynamics beyond purely technical or cyclical factors.

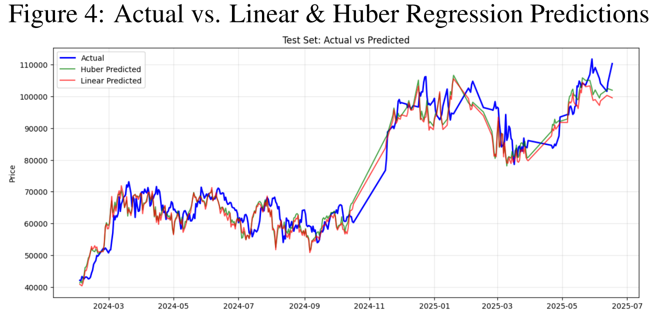
5 Experiments

To evaluate our Bitcoin price prediction approach, we compared four models: Linear Regression, Huber Regression, Random Forest, and LSTM. All experiments used a 70/20/10 chronological train/validation/test split spanning 2014-2025. The test set presented a challenging extrapolation task, with average prices of $72,700 compared to the training set's $11,133. We engineered 21 features including lagged prices, technical indicators (SMA, EMA, RSI, MACD), momentum metrics, volatility measures, and volume data. Isolation Forest outlier detection removed 5% of extreme data points, improving all models by 3-5% in R^2 score.

5.1 Baseline Linear Models

Linear Regression and Huber Regression (Figure 4), established strong baselines with standardized features. Huber used $\epsilon = 1.35$ for outlier robustness and $\alpha = 0.0001$ for L2 regularization. Both achieved validation $R^2 = 0.92$ with MAE around $1,640. On the test set, Huber slightly outperformed Linear with $R^2 = 0.91$ (MAE = $\$4,028$) versus $R^2 = 0.90$ (MAE = $\$4,169$), representing 5.5-5.7% prediction error. The strong test performance demonstrated successful extrapolation to prices six times higher than training data, with Huber showing superior robustness to volatility.

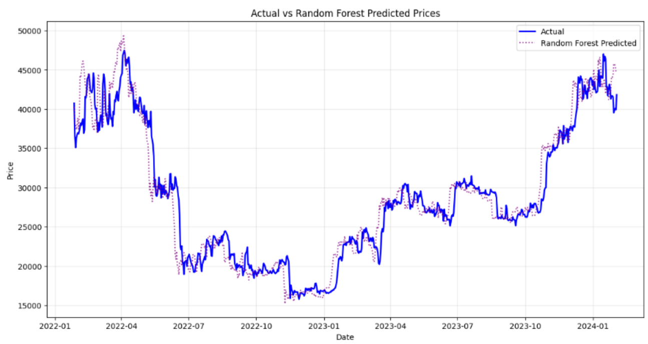
Figure 4: Actual vs. Linear & Huber Regression Predictions



5.2 Random Forest With Percentage Change

Random Forest (Figure 5) initially failed catastrophically when predicting absolute prices (test $R^2 = -18.92$) because tree-based models cannot extrapolate beyond their training range. We implemented a percentage change approach where the model predicts relative change from current price instead of absolute price. Using `GridSearchCV` with 3-fold cross-validation, we found optimal hyperparameters: 200 trees, `max_depth=30`, `min_samples_split=2`, `min_samples_leaf=2`, and `max_features='sqrt'`. With this transformation, Random Forest achieved validation $R^2 = 0.92$ (MAE $= \$1,632$) and test $R^2 = 0.86$ (MAE $= \$4,961$, with 6.8% error). Feature importance analysis revealed Volume BTC (8.2%) and `volume_lag_1` (6.5%) as most predictive, indicating trading volume's importance for ensemble methods.

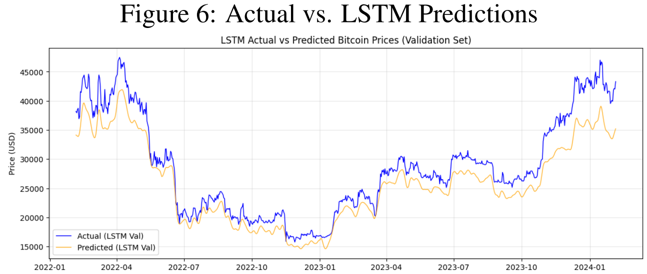
Figure 5: Actual vs. Random Forest Predictions



5.3 LSTM Neural Network

We implemented an LSTM (Figure 6) with two layers (128 and 64 units) followed by dropout (0.3) and dense layers, using a 60-day lookback window with 27 features. We experimented with different configurations: a 30-day lookback with LSTM(64, 32) achieved validation $R^2 = 0.81$, while extending to 60 days with LSTM(128, 64) improved to $R^2 = 0.84$. Adding enhanced features (ATR, Bollinger Bands) further improved to $R^2 = 0.85$ (MAE $= \$2,756$). The model trained using *Adam* optimizer, a batch size of 32, and early stopping. On the test set, LSTM achieved $R^2 = 0.88$ with MAE $= \$3,892$ (5.4% error), showing the lowest absolute dollar error despite slightly lower R^2 than linear models.

Figure 6: Actual vs. LSTM Predictions



Final Test Rankings

All models achieved 5.4-6.8% prediction error [see Table 1], demonstrating strong performance for 7-day cryptocurrency forecasting. Residual analysis showed approximately normal error distributions with slight heteroscedasticity at higher prices. Models successfully tracked both upward trends and consolidation periods throughout 2024-2025 but struggled during extreme volatility spikes exceeding 20% daily movement.

Key takeaways from our experiments:

Table 1: Model Performance Comparison

Model	R^2	MAE
Huber	0.91	$4,028
Linear	0.90	$4,169
LSTM	0.88	$3,892
Random Forest	0.86	$4,961

1. Well-tuned linear models can match complex methods with proper feature engineering.
2. Percentage change transformation is essential for tree-based models on financial data, outlier removal significantly improves robustness.
3. LSTM's temporal modeling provides only marginal gains over statistical baselines for medium-term forecasts.

6 Conclusion

In this experiment, we successfully predicted Bitcoin prices seven days ahead using four machine learning models: Huber Regression, Linear Regression, Random Forest, and LSTM. All models achieved strong performance with R^2 scores above 0.86 on unseen test data from 2024-2025. Huber Regression performed best with a test R^2 of 0.91 and MAE of $4,028 (~5.5% error), demonstrating that robust statistical methods can match more complex approaches when handling volatile financial data. Random Forest required a percentage change approach to extrapolate beyond training price ranges, revealing important limitations of tree-based models compared to linear methods.While 5-7% prediction error represents solid performance for cryptocurrency forecasting, significant challenges remain. Bitcoin's extreme volatility during market crashes, regulatory changes, and unexpected events continues to challenge even well-tuned models. Our models capture trends and momentum but cannot predict black swan events. For traders and investors, these models provide useful tools for identifying medium-term trends, but should be combined with risk management and market awareness—not used as standalone predictors. Future work could incorporate sentiment analysis, test ensemble methods, and evaluate performance on other cryptocurrencies like Ethereum. This experiment demonstrates that machine learning offers meaningful insights for Bitcoin price prediction. While perfect forecasting remains elusive, our models provide practical accuracy that could assist informed investment decisions when paired with sound financial judgment.

References

[1] Anton Badev & Matthew Chen (2014) Bitcoin: Technical Background and Data Analysis. *Finance and Economics Discussion Series*, 2014-104. Washington: Board of Governors of the Federal Reserve System.