

Final Project Report

Bitcoin Volatility Prediction Using Machine Learning

1. Project Overview

Cryptocurrency markets are known for extreme price fluctuations and uncertainty. Volatility prediction plays a crucial role in risk management, trading strategies, and financial decision-making.

This project aims to predict next-day Bitcoin volatility using historical market data and machine learning techniques. The system processes historical price and volume data, computes volatility using statistical methods, engineers relevant features, and applies a regression-based machine learning model to generate predictions.

2. Dataset Description

- The dataset consists of historical cryptocurrency market data.
- It includes OHLC prices, trading volume, market capitalization, and cryptocurrency identifiers.
- For this project, the scope is limited to Bitcoin to ensure clarity and reliable analysis.
- The data is treated as a time-series dataset, maintaining strict chronological order.

3. Methodology

3.1 Volatility Calculation

- Daily log returns are calculated from closing prices.
- Volatility is computed as a 7-day rolling standard deviation of log returns.
- The prediction target is defined as next-day volatility.

3.2 Feature Engineering

- To improve prediction accuracy, the following features are used:
- Lagged volatility values (1, 3, 7, 14 days)
- Absolute returns
- High–Low price range
- Liquidity ratio (volume to market capitalization)

These features capture temporal dependency, price movement, and market liquidity.

3.3 Model Selection

- Random Forest Regressor is chosen due to its ability to model non-linear relationships and robustness to noise.
- A time-based train-test split (80% training, 20% testing) is applied to prevent data leakage.

4. Model Evaluation

The model performance is evaluated using standard regression metrics:

- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R^2 Score

Key observations:

- Test RMSE is comparable to or slightly lower than Train RMSE.
- This indicates good generalization and absence of overfitting.
- The model significantly outperforms a naïve baseline that uses previous-day volatility.

5. Results and Analysis

- The model successfully captures volatility clustering behavior.
- Feature importance analysis shows that recent volatility (lag-1) is the most influential predictor.
- Actual vs predicted volatility plots demonstrate that the model closely follows real volatility trends.
- Hyperparameter tuning results in marginal performance improvement, confirming model stability.

6. Key Insights

- Volatility exhibits strong temporal dependency.
- Lag-based features are critical for accurate volatility prediction.
- Random Forest models are effective for short-term financial time-series forecasting.
- Proper time-based data splitting is essential to avoid misleading results.

7. Limitations

- The model is trained only on historical price-based features.
- External factors such as news events, macroeconomic indicators, and market sentiment are not included.
- Predictions are limited to short-term (next-day) volatility.

8. Future Scope

- Extend the model to multiple cryptocurrencies.

- Incorporate sentiment analysis and external economic indicators.
- Experiment with deep learning models such as LSTM for long-term forecasting.
- Deploy the model as a real-time volatility monitoring system.

9. Conclusion

This project demonstrates a complete and well-structured machine learning pipeline for Bitcoin volatility prediction. By combining financial theory with machine learning techniques, the system achieves reliable and interpretable predictions. The results validate the effectiveness of feature engineering and ensemble learning for modeling cryptocurrency market volatility.