CS 7641 ML Project Final Report

Team Member Contribution

Contributions	Member
How-Shing Wang	dataset description, Autoformer, expected results
Hsun-Heng Lo	motivation, D-linear
Hung-Yu Shih	evaluation metric, Transformer
Yan-Tong Lin	problem description, literature review, Mamba
Yao-Ting Huang	data preprocessing, LSTM

Introduction/Background

Literature Review

Recent success in transformers has shown their effectiveness in various sequence modeling tasks[1]. Their potential application to high-noise time series prediction, such as crypto markets, is promising. Various specializations of transformers show competitive results on time series data, including Autoformer (Wu et al., 2021[2]). However, some studies suggest linear methods may still outperform in some time series prediction tasks[3]. In this work, we evaluate multiple architectures in such challenging conditions.

Bhandari[4] conducted a comparative analysis of single-layer versus multilayer Long Short-Term Memory (LSTM) neural network architectures to forecast the next-day closing price of the S&P 500 index. The study incorporated nine predictors derived from fundamental market data, macroeconomic indicators, and technical analysis metrics, including the Moving Average Convergence Divergence (MACD). The findings revealed that the single-layer LSTM model achieved superior prediction accuracy and a better fit to actual market behavior compared to its multilayer counterparts.

Chen et al.[5] employed LSTM networks to predict stock returns within the Chinese market by transforming historical stock data into 30-day sequences, utilizing ten learning features that encompassed both the stock's and the market index's OHLCV (Open, High, Low, Close, Volume) data. Their LSTM-based approach enhanced prediction accuracy from a baseline of 14.3% (random prediction) to 27.2%, demonstrating the model's effectiveness in capturing the complexities of the Chinese stock market.

Nelson et al.[6] utilized LSTM neural networks to forecast stock price directions (upward or downward) by leveraging 175 features, which included historical price data and a comprehensive set of technical analysis indicators. The model was trained on 15-minute interval data from Brazilian stocks. The LSTM-based investment strategy yielded positive financial returns with lower maximum drawdowns, indicating enhanced profitability and reduced risk compared to traditional investment approaches.

Dataset Description

Our study focuses on the top 15 cryptocurrencies, selected based on their market capitalization and trading volume, over the period from January 1, 2023, to November 1, 2024. We utilized K-line data retrieved from the Binance API (https://developers.binance.com/docs/derivatives/coin-margined-futures/market-data/Kline-Candlestick-Data), encompassing high-frequency trading intervals of 1 minute, 5 minutes, and 15 minutes. The dataset includes a comprehensive set of indicators such as Open, High, Low, Close and Volume.

Problem Definition

Problem

This project aims to compare neural architectures for their predictive ability in high-noise, low-data environments like crypto markets.

Motivation

The field of cryptocurrencies presents a distinctive environment where individual investors may possess inherent advantages over traditional stock markets. Unlike stock markets, which have been extensively analyzed and monitored by sophisticated trading firms employing advanced strategies, the cryptocurrency market remains relatively nascent and less saturated with institutional participants. This disparity implies that individual investors operating within the cryptocurrency domain often have access to similar informational resources as their institutional counterparts, given the transparent and publicly accessible nature of crypto trading data. Consequently, the competitive edge typically held by large trading companies in stock markets due to their proprietary information is less pronounced in the cryptocurrency arena.

Moreover, the cryptocurrency market is characterized by its pronounced volatility, which poses significant challenges for accurate price prediction and risk management. Traditional financial markets, despite their volatility, benefit from established analytical frameworks and a wealth of historical data that inform trading strategies. In contrast, the rapid and often unpredictable fluctuations in cryptocurrency prices necessitate more agile and adaptive analytical approaches. This is where machine learning (ML) techniques emerge as a transformative tool, offering the capability to process and analyze vast amounts of data swiftly and efficiently.

Machine learning algorithms allow individual investors to perform comparative analyses and real-time assessments of market trends, enabling more informed and timely decision-making. Techniques such as neural networks, ensemble methods, and transformers can be employed to enhance predictive accuracy, optimize trading strategies, and manage risks more effectively. Furthermore, the application of ML in cryptocurrency trading can uncover latent market inefficiencies and exploit them before they are arbitraged away, thereby offering individual investors a potential strategic advantage.

In essence, the integration of machine learning into cryptocurrency trading not only addresses the challenges posed by market volatility but also democratizes access to sophisticated analytical tools. This empowers individual investors to compete more effectively in a space that is still evolving and less dominated by institutional heavyweights.

Methods

Data Preprocessing Methods

- Normalization : Ensures all features are on a similar scale and reduce bias towards high-magnitude features.
- Log-Transform : Crypto data often exhibit significant skewness and volatility. To address these characteristics, a logarithmic transformation was applied to the crypto returns, calculated as: $\log y_t - \log y_{t-1}$ This transformation stabilizes the variance and normalizes the distribution of returns.
- PCA (Principal Component Analysis): Reduces dimensionality, addresses the issue of multicollinearity and highlights the most impactful features.
- Seasonal Trend Decomposition: Separates trend, seasonality, and noise. Digital currencies may exhibit seasonal trends due to factors like market cycles, investor sentiment or holidays.

ML Algorithms/Models Identified

- D-Linear : Linear regression uses the linear trends of the observed data to predict future price movements. Often used as a baseline model in cryptocurrency analysis. Linear regression provides valuable insights and serve as a starting point for exploratory data analysis of the market.
- LSTM : LSTM (Long Short-Term Memory) is a type of recurrent neural networks (RNNs) designed to handle sequential data, making them well-suited for time series prediction. LSTM uses three gates (input, forget, and output) to control the flow of information, enabling it to remember important patterns over long time sequences while forgetting irrelevant data.
- Transformer : Tranformer is a powerful deep learning model designed for handling sequencial data. Utilizing the self-attention mechanism, the model is able to capture intra-token relations and excel in tasks like Natural Language Processing and Computer Vision.
- Autoformer : Researches suggest that Transformers's performance suffers in long-term time series forecasting. To overcome this, autoformer proposes series decomposition and auto-correlation, achieving $O(L \log L)$ time complexity and better performance.
- Mamba : A neural architecture derived from linear RNNs and State-Space Model discretization. It claims to rival transformers in training efficiency and modeling capabilities for language tasks, with its state-space foundation making it especially suited for time series data.

Unsupervised Learning

- We use PCA in some of the feature processing as our unsupervised learning method.

Results and Discussion

Quantitative Metrics

- MSE: The Mean Square Error between the ground truth and the predicted time series.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - f_i)^2$$

- R2: The R2 score is defined by 1 minus the residual sum of square divided by the total sum of square. Which is

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$$

This can be interpreted as how well the variation of the data is explained by the model.

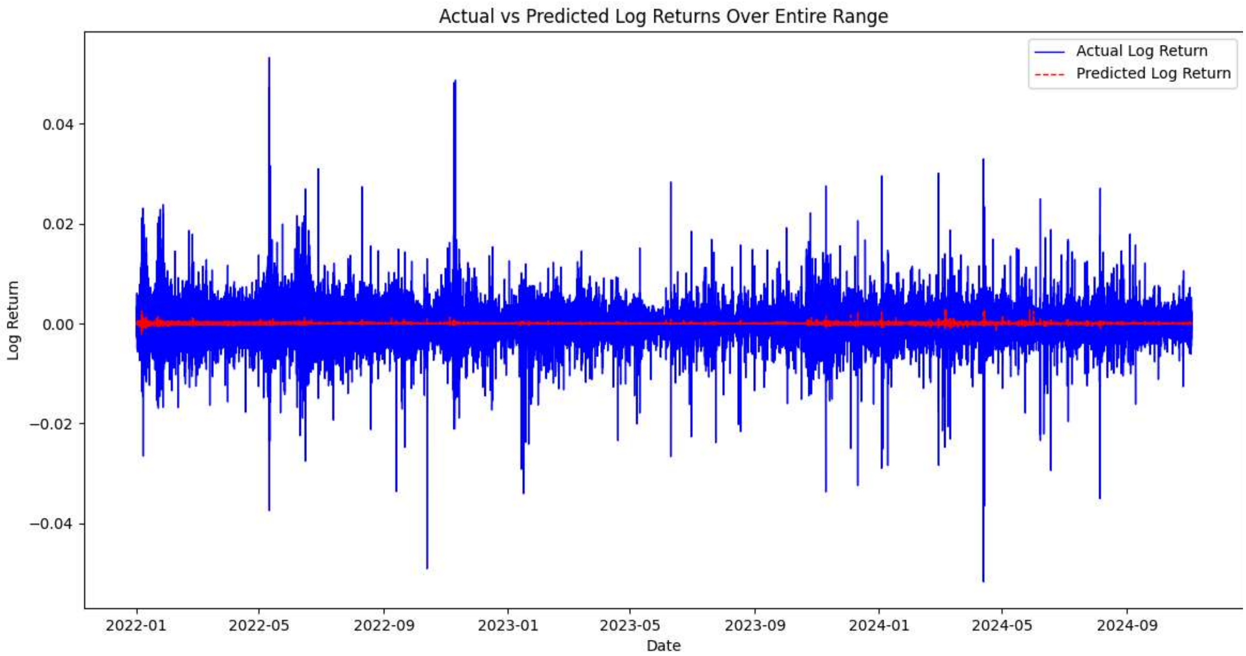
- Sharp Ratio: Sharpe ratio is defined by

$$S_p = \frac{R_p - R_f}{\sigma_p}$$

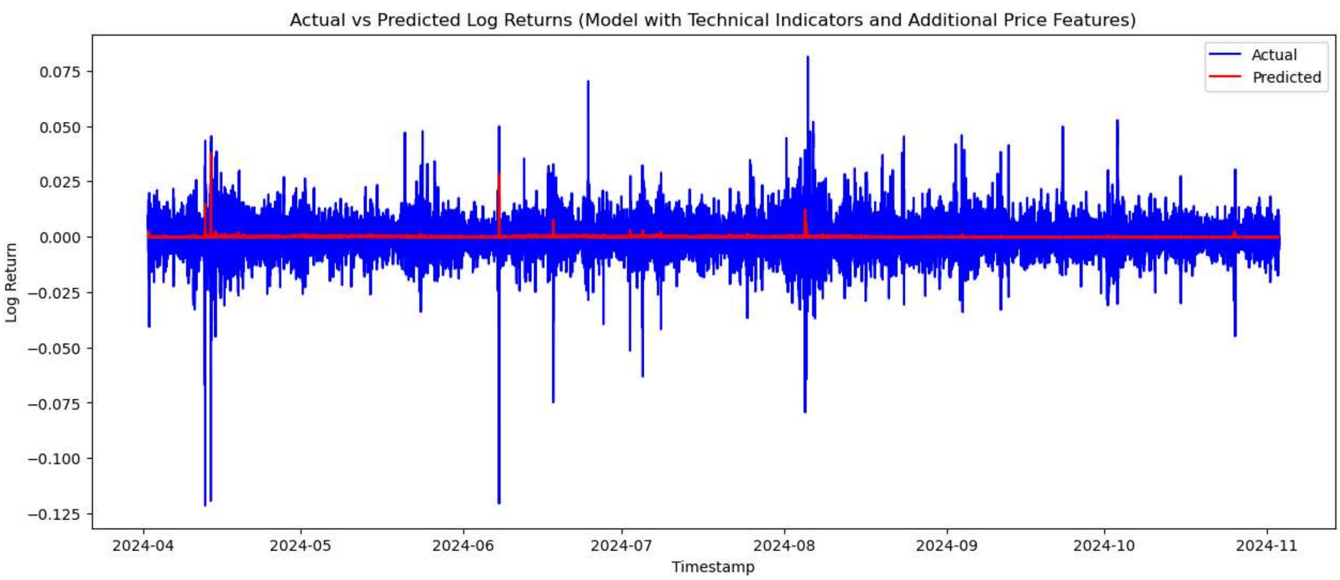
It is used to evaluate how "good" an investment is by comparing its expected return and risk.

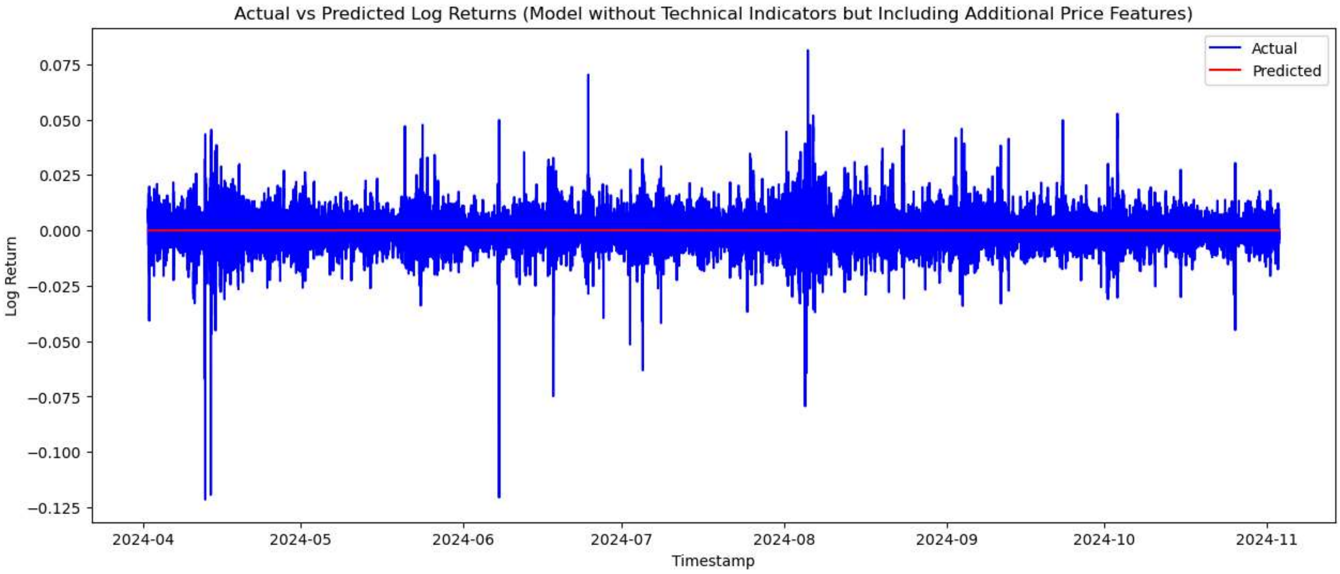
Final Result

- Linear Regression
  - Visualization
    The blue line is the ground truth, and the red line is the model prediction result.



  - Evaluation:
    - MSE: 1.68×10^{-6}
    - R2 score: -0.0016
  - Analysis:
    The linear regression model performed poorly on predicting the log return of the original time series data. Despite achieving a low Mean Squared Error (MSE) the r squared score was negative (-0.0016), indicating that the model failed to explain the variance in the data and performed worse than a simple baseline mean predictor. This poor performance can be attributed to several factors:
    1. Linear regression is a simple model that assumes a linear relationship between input features and the target variable. Time series data, especially financial data like log returns, often have complex and nonlinear patterns that linear regression cannot capture effectively.
    2. The data used consisted of minute-by-minute cryptocurrency prices, which are highly volatile and noisy. This randomness introduces significant variability that a linear regression model struggles to handle, leading to underperformance.
  - Future work:
    In future works for linear regression. We can look to preprocess the data with more engineered features aside from adding lag. Adding features based on the market sentiment could prove to be useful, as the sentiment features can be closely correlated to the price. Incorporating different domain knowledge other than simply the high, low, close, and volume into feature design can significantly enhance the model's effectiveness.
- LSTM
  - Visualization



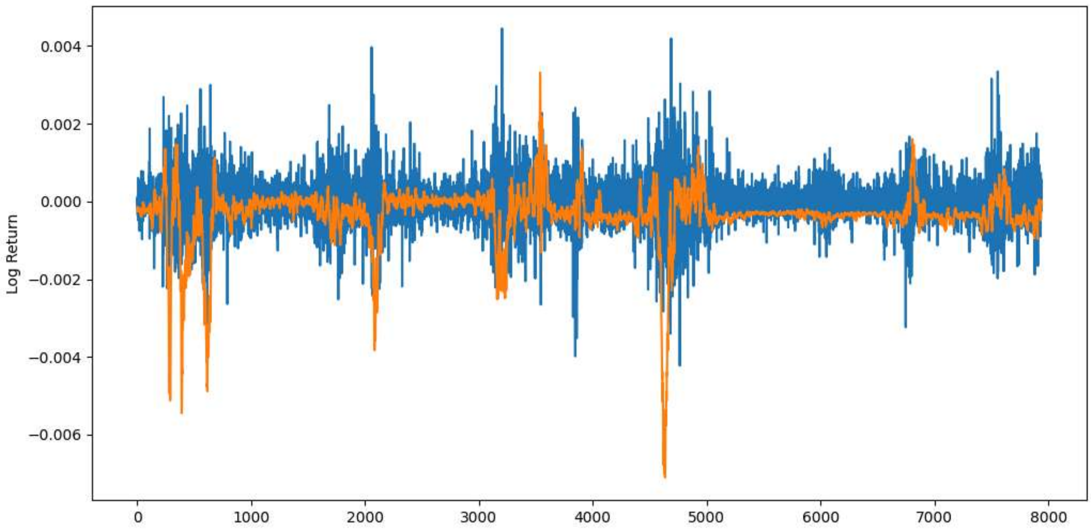


- Evaluation:
  - Without technical indicators
    - MSE: 0.000034 (RMSE: 0.0058)
    - R2 score: 9.17 $\times 10^{-5}$
  - With technical indicators
    - MSE: 0.000034 (RMSE: 0.0058)
    - R2 score: 1.8 $\times 10^{-4}$
- Analysis: The models were trained using different feature sets to evaluate the impact of incorporating technical indicators (SMA_10, SMA_50, RSI, MACD, MACD_signal, MACD_diff). The visulization of the model without technical indicators suggests the predicted log returns are almost identical to zero. This indicates that the model is not capturing any significant patterns or trends in the data. On the other hand, the inclusion of technical indicators results in a slight improvement in the model's predictions. While the predictions still barely explains the variance, there is a marginal increase in variability captured by the model.

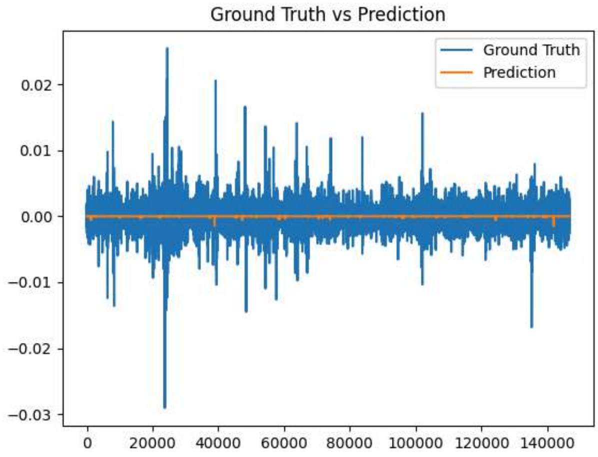
  The minimal changes in MSE and RMSE across models with and without technical indicators suggest that the overall error magnitude remains low, but this is misleading due to the model's lack of variability in predictions.
- Future work: Building upon the observed enhancements in model performance through the incorporation of additional features, future research endeavors should focus on further augmenting the predictive capabilities of LSTM models. This can be achieved by integrating external data sources, including macroeconomic indicators and sentiment analyses derived from social media platforms and news outlets. Such data can provide a more holistic view of the factors influencing cryptocurrency price movements.

  Moreover, the inherent high noise levels associated with one-minute interval predictions present significant challenges to accurate forecasting. To address this, it is recommended to employ advanced scaling methodologies that can better normalize the data and reduce variability. Alternatively, aggregating log returns over a 15-minute period may help in diminishing bias and enhancing the stability of predictions. These approaches are expected to contribute to more reliable and robust model performance in high-frequency trading environments.

- Transformer
  - Visualization
    The blue line is the ground truth, and the orange line is the model prediction result.

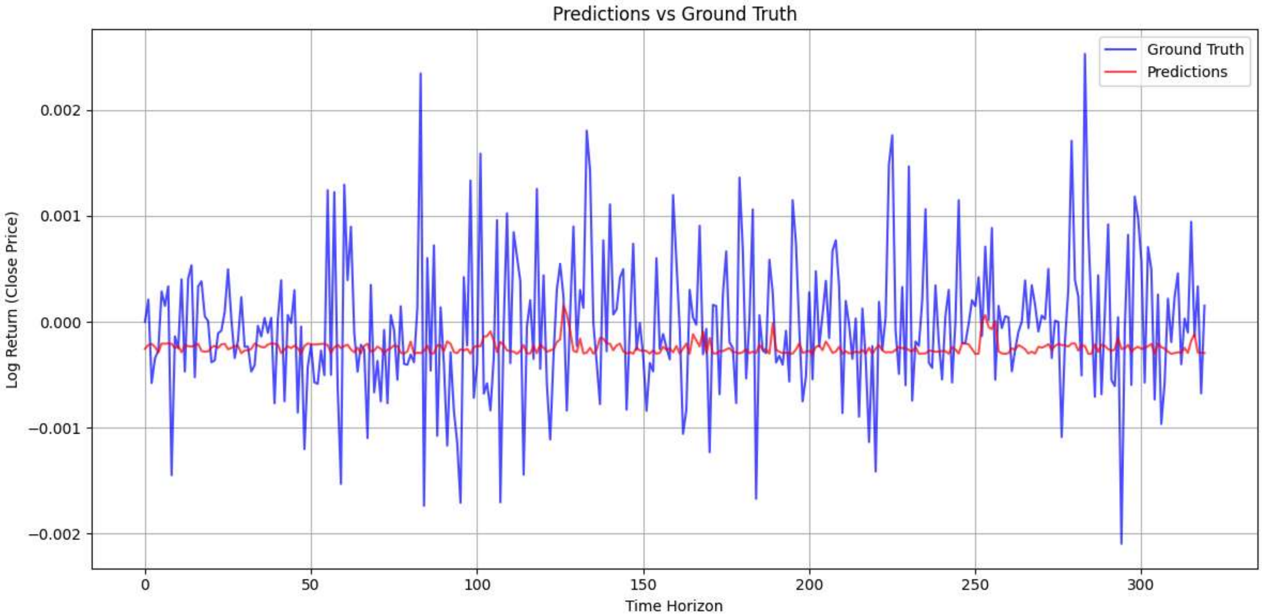


  - Evaluation:
    - MSE: 9.91 $\times 10^{-7}$
    - R2 score: -1.859641
  - Analysis:
    The transformer model performed poorly on predicting the log return of the original time series data. Despite the low testing loss, the R2 score was very low, indicating that the model failed to learn the data's patterns effectively. After inspecting the visualization result, we concluded that this is due to underfitting. There are a few key reasons that cause the model to underfit:
    1. The self-attention mechanism of transformer is permutation-invariant. Although the positional embedding can preserve some of the information of the order, the information loss might still be too much. Especially for the task of time series prediction, in which the order of the data is crucial for the analysis.
    2. Predicting log-return of financial data is inherently challenging, as changes in the data can often be highly random and unpredictable.
    3. The data we used is minute-by-minute cryptocurrency prices, which is often noisy and volatile, making it hard for the model to predict.
  - Future work:
    We originally thought the underfitting is due to the model being too small and lacking flexibility. After increasing the model size by more than 40 times, the performance had a noticeable improvement, but it's still far from enough, and training an even larger model will be computationally difficult. If we want to achieve a better result using the transformer model, we may need to explore more advanced feature engineering techniques to preprocess the data, making it easier for the model to interpret and analyze.
- Autoformer



- Visualization:
- Evaluation:
  - MSE: 1.07222e-06
  - R2 score: -9.547e-05
  - Directional Accuracy: 0.4224
- Analysis:
  Despite incorporating time features into the model, autoformer still failed to capture the signal in the data.
  1. The low directional accuracy and r2 implies that the model is similar to a random model
  2. Data is too noisy, more robust features instead of raw features need to be developed to boost the performance.
- Future work:
  Despite using hyperparameter tuning framework like ray to tune the learning rate, we might need to also tune on model architecture, such as the number of layers or feed-forward dimension. Or we can try increasing the lookback window of the data.

- Mamba
  - Visualization



  - Evaluation:
    - Training Set
      - MSE: 7.13e-7
      - RMSE: 0.00844
      - R2 score: -0.2141
    - Validation Set
      - MSE: 6.29e-7
      - RMSE: 0.00793
      - R2 score: -0.1037
  - Discussion:
    - Performance
      - The validation metrics are similar to the training metrics, suggesting that the model doesn’t overfit the training data.
      - Both performance is suboptimal for both training and validation given the negative R2.
    - Reasons
      - The training and validation metrics are close, indicating minimal overfitting.
      - Given that the model size is considerable large compared to the input feature, it may because that the features used are insufficient or uninformative for the task
    - Future Works
      - Incorperating more data sources, e.g. onchain transaction volume, large transfers to exchange
      - Trying different time horizon (e.g. lower frequency or predict into later in the future)
      - Feature engineering (which we try to avoid by using advanced sequence, but may be actually necessary)

Comparison

The following is the comparison of the model performance.

Model	MSE	R2
Linear	1.68×10^{-6}	-0.0016
LSTM	3.4×10^{-4}	9.17×10^{-5}
Transformer	9.91×10^{-7}	-1.8596
Autoformer	1.07×10^{-6}	-9.54×10^{-5}
Mamba	6.29×10^{-7}	-0.1037

- Despite experimenting with various model sizes and architectures, the overall performance of all tested approaches remained suboptimal, indicating the inherent challenges in applying machine learning to high-frequency cryptocurrency time series data.
- The information contained in the input data, largely based on raw price and volume indicators, might be inadequate for the prediction task, suggesting a need for more robust feature engineering and integration of external data sources.
- The low signal-to-noise ratio and the high volatility characteristic of cryptocurrency markets make accurate predictions particularly difficult, as the models struggle to distinguish meaningful patterns from random fluctuations.

Future Work

- Explore more advanced feature engineering method to make the data more interpretable to the models.
- Incorporating external data sources, such as such as macroeconomic indicators, social media sentiment analysis, and blockchain on-chain activity.
- Experiment on data with different frequency (5 min, 15 min, 1 hour, or predict aggregated value over a certain time span)

References

1. Vaswani, A., et al. (2017). "Attention is all you need." ↩

2. Wu, H., et al. (2021). "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting." ↩

3. Zeng, Ailing, et al. (2023). "Are Transformers Effective for Time Series Forecasting? A comprehensive study." ↩

4. Bhandari, Hum Nath, et al. "Predicting stock market index using LSTM." Machine Learning with Applications 9 (2022): 100320. ↩

5. Chen, Kai, Yi Zhou, and Fangyan Dai. "A LSTM-based method for stock returns prediction: A case study of China stock market." 2015 IEEE international conference on big data (big data). IEEE, 2015. ↩

6. Nelson, David MQ, Adriano CM Pereira, and Renato A. De Oliveira. "Stock market's price movement prediction with LSTM neural networks." 2017 International joint conference on neural networks (IJCNN). Ieee, 2017. ↩