SemeAd 2024
XXVII Seminários em Administração

05, 06, 07 e 08 de novembro de 2024
ISSN 2177-3866

Application of Stacking Models for Crypto Price Prediction and Portfolio Optimization Based on Risk Metrics

JOSÉ ERASMO SILVA
UNIVERSIDADE FEDERAL DA BAHIA (UFBA)

SALATIEL SANTANA DE SOUZA
UNIVERSIDADE ESTADUAL DE SANTA CRUZ (UESC)

Application of Stacking Models for Crypto Price Prediction and Portfolio Optimization Based on Risk Metrics

Introduction

Cryptocurrencies have emerged as a highly volatile and fast-growing asset class, attracting investors, speculators, and researchers (Sebastião & Godinho, 2021). The volatility and speculative nature of cryptocurrencies, such as Bitcoin, Ethereum, and others, present significant challenges, but they also offer unique opportunities for the application of advanced price prediction models and trading strategies (Ahmed, 2024; Otabek & Choi, 2024).

The existing literature on cryptocurrency price prediction often focuses on the effectiveness of different predictive models, utilizing error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) (AlMadany, Hujran, Naymat, & Maghyereh, 2024; Muchtadi-Alamsyah et al., 2024). In addition, classification metrics such as accuracy, sensitivity, specificity, and the area under the ROC curve (AUC-ROC) are widely used to evaluate the performance of price prediction models (Chandra, Tyagi, Gupta, Saxena, & Kharaliya, 2024; Singla, Gill, Chauhan, Pokhariya, & Lande, 2024). These metrics help to understand the ability of models to correctly differentiate between positive and negative price movements, offering a more practical insight into trading strategies (Singla et al., 2024).

However, one crucial aspect often overlooked is how these forecasts translate into tangible returns for investors. The accuracy of forecasts is essential, but their practical application in trading strategies and their impact on portfolio returns are equally important (Guo, Sang, Tu, & Wang, 2024; Li & Ma, 2024). Furthermore, while many studies seek to optimize error and ranking metrics, few conduct investment simulations to assess the actual performance of trading strategies based on these predictions (Arslan, 2024; Tri Wahyuni et al., 2024). Investment simulations provide a practical insight into how price predictions can be used to make trading decisions and maximize returns (Brini & Lenz, 2024; Kim, Jeong, & Jeong, 2024).

Another critical aspect that often does not receive proper attention is minimizing risk in a highly volatile cryptocurrency market. Implementing diversification techniques, especially utilizing advanced algorithms such as Hierarchical Risk Parity (HRP), can be vital to building more resilient portfolios and reducing exposure to significant losses (Jeleskovic, Latini, Younas, & Al-Faryan, 2024; Majumder, 2024). Recent studies also highlight the effectiveness of models such as GARCH-Copula and entropy-based techniques for portfolio optimization, which show significant improvements in risk-adjusted performance (Giunta, Orlando, Carleo, & Ricci, 2024; He & Hamori, 2024).

Additionally, the integration of macroeconomic variables and sentiment analysis has shown a significant impact on forecasts and trading strategies. Recent research shows that variables such as the consumer confidence index and the consumer price index, as well as sentiments drawn from social media platforms, can significantly improve the accuracy of predictive models (Anand & Arya, 2024; Tzeng & His, 2024). These findings suggest combining traditional financial data with sentiment analysis can provide a competitive advantage in trading strategies.

This paper addresses these gaps in the literature by proposing an integrated approach that connects the ranking metrics of price predictions with returns for investors, conducts investment simulations, and explores diversification techniques for risk minimization. In particular, the present study seeks to demonstrate the relationship between the accuracy of forecasts and actual returns for investors by assessing how different metrics affect the performance of trading strategies through backtesting (Ahmed, 2024; Yang et al., 2023). Investment simulations were conducted to assess the impact of forecasts on portfolio performance, develop and test trading strategies based on price predictions, and analyze the results in terms of cumulative returns, Sharpe ratio, maximum drawdown, and other financial metrics (Dip Das, Thulasiram, Henry, & Thavaneswaran, 2024; Ni, Chiang, Day, & Chen, 2024). Risk minimization was also explored by using diversification techniques, implementing and comparing optimized portfolios using HRP with traditional portfolios, evaluating volatility reduction, and improving risk-adjusted performance (Giunta et al., 2024; He & Hamori, 2024). By integrating these approaches, it was sought to provide a more complete and practical view of cryptocurrency prediction and trading strategies, contributing to academic research and market practice.

Literature Review

The volatility of cryptocurrencies and their rapid price changes present considerable challenges for investors, but they also open up significant opportunities for applying advanced predictive models. This context sets the stage for the research presented in this paper. Sebastião e Godinho (2021) demonstrate the effectiveness of machine learning models, including staking techniques, in predicting cryptocurrency returns. These models, which combine the ability of several algorithms to capture complex and non-linear patterns, significantly improve the accuracy of forecasts and, consequently, the potential for investor returns. The application of staking models in cryptocurrencies allows the integration of multiple sources of information and predictive techniques, resulting in more robust trading strategies that are adaptive to the volatile dynamics of the cryptocurrency market.

Wiranata and Djunaidy (2021) systematically reviewed the literature on stock market forecasting techniques, analyzing 81 studies published between January 2015 and June 2020. Among the forecasting approaches, stacking was highlighted as an effective meta-learning technique. The base models included Random Forest (RF), Extra Trees (ERT), LightGBM (LGBM), Recurrent Neural Network (RNN), Bidirectional Recurrent Neural Network (BRNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The meta-model used was logistic regression (LR). The analysis revealed that by combining technical and macroeconomic data to predict indices such as the Nasdaq, Dow Jones, and S&P 500, stacking achieved an accuracy of 70.74%. The review highlighted that the stacking technique significantly improves the accuracy of predictions by optimally combining the base models' predictions, demonstrating a robust approach to predicting stock returns.

Muslim and Dasril (2021) proposed a framework for predicting the bankruptcy of companies based on the most influential characteristics, using XGBoost and a stacking ensemble. The data sample was obtained from Polish companies listed between 2000 and 2012, consisting of 65 characteristics and 42,627 records. For stacking, the base models included K-nearest neighbor, decision tree, support vector machines (SVM), and random forest, while the meta-learner was LightGBM. The trait selection process used the importance of XGBoost traits, selecting those with a weight greater than 10. The results showed that the stacking model outperformed the base models, achieving an accuracy rate of 97%, highlighting the effectiveness of stacking in improving the accuracy of predicting business failure.

In a novel approach, Zhao and Cheng (2022) applied a clustering technique to merge and refine multiple individual models for predicting linear and nonlinear stock returns. Their study, which focused on predicting the excess return of the U.S. market, introduced a wide range of base models, including linear regression, weighted least squares, Huber regression, kernel regression, nonparametric regression, LASSO, elastic net, gradient-boosted regression trees (GBRT), random forest (RF), complete subset regression, Mallow's model averaging, principal component regression (PCR), and neural networks (NN). The meta-models used were RF, GBRT, and NN. The results, which indicated that stacking with a slightly more complex meta-model outperformed traditional benchmarks in both in-sample and out-of-sample performance measures, such as the historical average and several other prediction combination models, are sure to pique the interest of academic researchers, financial analysts, and investors.

Rao et al., (2023) investigated the prediction of cryptocurrency time series using deep learning techniques in conjunction with ensemble learning methods. The study, which focused on predicting the hourly prices of Bitcoin, Ethereum, and XRP, combined traditional deep learning models (LSTM, BiLSTM, and convolutional layers) with ensemble methods such as averaging, bagging, and stacking. For stacking, they used logistic regression (LR) as a meta-learner. The potential benefits of their findings, which showed that stacking significantly improved prediction accuracy compared to individual models, are promising for financial analysts and investors. Metrics included RMSE, accuracy, AUC, and F1-score, highlighting the efficiency of ensemble models in predicting cryptocurrency prices.

The relationship between the accuracy of forecasts and actual returns for investors is a critical aspect that deserves greater attention. This topic, which is of utmost importance in the field of financial research, has been the focus of numerous studies, including AlMadany et al., (2024) and Muchtadi-Alamsyah et al., (2024) focus on evaluating predictive models using error metrics such as MSE, RMSE, MAE, and MAPE. However, as highlighted by Hewamalage, Ackermann, and Bergmeir, (2023) Using these metrics in isolation may not adequately reflect the actual impact for the investor.

Shintate and Pichl (2019) developed a classification framework for forecasting trends in high-frequency Bitcoin time series utilizing deep learning methods. They introduced the Random Sampling Method (RSM) to mitigate problems of non-stationarity and class imbalance in Bitcoin prices. The results showed that while the RSM outperformed the LSTM and MLP models in classification accuracy (with an F1 score of 0.5092 for BTCCNY and 0.5367 for BTCUSD), it was no more profitable than the buy-and-hold strategy, achieving lower returns during the testing period. This suggests that even sophisticated prediction models may not always result in superior trading performance, especially in highly volatile markets.

Nasirtafreshi (2022) compared machine learning, deep, and ensemble models for cryptocurrency price prediction. They evaluated models such as ARIMA, k-NN, SVR, RF, LSTM, GRU, TCN, and TFT in various cryptocurrencies. Even though LSTM achieved the best predictive performance with an average RMSE of 0.0222 and MAE of 0.0173, their trading simulations revealed that these predictive models did not consistently lead to improved trading strategies. The authors noted that cryptocurrencies' high volatility and unpredictable nature often resulted in discrepancies between predicted and actual prices, resulting in trading strategies that did not outperform basic approaches like the buy-and-hold strategy.

Previous studies have noted that the results do not exceed benchmarks, even with complex models and interesting error adjustments. However, the potential of predictive models to maximize returns and minimize risk is a beacon of hope. Analyzing how different error or classification metrics behave is essential, but just as important is analyzing the performance of trading strategies to validate the effectiveness of predictive models. In this sense, detailed backtesting simulations can provide insight into the impact of forecasts on portfolio performance, assisting investors in maximizing returns and managing risk.

Conducting investment simulations is essential for evaluating the performance of cryptocurrency price predictions. Many studies have highlighted the importance of validating predictive models using backtesting, a process that provides a solid foundation for the reliability of the models. This allows for analyzing cumulative returns, Sharpe ratio, maximum drawdown, and other financial metrics. These simulations not only validate the models but also provide a practical insight into how predictions can be used to make well-informed trading decisions, empowering investors to maximize returns and minimize risk. The results of these simulations are crucial for the validation of the models in real-world scenarios, offering a measure of the effectiveness of forecast-based trading strategies. (Arslan, 2024; Muchtadi-Alamsyah et al., 2024; Tri Wade et al., 2024).

Kim et al., (2023) Our investigation involved combining convolutional neural networks (CNN) with Grad-CAM to enhance the profitability of investment strategies applied to KOSPI 200 futures contracts. The goal was to boost the accuracy of future price direction forecasts and ensure the transparency of the decision process. Using KOSPI 200 futures contract data from January 2018 to December 2021, we applied the backtesting technique to compare the returns of the proposed strategy with the benchmark strategies. The backtesting results in 2021 revealed that the proposed strategy demonstrated higher returns and lower volatility than the benchmark strategies, underscoring the model's potential to deliver consistent returns and reduce risk in financial markets.

Zhao et al. (2023) proposed a portfolio selection method based on deep reinforcement learning (DRL). The study aimed to improve the accuracy of modeling nonlinear correlations between asset prices and optimize investment policy to maximize cumulative return. The research involved backtest experiments covering specific trading periods, using cryptocurrency datasets (Bitcoin and others), S&P 500 stocks, and ETFs. The study demonstrated that the proposed method, by employing a policy network based on attention and dilated causal convolution, outperformed traditional and recent DRL methods in terms of cumulative return (APV), Sharpe Ratio (SR), and Calmar Ratio (CR). The backtest results indicated that the proposed model could effectively capture asset price correlations, resulting in improved financial performance and risk management.

Parente et al., (2024) developed a trading algorithm for cryptocurrencies using a Multi-Layer Perceptron (MLP) neural network model. The study aimed to find a reliable and profitable model to predict the price direction of crypto assets by validating it through a backtest on three cryptocurrencies (Bitcoin, Ethereum, and Algorand) under different market conditions (bullish, bearish, and stable). The dataset consisted of historical data from hundreds of cryptoassets with a temporal resolution of 4 hours. The results of the backtest showed that the MLP model, combined with a 10% stop-loss, provided substantially positive investment returns, especially in the long run, with Ethereum achieving an ROI of 165.91%. Compared to a naïve model, MLP has demonstrated superiority in terms of profitability and stability of capital curves, validating the effectiveness of the proposed trading strategy.

Muchtadi-Alamsyah et al. (2024) investigated the risk prediction of cryptocurrencies using heteroscedastic models based on support vector regression (SVR) combined with GARCH and GJR-GARCH models. The study aimed to protect against the heteroscedastic risks of significant cryptocurrencies (Bitcoin, Ethereum, Tether, and

Binance Coin) before and during the COVID-19 pandemic. Using backtesting, SVR-based models were compared with maximum likelihood-based models under assumptions of normal, Student's t, and asymmetric Student's t distributions. The results showed that SVR-based models produced more accurate volatility predictions, especially for Tether, due to low volatility. In addition, Value at Risk (VaR) and Expected Shortfall (ES) predictions with SVR-based models showed better accuracy at 1%, 5%, and 10% significance levels when compared to MLE-based models. However, the McNeil and Frey tests indicated a lower performance in predicting ES at the 1% significance level. The study also noted increased market risks for Bitcoin and Binance Coin during the pandemic, while Ethereum and Tether showed decreased risks.

Risk minimization in a highly volatile cryptocurrency market is a critical aspect that requires advanced diversification techniques. Studies such as those by Majumder (2024) and Jeleskovic et al. (2024) explore distinct approaches to portfolio optimization. Majumder (2024) analyzed the hedge capacity of cryptocurrencies compared to gold against macroeconomic shocks in emerging economies, highlighting that different cryptocurrencies can serve as new "digital gold" depending on the specific economic context of each country. On the other hand, Jeleskovic et al. (2024) used GARCH-Copula models within the Markowitz framework for portfolio optimization, demonstrating that including cryptocurrencies can improve the stability and risk-return profile of the portfolio. The combination of accurate predictive models with advanced diversification strategies can provide a significant advantage for investors, balancing the pursuit of high returns with the need to minimize risks.

The application of advanced diversification techniques, such as Hierarchical Risk Parity (HRP), has shown promising results in minimizing risk and improving the risk-adjusted performance of cryptocurrency portfolios. By organizing assets in a hierarchical structure based on their correlations, HRP allows for a more balanced risk allocation while minimizing exposure to severe losses. This approach is especially relevant in cryptocurrencies, where price fluctuations can be highly volatile.

López de Prado (2016) presented the HRP as a solution to the Critical Line Algorithm's (CLA) practical problems, such as instability, concentration, and underperformance. HRP uses modern graph theories and machine learning to construct portfolios, addressing the flaws of CLA, which requires the inversion of an often unstable covariance matrix. The HRP operates in three stages: hierarchical grouping, quasi-diagonalization, and recursive bisection, allowing efficient resource allocation without needing matrix inversion. The empirical results demonstrated that the HRP outperforms the CLA and the Inverse-Variance Portfolio (IVP) regarding out-of-sample variance, providing more robust and diversified portfolios. Monte Carlo experiments indicated that HRP reduces out-of-sample variance compared to CLA and IVP, improving the Sharpe ratio and portfolio resilience against idiosyncratic and systemic shocks.

Jain and Jain (2019) investigated whether machine learning-based portfolios, such as HRP, can outperform traditional risk-based portfolios by considering the incorrect specification of the covariance matrix. The study used data from the NIFTY 50 index of the Indian National Exchange from November 2010 to December 2016 for the estimation period and from January 2017 to December 2017 for the valuation period. The HRP, Minimum Variance Portfolio (MVP), Inverse Volatility Weighted Portfolio (IVWP), Equal Risk Contribution Portfolio (ERC), and Maximum Diversification Portfolio (MDP) methods were compared. The results showed that the HRP, by avoiding the inversion of the covariance matrix, presented intermediate robustness to the incorrect covariance specification compared to the traditional methods. The HRP was less sensitive to the incorrect covariance specification than the MVP and MDP but not as robust as the IVWP.

The analysis of out-of-sample performance, using the Superior Predictive Ability Test (SPA), indicated that the HRP maintained a competitive performance in different horizons of rebalancing. Metrics used included portfolio variance, Conditional Value-at-Risk (CVaR), Herfindahl ratio, and Sharpe Ratio.

Sen et al. (2021) proposed a systematic approach to portfolio design using CLA and HRP algorithms across eight sectors of the Indian stock market. The sample included stock price data from January 1, 2016, to December 31, 2020, for training and from January 1, 2021, to August 26, 2021, for testing. The results indicated that while CLA outperformed in the training data, HRP outperformed CLA in the test data due to its robustness against instabilities in the covariance matrix, unlike CLA, which is sensitive to slight variations in return predictions. The HRP does not require the covariance matrix of returns to be invertible, being able to operate even with a singular matrix, using graph theory and machine learning techniques for weight allocation. The metrics used included volatility and Sharpe Ratio, and the backtesting results highlighted the greater efficiency of the HRP in the out-of-sample data.

Kim et al., (2024) examined the impact of incorporating cryptocurrencies into global asset portfolios utilizing ensemble approaches and a tracking strategy. The sample included daily adjusted price data from ETFs, U.S. Treasuries, gold, and cryptocurrencies from Jan. 10, 2018, to Jan. 11, 2023. MVP, MDP, ERCP, and HRP were used as benchmarks. The results highlighted HRP as a preferred strategy in the 3-month RTP strategy, outperforming HRP alone. Crypto allocation has improved the performance metrics of ensemble portfolios, but it has also increased risk. The metrics used included cumulative return, annualized return, annualized volatility, Sharpe ratio, maximum drawdown, and Calmar ratio.

Methodology

The methodology presented aims to predict the price movements of cryptocurrencies using a stacking approach of machine learning models. This approach combines several estimators to improve the robustness and accuracy of forecasts. The data includes technical indicators and market variables, applied to different cryptocurrencies over time. The utilization of stacked models aims to capture different market patterns and dynamics, as addressed by studies on the effectiveness of ensemble and machine learning techniques (AlMadany et al., 2024; Yang et al., 2023).

Data

The data was extracted from the Binance exchange through its API. The exchange had 507 cryptocurrencies listed on 07/14/2024, and it was decided to use those that presented the 50 most significant data periods available. Except for BTC-USD and LTC-USD pairs, which have data available since 2014, the daily data for the other cryptocurrency pairs refers to the period from November 2017 to July 2024. The available period was used if the cryptocurrency did not present data for the entire period.

Studies like this are expected to observe the selection of cryptocurrencies based on their market value. However, it was decided not to adopt this criterion so as not to run the risk of introducing a bias in the sample since they are the currencies that have appreciated the most, that is, if it is intended here to create a robust and profitable cryptocurrency portfolio, It would not make sense to select the ones that appreciated the most, after all, the result would be artificially positive. The sample included daily prices and technical indicators calculated with different time windows (6, 12, and 18 periods).

The variables used include RSI, MACD, ADX, CCI, and EMA indicators. In addition, temporal variables such as day, month, year, and day of the week were included. Previous studies highlight the importance of integrating technical and temporal variables to improve the accuracy of forecasts. Finally, the target variable was created based on the return of one day ahead. The target variable received a value of 1 for a positive return and 0 for a value less than or equal to zero (Tri Wahyuni et al., 2024; Tzeng & Su, 2024).

Preprocessing

Data preprocessing is essential for converting raw data, which is often inconsistent, unformatted, and incomplete, into well-formed data. It is necessary to use various pre-processing techniques to ensure that the data is in the appropriate format since redundant or erroneous data can lead to incorrect analysis and insights. Machine learning algorithms are ineffective on raw data, making pre-processing a vital step. Data manipulation, a crucial phase in data mining, helps make sense of raw data by allowing for the integration of data from multiple sources and its incorporation into a compatible system (Elsayed, Elaleem, & Marie, 2024; Rajpurohit, Mhaske, Gaikwad, Ahirrao, & Dhamale, 2023).

The data was organized by ticker (identifier of each cryptocurrency) and sorted by date. Temporal variables (day, month, year, day of the week) were extracted from the dates. Time counting variables were created for each Ticker to capture the effect of time variation, considering that the models used are not naturally built to work with time series. Null values have been zeroed. To deal with the class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) (Thanh-Long, Minh, & Hong-Chuong, 2022; Premalatha, Priyanka, & Chaitya, 2023). According to Yotsawat et al. (2023) when using SMOTE, attention should be paid to introducing noisy data, considering that the main idea of oversampling is to create generic data to compensate for the classes' imbalance. As the data in this article are slightly unbalanced, the amount of generic data created was negligible. Table 1 shows the amount of each class per cryptocurrency.

Table 1

Number of observations for each of the target variable classes

Ticker	0	1	total_count	percent_0	percent_1
BNT-USD	278	276	554	50.180505	49.819495
BTC-USD	359	348	707	50.777935	49.222065
BTG-USD	391	316	707	55.304102	44.695898
BTS-USD	383	324	707	54.17256	45.82744
CHESS-USD	33	40	73	45.205479	54.794521
CVC-USD	352	355	707	49.787836	50.212164
DASH-USD	335	372	707	47.38331	52.61669
DATA-USD	358	349	707	50.636492	49.363508
DCR-USD	363	344	707	51.343706	48.656294
DENT-USD	344	363	707	48.656294	51.343706
DGB-USD	360	347	707	50.919378	49.080622
DNT-USD	390	317	707	55.162659	44.837341
DOGE-USD	357	350	707	50.49505	49.50495
ENJ-USD	342	365	707	48.373409	51.626591
EOS-USD	350	357	707	49.50495	50.49505

ETC-USD	363	344	707	51.343706	48.656294
ETH-USD	352	355	707	49.787836	50.212164
FIRO-USD	337	370	707	47.666195	52.333805
FUN-USD	96	95	191	50.26178	49.73822
GAS-USD	357	350	707	50.49505	49.50495
GLM-USD	341	366	707	48.231966	51.768034
GNO-USD	358	349	707	50.636492	49.363508
HC-USD	381	326	707	53.889675	46.110325
ICX-USD	343	364	707	48.514851	51.485149
KMD-USD	355	352	707	50.212164	49.787836
LTC-USD	340	367	707	48.090523	51.909477
MILLION-USD	87	104	191	45.549738	54.450262
MTL-USD	335	372	707	47.38331	52.61669
NEBL-USD	365	342	707	51.626591	48.373409
NEO-USD	345	362	707	48.797737	51.202263
NULS-USD	343	364	707	48.514851	51.485149
OAX-USD	365	342	707	51.626591	48.373409
OMG-USD	347	360	707	49.080622	50.919378
PHB-USD	355	351	706	50.283286	49.716714
POWR-USD	335	372	707	47.38331	52.61669
QTUM-USD	336	371	707	47.524752	52.475248
REP-USD	382	325	707	54.031117	45.968883
REQ-USD	363	344	707	51.343706	48.656294
RLC-USD	342	365	707	48.373409	51.626591
STEEM-USD	331	376	707	46.817539	53.182461
STORJ-USD	345	362	707	48.797737	51.202263
STRAX-USD	357	348	705	50.638298	49.361702
VGX-USD	383	324	707	54.17256	45.82744
WAVES-USD	354	353	707	50.070721	49.929279
WTC-USD	343	338	681	50.367107	49.632893
VIEW-USD	349	358	707	49.363508	50.636492
XLM-USD	349	358	707	49.363508	50.636492
XMR-USD	316	391	707	44.695898	55.304102
ZRX-USD	352	355	707	49.787836	50.212164

Models

This article proposes using the general stacking approach to predict cryptocurrency movements. Intuitively, this algorithm fits a top-level model over a group of lower-level models to produce a refined forecast. The higher-level model is called the meta-model, which in the present study used Catboost.

The main difference between stacking and traditional prediction combination methods lies in two aspects: (1) the meta-model in stacking can be a complex function, while traditional methods use a simple function, such as mean or median; (2) lower models in stacking can be linear or nonlinear, while traditional approaches intensively use linear models.

The present study used XGBoost, Random Forest, LightGBM, Extra Trees, Support Vector Machine, and Gradient Boosting Machine as inferior models. Stacking is best suited for prediction problems with noisy, non-stationary, regime-shifting data generation processes. Although it is possible to optimize the hyperparameters of this set of models, we chose to use the default hyperparameters(Zhao & Cheng, 2022).

Training and Testing Framework

The data were divided into training, validation and test sets using cross-validation in training and mobile windows in validations and tests. The model was initially trained with all available currency pairs, i.e., the 50 pairs collected. After the training and validation stage, carried out from November 2017 to July 2023, only the best results (auc_roc > 0.50) were separated based on the auc_roc metric. The purpose of the separation was to confirm whether the performance would be repeated in the following period, in which the model was trained and tested, which comprised the period from August 2023 to July 2024. The selected assets are in bold in Table 2.

Table 2

Metrics obtained in the training and validation period that served as a filter for the training and testing period

Ticker	best_threshold	accuracy	f1_score	precision	recall	specificity	npv	roc_auc
BNT-USD	**0.1**	**0.493113**	**0.660517**	**0.493113**	**1**	**0**	**0**	**0.510718**
BTC-USD	0.1	0.484848	0.597849	0.476027	0.803468	0.194737	0.521127	0.496501
BTG-USD	**0.36**	**0.449036**	**0.609375**	**0.440678**	**0.987342**	**0.034146**	**0.777778**	**0.503921**
BTS-USD	0.1	0.471074	0.640449	0.471074	1	0	0	0.47207
CHESS-USD	**0.31**	**0.541667**	**0.666667**	**0.52381**	**0.916667**	**0.166667**	**0.666667**	**0.625**
CVC-USD	0.1	0.495868	0.660482	0.495822	0.988889	0.010929	0.5	0.487037
DASH-USD	**0.11**	**0.528926**	**0.688525**	**0.527933**	**0.989529**	**0.017442**	**0.6**	**0.523591**
DATA-USD	0.1	0.517906	0.682396	0.517906	1	0	0	0.473191
DCR-USD	**0.33**	**0.479339**	**0.64**	**0.473239**	**0.988235**	**0.031088**	**0.75**	**0.548217**
DENT-USD	0.1	0.504132	0.67033	0.504132	1	0	0	0.471008
DGB-USD	**0.27**	**0.509642**	**0.673993**	**0.508287**	**1**	**0.005587**	**1**	**0.525079**
DNT-USD	**0.22**	**0.426997**	**0.596899**	**0.425414**	**1**	**0.004785**	**1**	**0.50814**
DOGE-USD	0.1	0.490358	0.658041	0.490358	1	0	0	0.459794
ENJ-USD	**0.1**	**0.509642**	**0.675182**	**0.509642**	**1**	**0**	**0**	**0.50823**
EOS-USD	**0.35**	**0.523416**	**0.660118**	**0.507553**	**0.94382**	**0.118919**	**0.6875**	**0.528394**
ETC-USD	**0.14**	**0.487603**	**0.650376**	**0.48324**	**0.994253**	**0.021164**	**0.8**	**0.521134**
ETH-USD	0.27	0.493113	0.648855	0.487106	0.971429	0.047872	0.642857	0.445502
FIRO-USD	0.22	0.517906	0.681239	0.518006	0.994681	0.005714	0.5	0.439483
GAS-USD	**0.21**	**0.498623**	**0.660448**	**0.493036**	**1**	**0.021505**	**1**	**0.559747**
GLM-USD	0.1	0.509642	0.672794	0.511173	0.983871	0.011299	0.4	0.467499
GNO-USD	**0.1**	**0.490358**	**0.658041**	**0.490358**	**1**	**0**	**0**	**0.529578**
HC-USD	0.18	0.46832	0.633776	0.463889	1	0.015306	1	0.479378
ICX-USD	**0.1**	**0.495868**	**0.662983**	**0.495868**	**1**	**0**	**0**	**0.537492**
KMD-USD	0.23	0.484848	0.647834	0.479109	1	0.020942	1	0.472422
LTC-USD	**0.1**	**0.526171**	**0.689531**	**0.526171**	**1**	**0**	**0**	**0.553452**
MTL-USD	0.1	0.53168	0.694245	0.53168	1	0	0	0.496952

NEBL-USD	0.23	0.473829	0.63619	0.467787	0.994048	0.025641	0.833333	0.479548
NEO-USD	0.1	0.504132	0.67033	0.506925	0.989189	0	0	0.477316
NULS-USD	0.1	0.504132	0.67033	0.504132	1	0	0	0.462417
OAX-USD	0.1	0.476584	0.642857	0.475	0.994186	0.010471	0.666667	0.45638
OMG-USD	**0.36**	**0.523416**	**0.6742**	**0.51585**	**0.972826**	**0.061453**	**0.6875**	**0.556352**
PHB-USD	0.1	0.477901	0.646729	0.477901	1	0	0	0.415543
POWR-USD	0.1	0.526171	0.682657	0.525568	0.973684	0.034682	0.545455	0.495558
QTUM-USD	0.14	0.553719	0.708633	0.548747	1	0.024096	1	0.482203
REP-USD	0.24	0.473829	0.631985	0.461972	1	0.040201	1	0.484128
REQ-USD	0.28	0.493113	0.651515	0.484507	0.99422	0.036842	0.875	0.47481
RLC-USD	**0.1**	**0.528926**	**0.689655**	**0.526316**	**1**	**0.011561**	**1**	**0.51226**
STEEM-USD	**0.1**	**0.528926**	**0.691892**	**0.530387**	**0.994819**	**0**	**0**	**0.504877**
STORJ-USD	0.12	0.523416	0.684882	0.522222	0.994709	0.011494	0.666667	0.467098
STRAX-USD	**0.1**	**0.493113**	**0.660517**	**0.494475**	**0.994444**	**0**	**0**	**0.525106**
SYS-USD	**0.1**	**0.471074**	**0.640449**	**0.471074**	**1**	**0**	**0**	**0.530763**
VGX-USD	0.34	0.451791	0.618042	0.448468	0.993827	0.014925	0.75	0.438886
VIB-USD	**0.15**	**0.498623**	**0.656604**	**0.490141**	**0.994286**	**0.037234**	**0.875**	**0.509483**
WAVES-USD	0.1	0.490358	0.658041	0.490358	1	0	0	0.495263
WTC-USD	0.2	0.517906	0.677716	0.513966	0.994595	0.022472	0.8	0.489918
VIEW-USD	0.11	0.515152	0.676471	0.512535	0.994595	0.016854	0.75	0.49174
XLM-USD	**0.1**	**0.520661**	**0.684783**	**0.520661**	**1**	**0**	**0**	**0.517667**
XMR-USD	0.1	0.550964	0.700917	0.553623	0.955	0.055215	0.5	0.493558
ZRX-USD	0.16	0.504132	0.669118	0.502762	1	0.005525	1	0.479752

Table 2 notes that the two oldest and most traditional cryptocurrencies were part of the training and validation stage but did not enter the training and testing stage, as they did not reach the minimum required in the auc_roc. It is worth noting that auc_roc is a widely used evaluation metric for classification models, and values below 0.50 mean that the model's classification power is no better than a random guess. Despite this, Bitcoin was used as a benchmark for evaluating portfolios.

The literature often shows that the order of the data must be respected to use cross-validation with time series data. Despite this, some authors have argued that this order is not necessary and that using conventional cross-validation (with random order) can improve the performance of models (Hewamalage et al., 2023). A similar idea is also seen when training time series using BiLSTM models, in which the models are trained not only in the original sense of the series but also in reverse (Mizdrakovic et al., 2024; Zhang, Ye, & Lai, 2023). Notably, despite adopting this measure for cross-validation, the series' original order was maintained in the context of moving windows.

As daily data was used, it was decided to move the moving windows at the same frequency, keeping the start date fixed and moving only the end date. In this sense, right after the last date of the training set, the test is carried out later. The process was repeated over several iterations, moving the time window forward with each step. In each iteration, the training data was balanced using SMOTE. The base models were trained on the balanced dataset. The stacking model was trained using the predictions of the base models as features, which is the central idea of stacking.

Model evaluation metrics

After the models are executed, cryptocurrencies receive ratings based on their returns. Those classified as having positive returns and presenting a positive value are called "True Positive (TP)." Those classified as having positive returns but which presented negative returns are called "False Positive (FP)." Cryptocurrencies classified as having negative returns and showing positive returns are called "False Negative (FN)." Finally, those classified as having negative returns and that showed negative returns are called "True Negative (TN)." Based on these classifications, the following indicators could be calculated:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{F1 Score} = 2\,\frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5}$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{6}$$

Accuracy measures the proportion of all correct predictions (true positives and true negatives) to the total number of predictions made. This metric provides an initial insight into the model's performance, indicating how often it gets its predictions right.

Accuracy evaluates the ratio of correct positive predictions to the total positive predictions. High accuracy indicates that few examples classified as positive are actually negative (low false positive rate), which is an important metric when the cost of false positives is high.

Sensitivity, or recall, measures the ratio of correct positive examples identified by the model to the total number of actual positive examples. High sensitivity indicates that few real positive examples are missed (low false negative rate), which is crucial in contexts where it is essential to identify the most favorable cases.

The F1 Score is the harmonic mean of accuracy and sensitivity. This metric is useful when a balance between accuracy and recall is required, especially in unbalanced classes.

Specificity measures the proportion of true negatives correctly identified by the model about the total number of real negative examples. High specificity indicates that few actual negative examples are classified as positive (low false positive rate), which is essential when the cost of false positives is significant.

Negative Predictive Value (NPV) evaluates the ratio of correct negative predictions to total negative predictions. High NPV indicates that few examples classified as negative are actually positive (low false negative rate), being helpful when it is essential to confirm negative results.

Finally, AUC-ROC is a metric that evaluates the performance of a classification model at all possible classification thresholds. The Receiver Operating Characteristic (ROC) curve is a graph that shows the rate of true positives (sensitivity) versus the rate of false positives (1 - specificity) for different thresholds. The area under the curve (AUC) quantifies the model's ability to distinguish between positive and negative classes. An AUC of 1 indicates a perfect model, while an AUC of 0.5 indicates a model with no discriminative ability better than chance.

Implementation

The forecasting and portfolio allocation models were implemented using the Python programming language and libraries such as riskfolio-lib and vectorbtpro. The daily returns of the selected cryptocurrencies were calculated to serve as the basis for the analysis.

Portfolio allocation was optimized using the Hierarchical Risk Parity (HRP) method. HRP is an asset allocation technique that relies on the hierarchy of assets to build a robust portfolio. This method groups assets into clusters, minimizing the portfolio's total risk. Risk was measured using the Conditional Value at Risk (CVaR) metric, which measures the expected risk at the tails of the return distribution. CVaR is particularly useful for assets like cryptocurrencies, which can experience extreme price movements.

In addition to CVaR, the default HRP metric, other risk metrics, such as volatility (vol), Max Drawdown (MDD), Absolute Average of Differences (MAD), and Value at Risk (VaR), were used to calculate the portfolio's asset weights. These risk metrics are essential for capturing different aspects of the assets' risk behavior, providing a comprehensive view of the portfolio's risk.

Volatility (vol) measures the variability of an asset's or portfolio's returns, quantifying the deviation of returns from their average. Volatility is critical to understanding the total risk associated with the portfolio. Value at Risk (VaR) estimates the maximum potential loss of a portfolio over a given period with a certain level of confidence, while CVaR averages losses that exceed VaR, offering a detailed analysis of tail risk. The Historical High Drop (MDD) represents the most considerable peak-to-trough percentage loss in a portfolio's value over a given period and is crucial to avoid significant capital losses. The Mean Absolute Deviation (MAD) measures the average of the absolute differences between the actual returns and the average of the returns, capturing variability without considering the direction.

In this way, a portfolio was created for each risk metric. The assets were entered into the portfolio bought or shorted, depending on the prediction made by the staking model. The weights of each asset in the portfolio were defined using the HRP and varied according to the risk intended to be minimized. The ideal weights for the test period were calculated in the training and validation period and are available in Table 3.

Table 3

Cryptocurrency pairs selected for the training and testing stage with the weights of the allocation of each asset in the portfolio according to different risk metrics

Ticker	Vol	VaR	CVaR	MDD	MAD
BNT-USD	5.79%	5.56%	5.74%	7.16%	5.98%
BTG-USD	4.54%	4.02%	3.66%	2.39%	4.06%
CHESS-USD	3.63%	3.24%	3.44%	2.26%	3.39%
DASH-USD	5.76%	6.34%	6.08%	4.32%	6.39%
DCR-USD	3.51%	2.97%	2.99%	2.43%	3.18%
DGB-USD	3.46%	3.06%	3.02%	2.55%	3.30%
DNT-USD	5.93%	6.04%	6.46%	6.43%	5.54%
ENJ-USD	6.09%	6.09%	6.04%	7.20%	6.22%
EOS-USD	6.86%	7.04%	6.87%	8.57%	6.95%
ETC-USD	5.57%	6.50%	6.07%	6.17%	5.60%
GAS-USD	3.22%	2.68%	2.93%	3.25%	2.86%
GNO-USD	5.67%	5.83%	5.79%	4.70%	5.93%
ICX-USD	5.72%	5.90%	6.02%	7.19%	5.76%
LTC-USD	3.74%	3.83%	3.83%	3.80%	3.80%
OMG-USD	3.41%	3.21%	3.17%	3.41%	2.95%
RLC-USD	5.66%	5.63%	5.70%	4.89%	5.51%
STEEM-USD	2.53%	2.79%	2.92%	2.92%	2.87%
STRAX-USD	6.83%	6.53%	6.92%	7.00%	6.66%
SYS-USD	3.96%	3.56%	3.63%	2.96%	3.85%
VIB-USD	5.08%	6.30%	5.71%	7.68%	6.38%
XLM-USD	3.05%	2.88%	3.00%	2.70%	2.84%

Asset weights have been optimized for each risk metric, resulting in different portfolio allocations seeking to minimize risk per the metric considered. This approach allows for detailed and personalized portfolio analysis, ensuring that different aspects of risk are contemplated and providing a more robust and efficient asset allocation.

Generating Buy and Sell Signals

The buy and sell signals were generated based on the predictions of the stacking models, as highlighted in the previous subsection. Depending on the forecast, these signals were used to create long and short positions on the assets. To ensure that there was always an open position, composite signals were created that held a long or short position continuously. The results of the stacking models are given in terms of probability. From it, a cutoff point is defined to classify events and non-events. In this sense, the cutoff point was used, and the F1-Score was maximized. Table 4 shows the cutoff points used for each asset.

Table 4

Cutoff points used for event classification (buy signals) and non-event (sell or short sell signals)

Ticker	best_threshold	accuracy	f1_score	precision	recall	specificity	npv	roc_auc

BNT-USD	0.1	0.507853	0.673611	0.507853	1	0	0	0.555495
BTG-USD	0.33	0.479651	0.636917	0.468657	0.993671	0.043011	0.888889	0.523853
CHESS-USD	0.13	0.591837	0.736842	0.583333	1	0.047619	1	0.530612
DASH-USD	0.1	0.520349	0.684512	0.523392	0.98895	0	0	0.495475
DCR-USD	0.38	0.531977	0.67992	0.519757	0.982759	0.070588	0.8	0.509398
DGB-USD	0.37	0.505814	0.650206	0.489164	0.969325	0.088398	0.761905	0.533437
DNT-USD	0.1	0.473837	0.642998	0.473837	1	0	0	0.48307
ENJ-USD	0.16	0.526163	0.688337	0.524781	1	0.006098	1	0.516159
EOS-USD	0.34	0.540698	0.687747	0.53211	0.972067	0.072727	0.705882	0.51417
ETC-USD	0.2	0.5	0.664062	0.497076	1	0.011494	1	0.50284
GAS-USD	0.1	0.502907	0.669246	0.502907	1	0	0	0.5012
GNO-USD	0.16	0.5	0.664063	0.498534	0.994152	0.011561	0.666667	0.449346
ICX-USD	0.1	0.534884	0.69697	0.534884	1	0	0	0.486821
LTC-USD	0.1	0.511628	0.676923	0.511628	1	0	0	0.508185
OMG-USD	0.1	0.511628	0.676923	0.511628	1	0	0	0.518804
RLC-USD	0.1	0.505814	0.671815	0.507289	0.994286	0	0	0.545258
STEEM-USD	0.1	0.52907	0.692015	0.530612	0.994536	0	0	0.498252
STRAX-USD	0.1	0.491228	0.658824	0.491228	1	0	0	0.515463
SYS-USD	0.37	0.523416	0.667946	0.510264	0.966667	0.087432	0.727273	0.576715
VIB-USD	0.16	0.540698	0.696154	0.535503	0.994505	0.030864	0.833333	0.511057
XLM-USD	0.23	0.497093	0.660118	0.494118	0.994083	0.017143	0.75	0.511344

Table 4 shows that not all assets maintained roc_auc values above 0.50. Despite this, values above 0.50 will no longer be filtered for the test period, as was done in the validation period, at the risk of introducing a bias towards the assets with the best ratings in the forecasts.

Transaction and slippage fees have been configured to simulate a realistic trading environment. The transaction fee was 0.1% per trade, while the slippage was 0.05%. In addition, a 5% stop-loss has been implemented to limit losses on adverse position element (Parente et al., 2024). Thus, the returns presented are already with these adjustments.

Mann-Whitney mean difference tests were performed for returns and risk metrics to assess the forecasts' effectiveness and the portfolio's robustness. These tests allowed us to verify whether the forecasts and portfolio allocations resulted in significantly different returns from historical averages and whether they minimized risk as expected.

Results

The results presented in Table 5 show a detailed descriptive analysis of the assets selected for the study. This analysis is crucial to understanding the behavior of assets over the period studied, including variables such as start and end date, minimum price, maximum price, average price, closing price, and total return.

From the descriptive analysis of the assets, significant variability in returns is observed, with some assets showing sharp negative returns, such as WTC-USD and XEM-USD. In contrast, others, such as BTC-USD and CHESS-USD, have shown impressive returns. This variability highlights the volatile and unpredictable nature of the cryptocurrency market, reinforcing the importance of robust risk management strategies. It is worth noting that Table 5 presents the complete list of assets, but not all of them were allocated for testing, as mentioned in the methodological section. For example, despite not being allocated in the test wallet, Bitcoin was used as a benchmark (buy & hold) because it is the most representative currency in the cryptocurrency market.

Table 5

Descriptive analysis of assets

Ticker	Start Date	End Date	Min Close	Max Close	Mean Close	Last Close	Total Return (%)
BNT-USD	09/11/2017	30/06/2024	0.1371	10.0960	1.7058	0.6311	-68.95%
BTC-USD	17/09/2014	30/06/2024	178.1030	73083.5000	16876.2936	62678.2930	13605.15%
BTG-USD	09/11/2017	30/06/2024	4.9276	453.4550	35.0497	25.5122	-83.72%
BTS-USD	09/11/2017	30/06/2024	0.0024	0.8919	0.0566	0.0024	-97.12%
CHESS-USD	09/11/2017	30/06/2024	0.0005	0.1213	0.0122	0.0834	4118.41%
CVC-USD	09/11/2017	30/06/2024	0.0145	1.3479	0.1748	0.1066	-66.17%
DASH-USD	09/11/2017	30/06/2024	23.0960	1550.8500	142.9235	24.7961	-92.39%
DATA-USD	03/11/2017	30/06/2024	0.0059	0.3431	0.0581	0.0457	-28.24%
DCR-USD	09/11/2017	30/06/2024	9.2874	246.9047	45.8272	15.2379	-53.91%
DENT-USD	09/11/2017	30/06/2024	0.0001	0.0989	0.0031	0.0010	148.42%
DGB-USD	09/11/2017	30/06/2024	0.0031	0.1571	0.0228	0.0083	-24.57%
DNT-USD	09/11/2017	30/06/2024	0.0029	0.4033	0.0597	0.0450	-8.61%
DOGE-USD	09/11/2017	30/06/2024	0.0010	0.6848	0.0683	0.1243	8684.81%
ENJ-USD	09/11/2017	30/06/2024	0.0182	4.6858	0.5294	0.1942	701.75%
EOS-USD	09/11/2017	30/06/2024	0.5346	21.5426	3.4726	0.5772	-51.12%
ETC-USD	09/11/2017	30/06/2024	3.4724	134.1018	20.2486	23.6685	66.57%
ETH-USD	09/11/2017	30/06/2024	84.3083	4812.0874	1381.2803	3432.8892	969.82%
FIRO-USD	09/11/2017	30/06/2024	1.2052	142.4340	9.0494	1.2428	-93.18%
FUN-USD	09/11/2017	30/06/2024	0.0013	0.1926	0.0134	0.0038	-77.61%
GAS-USD	09/11/2017	30/06/2024	0.7567	86.0602	6.5467	3.7424	-83.75%
GLM-USD	09/11/2017	30/06/2024	0.0254	1.0884	0.2502	0.3529	63.85%
GNO-USD	09/11/2017	30/06/2024	8.7952	580.7632	130.9662	285.4750	249.73%
HC-USD	09/11/2017	30/06/2024	0.0214	37.0015	2.0322	0.0214	-99.81%
ICX-USD	09/11/2017	30/06/2024	0.1097	12.1884	0.8536	0.1611	-86.83%
KMD-USD	09/11/2017	30/06/2024	0.1692	11.4424	1.0235	0.3130	-88.44%
LTC-USD	17/09/2014	30/06/2024	1.1570	386.4508	69.0613	75.2800	1388.17%
MILLION-USD	09/11/2017	30/06/2024	2.1662	256.6360	36.2284	17.0901	-72.17%
MTL-USD	09/11/2017	30/06/2024	0.1516	11.2406	1.4945	1.1570	-78.66%
NEBL-USD	09/11/2017	30/06/2024	0.0061	44.6762	1.8742	0.0070	-99.86%
NEO-USD	09/11/2017	30/06/2024	5.3772	187.4050	23.6681	11.5971	-63.65%
NULS-USD	09/11/2017	30/06/2024	0.1148	8.2295	0.7068	0.3515	-61.46%
OAX-USD	09/11/2017	30/06/2024	0.0210	2.2315	0.2089	0.1775	-63.53%
OMG-USD	09/11/2017	30/06/2024	0.3097	25.7170	3.7852	0.3472	-95.67%
PHB-USD	09/11/2017	30/06/2024	0.0014	3.7563	0.3052	1.8996	3552.07%
POWR-USD	09/11/2017	30/06/2024	0.0349	1.8084	0.2484	0.2172	15.89%
QTUM-USD	09/11/2017	30/06/2024	1.0419	94.6719	6.2783	2.6133	-78.11%
REP-USD	09/11/2017	30/06/2024	0.4662	108.4720	15.9724	0.8370	-95.62%
REQ-USD	09/11/2017	30/06/2024	0.0058	0.9884	0.1036	0.1096	61.18%
RLC-USD	09/11/2017	30/06/2024	0.1541	11.6510	1.4777	2.1061	225.82%
STEEM-USD	09/11/2017	30/06/2024	0.1128	8.0308	0.5907	0.2014	-79.63%

STORJ-USD	09/11/2017	30/06/2024	0.0640	3.2149	0.6006	0.3852	-34.74%
STRAX-USD	09/11/2017	30/06/2024	0.0436	21.7483	1.5224	0.0475	-98.73%
VGX-USD	09/11/2017	30/06/2024	0.0173	11.0170	1.0286	0.0941	-90.68%
WAVES-USD	09/11/2017	30/06/2024	0.5334	54.6127	6.0451	1.0015	-79.59%
WTC-USD	09/11/2017	30/06/2024	0.0007	41.7283	2.0879	0.0072	-99.88%
VIEW-USD	09/11/2017	30/06/2024	0.0138	1.8427	0.1302	0.0148	-93.31%
XLM-USD	09/11/2017	30/06/2024	0.0282	0.8962	0.1725	0.0911	128.03%
XMR-USD	09/11/2017	30/06/2024	33.0103	483.5836	150.0215	167.9201	39.03%
ZRX-USD	09/11/2017	30/06/2024	0.1376	2.3675	0.5382	0.3677	59.48%

Table 6 presents a comparative analysis between different investment strategies, including Bitcoin, an equal-weighted portfolio, and several portfolios created to minimize specific risk metrics (Vol, VaR, CVaR, MDD, MAD).

Table 6

Performance of Wallets Optimized for Minimization of Different Risk Metrics Compared to Equal-Weights Portfolio and Bitcoin

Metric	Bitcoin	Equal Weights	Vol	VaR	CVaR	MDD	MAD
Start Index	03/08/2023	03/08/2023	03/08/2023	03/08/2023	03/08/2023	03/08/2023	03/08/2023
End Index	10/07/2024	10/07/2024	10/07/2024	10/07/2024	10/07/2024	10/07/2024	10/07/2024
Total Duration	343 days	343 days	343 days	343 days	343 days	343 days	343 days
Total Return [%]	97.634968	46.2238	195.60748	181.92389	186.97844	141.21836	187.98737
Annualized Return [%]	106.46216	49.83122	216.88894	201.3029	207.05466	155.23378	208.20354
Annualized Volatility [%]	48.972075	60.672031	57.926323	56.264764	57.281369	54.611416	56.769666
Max Drawdown [%]	23.581781	47.231253	28.28888	27.884396	28.230427	28.881436	27.902455
Max Drawdown Duration	120 days	122 days	60 days	94 days	59 days	94 days	54 days
Sharpe Ratio	1.72421	0.967938	2.278526	2.240018	2.242967	1.989048	2.264826
Calmar Ratio	4.514594	1.055048	7.666933	7.219195	7.33445	5.374863	7.461836
Omega Ratio	1.29994	1.153472	1.439374	1.417515	1.42488	1.344596	1.428461
Sortino Ratio	2.74162	1.445659	3.72533	3.614117	3.636485	3.05643	3.671115
Skew	0.375042	0.250264	0.839841	0.711062	0.773678	0.211308	0.756005
Kurtosis	2.39125	2.531052	8.751315	7.457916	8.180204	4.11657	7.986935
Tail Ratio	1.148448	0.973382	1.368192	1.33518	1.233234	1.313212	1.309507
Common Sense Ratio	1.492914	1.122768	1.96934	1.892638	1.757209	1.76574	1.87058
Value at Risk	-0.040773	-0.051011	-0.036017	-0.035977	-0.037015	-0.038739	-0.036929

Note: Optimized portfolios are designed to minimize different risk metrics: Vol (Volatility): A measure of the variation in the returns of an asset or portfolio over time. Lower volatility indicates lower risk; VaR (Value at Risk) Estimates the maximum potential loss of a portfolio in a given period and with a certain level of confidence; CVaR (Conditional Value at Risk): Also known as Expected Shortfall, it measures the expected loss considering that the VaR has been exceeded; MDD (Max Drawdown): Refers to the most significant peak-to-trough drop in a portfolio, indicating the portfolio's worst performance over a period; MAD (Mean Absolute Deviation): Average of the absolute differences between a portfolio's returns and its average, indicating the dispersion of returns

Bitcoin had a total return of 97.63% and an annualized return of 106.46%, significantly higher than the total return of 46.22% and the annualized return of 49.83% of the equal-weights portfolio. This demonstrates Bitcoin's strong performance during the analyzed period. Additionally, Bitcoin's annualized volatility was 48.97%, while that of the equal-weight wallet was 60.67%, indicating more significant uncertainty and risk in the equal-weight wallet compared to Bitcoin. The Max Drawdown analysis showed that Bitcoin experienced a 23.58% drop within 120 days, while the equal-weights wallet had a 47.23% drop with a duration of 122 days, suggesting Bitcoin's greater resilience during adverse market periods.

In terms of Sharpe and Calmar ratios, Bitcoin had values of 1.72 and 4.51, respectively, compared to the values of 0.97 and 1.05 of the equal-weighted portfolio. These results indicate that Bitcoin has provided significantly better risk-adjusted returns. Other risk metrics, such as Omega Ratio, Sortino Ratio, Tail Ratio, and Common Sense Ratio, also showed superior values for Bitcoin, reaffirming its advantage in terms of risk-adjusted performance.

Portfolios created to minimize specific risk metrics also showed remarkable performance. The vol portfolio, created to minimize volatility, had an annualized return of 216.89% with a volatility of 57.93%. The Sharpe ratio of this portfolio was the highest among all strategies at 2.28, suggesting excellent risk-adjusted performance. Portfolios created to minimize VaR and CVaR had annualized returns of 201.30% and 207.05%, respectively, with volatilities of approximately 56%. Both Calmar ratios greater than 7 indicate high efficiency in risk management.

The MDD and MAD portfolios, created to minimize the maximum drop and mean absolute deviation, showed annualized returns of 155.23% and 208.20%, respectively. The MAD portfolio's Calmar ratio was the second highest, 7.46, after the Vol portfolio's, indicating a good risk-reward ratio.

These findings are consistent with research by Giunta et al., (2024), which highlighted the usefulness of entropy-based measures for selecting low-risk portfolios in high-volatility scenarios. The effectiveness of volatility prediction models is corroborated by the study by Tzeng e Su (2024), which demonstrated that macroeconomic variables can significantly improve the prediction of cryptocurrency volatility. The wallet weighted to minimize MDD showed a significant reduction in maximum drawdown compared to Bitcoin, indicating more effective risk management. This result is in line with the findings of Muchtadi-Alamsyah et al. (2024), who used support vector machine regression (SVR) models combined with GARCH to predict volatility with greater accuracy and manage risk more efficiently.

Table 7

Significance about Bitcoin

Wallet	U-statistic	p-value
Vol	61197	0.36073
VaR	61225	0.35509
CVaR	61204	0.35932
MDD	61076	0.38575
MAD	61226	0.35489

Note: The U-statistic and p-value values refer to the Mann-Whitney U test to compare wallets optimized against Bitcoin. A p-value of less than 0.05 would indicate a statistically significant difference between the optimized wallet and Bitcoin. However, p-values greater than 0.05 suggest that there is no significant difference between the optimized wallets and Bitcoin for the metrics analyzed

The statistical significance analyses in Tables 7 and 8 show that the differences in Vol, VaR, CVaR, MDD, and MAD wallets about Bitcoin are not statistically significant, with p-values greater than 0.35. Similarly, comparisons with the equal-weighted portfolio did not show statistical significance, with p-values higher than 0.39. This suggests that while there are differences in performance between portfolios, they are not statistically robust. This result can be attributed to the high volatility inherent in the cryptocurrency market, as discussed by Afshan et al. (2024), which highlighted the strong response of cryptocurrencies to external events and market shocks.

Table 8

Significance of Equal Weights

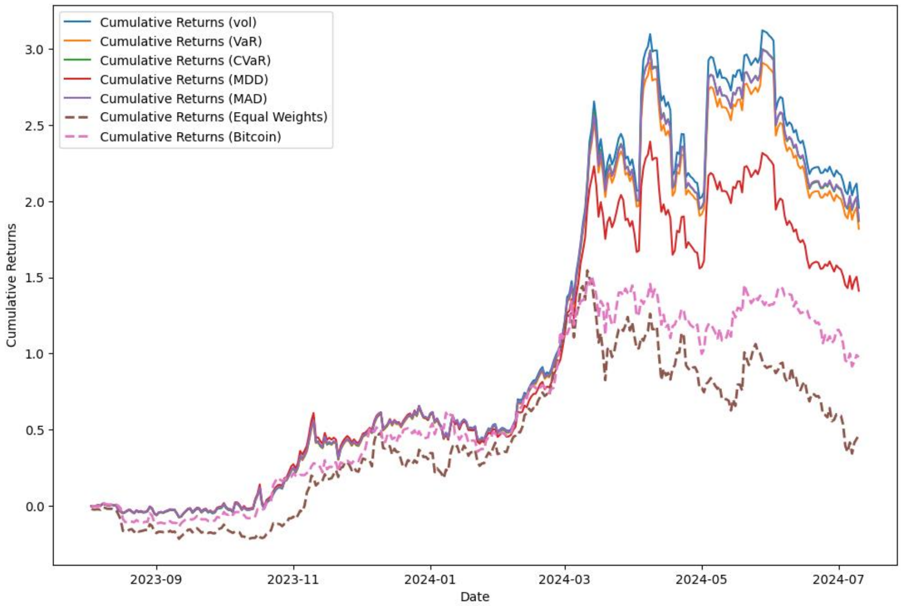
Wallet	U-statistic	p-value
Vol	60965	0.40961
VaR	61007	0.40048
CVaR	61011	0.39962
MDD	60853	0.43455
MAD	61023	0.39704

Note: The U-statistic and p-value values refer to the Mann-Whitney U test to compare optimized wallets against equal-weighted wallets. A p-value of less than 0.05 would indicate a statistically significant difference between the optimized and equal-weighted portfolios. However, p-values greater than 0.05 suggest that there is no significant difference between the optimized portfolios and the equal-weighted portfolio for the analyzed metrics

Despite its high volatility, the results show that Bitcoin offered superior risk-adjusted returns compared to the equal-weight portfolio. Portfolios that minimize specific risk metrics have demonstrated excellent performance, especially regarding annualized return and risk ratios. However, the lack of statistical significance in the comparisons highlights the need for caution when interpreting these results. While these portfolios look promising, statistical robustness is critical to confirming the validity of these results.

Figure 1

Cumulative return on portfolios



Note: The chart above shows the cumulative returns of the different optimized wallets (Vol, VaR, CVaR, MDD, MAD) compared to an equal-weight wallet and Bitcoin from August 2023 to July 2024. Cumulative returns have been calculated based on daily closing prices.

The results suggest that diversification and optimization based on risk metrics can significantly improve the performance of crypto portfolios. This is consistent with the findings of Seabe et al. (2024), which emphasized the predictive power of momentum and value factors in predicting crypto returns and Kim et al., (2024), which demonstrated that increasing allocation to cryptocurrencies can improve portfolio performance metrics despite increased volatility. From Figure 1, it is possible to observe the accumulated return over time.

The results of this study point to several directions for future research and highlight some limitations that should be considered. First, asset diversification and the inclusion of macroeconomic variables can improve the accuracy of forecasts and the robustness of investment strategies, as suggested by Tzeng e Su, (2024) e por Kim et al. (2024). Integrating liquidity metrics and considering the impact of external events, as highlighted by Ahmed (2024) and Afshan et al. (2024), are also crucial for a more comprehensive analysis.

In addition, the application of hybrid models that combine statistical and machine learning techniques, as demonstrated by Muchtadi-Alamsyah et al., (2024), can provide more accurate and efficient predictions. However, the cryptocurrency market's high volatility and dynamic nature pose ongoing challenges. Therefore, future research should focus on generalizing the models under different market conditions and conducting detailed backtesting analyses to validate the effectiveness of investment strategies.

Conclusion

The main objective of this study was to explore the application of stacking models to predict cryptocurrency price movements and optimize portfolio allocation based on different risk metrics. It also sought to evaluate the effectiveness of these investment strategies compared to Bitcoin's performance and an equal-weight portfolio.

The results demonstrated that while Bitcoin showed a substantially higher annualized return and better risk-adjusted performance than the equal-weights portfolio, portfolios created to minimize risk-specific metrics such as volatility, VaR, CVaR, MDD, and MAD showed remarkable performance. These optimized portfolios, especially those focused on minimizing volatility and mean absolute deviation, have shown superior annualized returns, reaffirming the effectiveness of risk diversification strategies in a highly volatile market like crypto. The vol portfolio, in particular, stood out with the highest Sharpe ratio, indicating excellent risk-adjusted performance.

However, statistical significance analyses revealed that the differences in Vol, VaR, CVaR, MDD, and MAD wallets relative to Bitcoin and the equal-weights wallet were not statistically significant, suggesting that the outperformance of these wallets may not be robust enough to be generalized. This observation is consistent with the high volatility inherent in the cryptocurrency market, which responds strongly to external events and market shocks.

Given this, the study's objectives were partially achieved. Stacking strategies and optimized portfolios have been shown to improve risk-adjusted performance, but the lack of statistical significance suggests the need for caution in generalizing these results. The practical implications of this study suggest that while advanced forecasting models and risk-minimizing strategies can improve portfolio performance, investors should be aware of the volatility and risks associated with cryptocurrencies.

References

Afshan, S., Yaqoob, T., Zaied, Y. Ben, Mishra, S., & Mishra, S. (2024). Oil shocks and currency behavior: A dual approach to digital and traditional currencies. *Global Finance Journal*, *62*. https://doi.org/10.1016/j.gfj.2024.101002

Ahmed, W. M. A. (2024). On the robust drivers of cryptocurrency liquidity: the case of Bitcoin. *Financial Innovation*, *10*(1). https://doi.org/10.1186/s40854-023-00598-9

AlMadany, N. N., Hujran, O., Naymat, G. Al, & Maghyereh, A. (2024). Forecasting cryptocurrency returns using classical statistical and deep learning techniques. *International Journal of Information Management Data Insights*, *4*(2). https://doi.org/10.1016/j.jjimei.2024.100251

Anand, H., & Arya, A. (2024). An Empirical Study of Financial BERT Models for Sentiment Analysis and Cryptocurrency Price Correlation. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–7. https://doi.org/10.1109/I2CT61223.2024.10543560

Arslan, S. (2024). Bitcoin Price Prediction Using Sentiment Analysis and Empirical Mode Decomposition. *Computational Economics*. https://doi.org/10.1007/s10614-024-10588-3

Brini, A., & Lenz, J. (2024). Pricing cryptocurrency options with machine learning regression for handling market volatility. *Economic Modelling*, *136*. https://doi.org/10.1016/j.econmod.2024.106752

Chandra, D., Tyagi, P., Gupta, R. S., Saxena, A. M., & Kharaliya, S. (2024). *Cryptocurrency Price Prediction using Machine Learning*. 258–264. https://doi.org/10.1109/incacct61598.2024.10551043

Dip Das, J., Thulasiram, R. K., Henry, C., & Thavaneswaran, A. (2024). Encoder–Decoder Based LSTM and GRU Architectures for Stocks and Cryptocurrency Prediction. *Journal of Risk and Financial Management*, *17*(5), 200. https://doi.org/10.3390/jrfm17050200

Elsayed, N., Elaleem, S. A., & Marie, M. (2024). Improving Prediction Accuracy using Random Forest Algorithm. *International Journal of Advanced Computer Science and Applications*, *15*(4). https://doi.org/10.14569/IJACSA.2024.0150445

Giunta, N., Orlando, G., Carleo, A., & Ricci, J. M. (2024). Exploring Entropy-Based Portfolio Strategies: Empirical Analysis and Cryptocurrency Impact. *Risks*, *12*(5), 78. https://doi.org/10.3390/risks12050078

Guo, L., Sang, B., Tu, J., & Wang, Y. (2024). Cross-cryptocurrency Return Predictability. *Journal of Economic Dynamics and Control*, 104863. https://doi.org/10.1016/j.jedc.2024.104863

He, X., & Hamori, S. (2024). The higher the better? Hedging and investment strategies in cryptocurrency markets: Insights from higher moment spillovers. *International Review of Financial Analysis*, 103359. https://doi.org/10.1016/j.irfa.2024.103359

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. https://doi.org/10.1007/s10618-022-00894-5

Jain, P., & Jain, S. (2019). Can machine learning-based portfolios outperform traditional risk-based portfolios? The need to account for covariance misspecification. *Risks*, *7*(3), 1–27. https://doi.org/10.3390/risks7030074

Jeleskovic, V., Latini, C., Younas, Z. I., & Al-Faryan, M. A. S. (2024). Cryptocurrency portfolio optimization: Utilizing a GARCH-copula model within the Markowitz framework. *Journal of Corporate Accounting and Finance*. https://doi.org/10.1002/jcaf.22721

Kim, M., Jin Jeong, Y., & Jeong, J. (2024). Two Empirical Studies of Portfolio Optimization Using Cryptocurrency Allocation Ratios. *IEEE Access*, *12*, 63827–63838. https://doi.org/10.1109/ACCESS.2024.3396495

Kim, S. H., Park, J. S., Lee, H. S., Yoo, S. H., & Oh, K. J. (2023). Combining CNN and Grad-CAM for profitability and explainability of investment strategy: Application to the KOSPI 200 futures. *Expert Systems with Applications*, *225*. https://doi.org/10.1016/j.eswa.2023.120086

Li, S., & Ma, J. (2024). The impact of sentiment and engagement of Twitter posts on cryptocurrency price movement. *Finance Research Letters*, *65*. https://doi.org/10.1016/j.frl.2024.105598

López de Prado, M. (2016). Building Diversified Portfolios that Outperform Out of Sample. *The Journal of Portfolio Management*, *42*(4), 59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Majumder, S. B. (2024). Is cryptocurrency a new digital gold? Evidence from the macroeconomic shocks in selected emerging economies. *Journal of Economic Studies*. https://doi.org/10.1108/JES-08-2023-0410

Mizdrakovic, V., Kljajic, M., Zivkovic, M., Bacanin, N., Jovanovic, L., Deveci, M., & Pedrycz, W. (2024). Forecasting bitcoin: Decomposition aided long short-term memory based time series modeling and its explanation with Shapley values. *Knowledge-Based Systems*, *299*. https://doi.org/10.1016/j.knosys.2024.112026

Muchtadi-Alamsyah, I., Viltoriano, R., Harjono, F., Nazaretha, M., Susilo, M., Bayu, A., Josaphat, B., Hakim, A., & Syuhada, K. (2024). Support vector regression-based heteroscedastic models for cryptocurrency risk forecasting. *Applied Soft Computing*, *162*. https://doi.org/10.1016/j.asoc.2024.111792

Muslim, M. A., & Dasril, Y. (2021). Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning. *International Journal of Electrical and Computer Engineering*, *11*(6), 5549–5557. https://doi.org/10.11591/ijece.v11i6.pp5549-5557

Nasirtafreshi, I. (2022). Forecasting cryptocurrency prices using Recurrent Neural Network and Long Short-term Memory. *Data & Knowledge Engineering*, 102009. https://doi.org/10.1016/j.datak.2022.102009

Ni, Y., Chiang, P., Day, M. Y., & Chen, Y. (2024). Using Big Data Analytics and Heatmap Matrix Visualization to Enhance Cryptocurrency Trading Decisions. *Applied Sciences (Switzerland)*, *14*(1). https://doi.org/10.3390/app14010154

Otabek, S., & Choi, J. (2024). From Prediction to Profit: A Comprehensive Review of Cryptocurrency Trading Strategies and Price Forecasting Techniques. *IEEE Access*, *12*, 87039–87064. https://doi.org/10.1109/ACCESS.2024.3417449

Parente, M., Rizzuti, L., & Trerotola, M. (2024). A profitable trading algorithm for cryptocurrencies using a Neural Network model. *Expert Systems with Applications*, *238*. https://doi.org/10.1016/j.eswa.2023.121806

Premalatha, G., Priyanka, R., & Chaitya, K. (2023). Feature selection for predicting bankruptcy: Comparative analysis. *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1–5. https://doi.org/10.1109/ICECCT56650.2023.10179633

Rajpurohit, A. S., Mhaske, H., Gaikwad, P. S., Ahirrao, S. P., & Dhamale, N. B. (2023). Data Preprocessing for Stock Price Prediction Using LSTM and Sentiment Analysis. *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, 1–6. https://doi.org/10.1109/ISCON57294.2023.10112026

Rao, K. R., Prasad, M. L., Kumar, G. R., Natchadalingam, R., Hussain, M. M., & Reddy, P. C. S. (2023). Time-Series Cryptocurrency Forecasting Using Ensemble

Deep Learning. *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, 1446–1451. https://doi.org/10.1109/ICCPCT58313.2023.10245083

Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2024). Optimizing Cryptocurrency Returns: A Quantitative Study on Factor-Based Investing. *Mathematics*, *12*(9). https://doi.org/10.3390/math12091351

Sebastião, H., & Godinho, P. (2021). Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, *7*(1). https://doi.org/10.1186/s40854-020-00217-x

Sen, J., Mehtab, S., Dutta, A., & Mondal, S. (2021). Hierarchical Risk Parity and Minimum Variance Portfolio Design on NIFTY 50 Stocks. *2021 International Conference on Decision Aid Sciences and Application (DASA)*, 668–675. https://doi.org/10.1109/DASA53625.2021.9681925

Shintate, T., & Pichl, L. (2019). Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning. *Journal of Risk and Financial Management*, *12*(1). https://doi.org/10.3390/jrfm12010017

Singla, M., Gill, K. S., Chauhan, R., Pokhariya, H. S., & Lande, J. (2024). Dogecoin price prediction – Through Utilization of XGBoost Model & GridSearchCV Classification Technique. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–6. https://doi.org/10.1109/I2CT61223.2024.10543645

Thanh-Long, N., Minh, T.-, & Hong-Chuong, L. (2022). A Back Propagation Neural Network Model with the Synthetic Minority Over-Sampling Technique for Construction Company Bankruptcy Prediction. *International Journal of Sustainable Construction Engineering and Technology*, *13*(3), 68–79. https://doi.org/10.30880/ijscet.2022.13.03.007

Tri Wahyuni, M., Ridwan, E., & Fitrizal Salim, D. (2024). US macroeconomic determinants of Bitcoin. *Investment Management and Financial Innovations*, *21*(2), 240–252. https://doi.org/10.21511/imfi.21(2).2024.19

Tzeng, K. Y., & Su, Y. K. (2024). Can U.S. macroeconomic indicators forecast cryptocurrency volatility? *North American Journal of Economics and Finance*, *74*. https://doi.org/10.1016/j.najef.2024.102224

Wiranata, R. B., & Djunaidy, A. (2021). The Stock Exchange Prediction using Machine Learning Techniques: A Comprehensive and Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informasi*, *14*(2), 91–112. https://doi.org/10.21609/jiki.v14i2.935

Yang, S., Ding, Y., Xie, B., Guo, Y., Bai, X., Qian, J., Gao, Y., Wang, W., & Ren, J. (2023). Advancing Financial Forecasts: A Deep Dive into Memory Attention and Long-Distance Loss in Stock Price Predictions. *Applied Sciences (Switzerland)*, *13*(22). https://doi.org/10.3390/app132212160

Yotsawat, W., Phodong, K., Promrat, T., & Wattuya, P. (2023). Bankruptcy prediction model using cost-sensitive extreme gradient boosting in the context of imbalanced datasets. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(4), 4683. https://doi.org/10.11591/ijece.v13i4.pp4683-4691

Zhang, J., Ye, L., & Lai, Y. (2023). Stock Price Prediction Using CNN-BiLSTM-Attention Model. *Mathematics*, *11*(9). https://doi.org/10.3390/math11091985

Zhao, A. B., & Cheng, T. (2022). Stock return prediction: Stacking a variety of models. *Journal of Empirical Finance*, *67*, 288–317. https://doi.org/10.1016/j.jempfin.2022.04.001

Zhao, T., Ma, X., Li, X., & Zhang, C. (2023). Asset correlation based deep reinforcement learning for the portfolio selection. *Expert Systems with Applications*, *221*. https://doi.org/10.1016/j.eswa.2023.119707