Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Leveraging Blockchain Analytics for Quantitative Insights in Decentralized Finance (DeFi) Markets

By

Chenyu Wang

Supervisor: Dr. Yang Su

Master of Science in Business Analytics

Trinity Business School

TRINITY COLLEGE

UNIVERSITY OF DUBLIN

Submitted in partial fulfilment of the requirements of the examination for MSc Business Analytics, Trinity College Dublin, July 2025.

Student declaration

I have read the University's code of practice on plagiarism. I hereby certify this material, which I now submit for assessment on the programme of study leading to the award of MSc Business Analytics is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited within the text of my work.

Student Name: Chenyu Wang

Student ID Number: 24336364

Student Signature: Chenyu Wang

Date: 15th July 2025

Abstract

Decentralized Finance (DeFi) has reshaped the financial landscape by enabling transparent, non-custodial, and programmable services through blockchain and smart contracts. This dissertation investigates the predictive power of blockchain analytics for short-term Ethereum (ETH) price forecasting within the DeFi ecosystem. It integrates on-chain indicators—such as Total Value Locked (TVL), gas fees, whale transactions, and smart contract activity—with technical indicators like RSI, MACD, and Bollinger Bands to build a feature-rich modeling framework.

Three machine learning models—Linear Regression, Random Forest, and XGBoost—were applied, alongside a stacking ensemble designed to combine complementary model strengths. Feature engineering captured both blockchain dynamics and price momentum. Models were evaluated using regression metrics and backtesting simulations reflecting real-world trading.

Empirical results show that Random Forest achieved the best statistical accuracy, while XGBoost delivered the highest portfolio returns and Sharpe Ratio. The stacking ensemble, however, provided no meaningful improvement and at times led to overfitting or diluted signals.

These findings highlight the importance of aligning model choice with forecasting goals—statistical precision versus trading performance—and challenge the assumption that complex ensembles always outperform strong individual models. This study offers a practical framework for using on-chain analytics in DeFi trading strategies, clarifying both the strengths and limits of blockchain-based features in financial prediction.

Keywords: blockchain analytics, decentralized finance, Ethereum forecasting, machine learning, XGBoost, Random Forest, feature engineering, quantitative modeling, backtesting, on-chain data

Acknowledgements

Time passes quickly. As my one-year Master's journey at Trinity College Dublin—and my 23rd year in life—draws to a close, I find myself filled with gratitude and reflection. This year has been more than an academic pursuit; it has been a time of growth, discovery, and unforgettable moments that will stay with me for life.

From my early days as an exchange student in Cork to life in Dublin, Ireland slowly revealed itself to me. I remember walking through Trinity's front arch as a tourist, wide-eyed at its historic charm. Back then, I could only imagine studying here. Today, I leave with knowledge, friendships, and experiences beyond what I hoped for.

I've been fortunate to meet many incredible people—friends and classmates who made this journey joyful and meaningful. Their support and warmth carried me through long nights of study and moments of doubt—often accompanied by shared laughter and, of course, the reliable comfort of a pint of Guinness (which surely deserves partial credit for this degree).

To Miss Geunju, Miss Bentley, Mr Patrick, and all those I worked with during the Master's programme: thank you for learning and growing with me. The lectures, group projects, and travels we shared are among my most cherished memories.

To my friends beyond the classroom: Mr Kipp and Mr Henry at the Binary Hub, Mr Mikael, Mr Robert, and Mr Kelvin from The Alibi Bar in Cork, and the musicians Mr Dan, Miss Camila, Miss Dafne, and others—thanks for making Ireland feel like home.

To my dearest family—thank you for your unwavering love and support across distance and time. Your belief in me carried me forward.

Finally, I offer my sincere thanks to my supervisor, Dr Yang Su. Her patient guidance helped shape this dissertation and guided me through to the end.

To everyone who walked with me on this path—I carry your kindness with me.

Thank you.

Contents

List of Figures 7

List of Tables 8

1. Introduction 9

    1.1 Context 9

    1.2 Research Problems 11

    1.3 Research Objectives 12

2. Literature Review 15

    2.1 Blockchain Technology and its Business Implications 15

    2.2 DeFi Protocols and Smart Contracts 18

    2.3 Cryptocurrency Market Behavior and Forecasting 21

    2.4 Conclusion 23

3. Methodology 25

    3.1 Methodological Framework Overview 25

    3.2 Data Collection and Preprocessing 27

    3.3 Feature Engineering 31

    3.4 Model Development 34

    3.5 Backtest 37

    3.6 Prediction 39

4. Data Analysis and Results 41

    4.1 Model Parameters 42

    4.2 Model Evaluation 44

        4.2.1 Predictive Accuracy Comparison 44

        4.2.2 Residual Analysis 45

        4.2.3 Feature Importance and Model Interpretability 47

    4.3 Backtest Result 50

    4.4 Prediction 60

    4.5 Conclusion 62

5. Discussion 63

5.1 Model Insights 63
5.2 Implications for DeFi Market Analysis 64
5.3 Limitations 65
6. Conclusion 66
6.1 Summary of Findings 66
6.2 Future Research Directions 67
6.3 Final Thoughts 68
7. Appendices: Codes, APIs and Datasets 70
8. References 72

List of Figures

Figure 1: Residuals Plot for XGB, RF and LR

Figure 2: Residuals Plot for STACKs

Figure 3: Backtest Portfolio Comparision

Figure 4: Portfolio Value For XGB, RF and LR

Figure 5: Portfolio Value For STACKs

Figure 6: Drawdown charts

Figure 7: Radar plots

Figure 8: The distribution of daily returns

Figure 9: Performance bar plots

Figure 10: Next-Day Price Prediction

List of Tables

Table 1: Smart Contract Use Cases in DeFi

Table 2: DeFi Risk Analysis

Table 3: Comparison of DeFi Applications

Table 4: Statistic Summary of Market Data

Table 5: Statistic Summary of On Chain Data

Table 6: Summary of Feature Types

Table 7: Comparison of Different Models

Table 8: Accuracy Metrics

Table 9: Random Forest Feature Importance

Table 10: XGBoost Feature Importance

Table 11: Linear Regression Coefficients

Table 12: Backtest Results

1. Introduction

1.1 Context

Blockchain-based cryptocurrencies have demonstrated how to securely implement traditionally centralized systems, such as currencies, in a decentralized fashion.(Gencer, 2018) This evolution has opened new possibilities for reconfiguring systems of trust, ownership, and governance across industries such as finance, healthcare, energy, and education (AlShamsi et al., 2023; Lizcano et al., 2020). What distinguishes blockchain from earlier digital systems is not simply its immutability or distributed structure, but its capacity to encode logic and incentives directly into transactional flows via smart contracts (Taherdoost, 2022).

One of the most significant applications of blockchain is DeFi, which removes the need for financial intermediaries by enabling peer-to-peer asset exchange and programmable economic relationships (Schär, 2021). Built primarily on the Ethereum network, DeFi protocols rely on smart contracts to support functions such as lending, trading, and liquidity provision. These architectures offer transparency and composability, yet they are not immune to emergent risks. Challenges such as rising gas fees during periods of network congestion, governance token concentration, and smart contract exploits highlight the tension between theoretical decentralization and practical centralization (Taherdoost, 2022).

Ethereum's role in this space makes it a critical locus for examining blockchain-based financial behavior. Common metrics like Total Value Locked (TVL), transaction throughput, gas fees, and the concentration of large token holders are often used as proxies for user activity and investor sentiment. However, the widespread adoption of these indicators has not been matched by critical scrutiny of their explanatory or predictive capacity. For instance, while TVL is frequently cited as evidence of protocol growth, it can be inflated through recursive lending or yield farming loops, which distort its representation of genuine user commitment (Schär, 2021).

As the scale and granularity of blockchain data increase, researchers have turned to machine learning (ML) models to extract signals from complex, high-frequency datasets. These efforts align with a broader movement to integrate data science into financial analysis, but the literature often falls short in addressing interpretability and economic validity. Black-box models may achieve short-term predictive gains, yet they risk obscuring causal mechanisms or overfitting to market noise—issues especially problematic in volatile, thinly regulated environments like DeFi.

Beyond predictive accuracy, the practical relevance of such models hinges on their integration into decision-making frameworks. Backtesting against historical data is a necessary, but not sufficient, condition for evaluating financial viability. As Lizcano et al. (2020) argue in the context of educational trust systems, the credibility of blockchain-based models depends not only on technical performance but also on their capacity to reflect systemic constraints and behavioral complexity. These insights are readily transferable to DeFi, where trust is encoded not through institutional guarantees but through code, incentives, and emergent user behavior.

This study builds on these concerns by critically assessing the utility of Ethereum-based on-chain indicators for short-term price prediction and trading strategy development. Instead of treating common metrics as inherently meaningful, it constructs a data-driven framework to test their relevance and interaction under real-time conditions. A range of machine learning models—including ensemble and stacking approaches—are applied and evaluated both statistically and financially, using rule-based backtesting to simulate real-world performance.

Ultimately, this research aims to clarify the extent to which blockchain-native features can inform quantitative models, and under what conditions they translate into actionable financial insights. By doing so, it contributes to an ongoing conversation at the intersection of blockchain analytics, financial systems, and data science.

1.2 Research Problems

Despite the rapid growth of DeFi, the analytical tools used to understand and forecast asset behavior within this ecosystem remain underdeveloped. Traditional financial models often struggle to adapt to the unique characteristics of blockchain-based markets, which operate continuously, lack centralized oversight, and generate vast, publicly accessible transactional datasets.

Recent literature identifies several unresolved methodological and practical challenges. First, there is no clear consensus on which blockchain-native metrics—such as gas fees, transaction volume, or Total Value Locked (TVL)—offer the most predictive power for asset pricing. While these variables are widely used in industry and academic studies, their efficacy under volatile or manipulated market conditions is not rigorously justified nor consistently evaluated across different contexts (Auer et al., 2023; Aramonte et al., 2021).

Second, although machine learning (ML) has shown promise in detecting patterns within high-dimensional financial data, its application to DeFi markets often suffers from a lack of interpretability and robustness. Many studies rely on black-box models, such as random forests or deep learning, that prioritize predictive accuracy without addressing the transparency required for real-world trading and risk management applications (Fischer et al., 2019; Stavroyiannis, 2018). This is particularly concerning in decentralized markets, where model explainability is crucial for trust and regulatory acceptance (Auer et al., 2023; Zhou et al., 2021).

Third, high statistical performance does not necessarily imply financial viability. Several works highlight the disconnect between in-sample accuracy and out-of-sample profitability (Fischer et al., 2019). Backtesting and performance benchmarking against trading strategies remain insufficiently explored, leaving a gap in validating whether ML-driven forecasts can support actionable investment decisions (Zhou et al., 2021).

These gaps motivate the following key research questions:

- Which on-chain metrics contain the most predictive information about Ethereum's short-term price movements? This aligns with the need to systematize and critically evaluate the signal quality of blockchain-native features (Auer et al., 2023).
- How do different machine learning models perform—both statistically and financially—when applied to Ethereum price prediction? Building on the contrast between interpretable models and black-box learners (Fischer et al., 2019; Stavroyiannis, 2018).
- What is the marginal contribution of feature engineering and model stacking to predictive accuracy and trading performance? This addresses the layered complexity and composability inherent in DeFi analytics (Auer et al., 2023).
- To what extent can explainable machine learning techniques reveal economically meaningful relationships in DeFi markets? Reflecting growing calls for model transparency in DeFi due to regulatory and systemic risk concerns (Aramonte et al., 2021; Zhou et al., 2021).

This study seeks to bridge the gap between predictive modeling and financial applicability by developing a holistic ML pipeline tailored to Ethereum-based DeFi markets. In doing so, it contributes to a deeper understanding of both the technical and economic dimensions of machine learning in decentralized environments.

1.3 Research Objectives

This dissertation investigates the extent to which blockchain-native data can support predictive modeling in the context of Ethereum price dynamics. It examines how machine learning techniques can extract meaningful signals from decentralized transaction environments and evaluates their practical relevance in a trading context. Rather than focusing solely on predictive accuracy, the study emphasizes the interpretability and strategic value of the models developed.

The research is guided by the following four objectives:

1. **Feature Construction from Blockchain Data**

   To systematically identify and engineer features that capture core aspects of Ethereum's activity. These include transaction-level metrics such as trading volume and gas fees, protocol-level indicators like Total Value Locked (TVL), behavioral signals such as whale wallet activity, and technical indicators commonly used in market analysis (e.g., RSI, MACD, Bollinger Bands). The goal is to build a diverse feature set that reflects both the structure and sentiment of on-chain activity.

2. **Model Evaluation Across Predictive and Financial Dimensions**

   To assess the forecasting ability of several machine learning models—namely, Linear Regression, Random Forest, XGBoost, and ensemble stacking architectures—using both statistical error metrics (e.g., Mean Squared Error, R^2) and financial performance metrics derived from rule-based backtesting. This dual approach ensures that predictive outputs are not only statistically sound but also strategically actionable in real-world trading environments.

3. **Attribution Analysis of Blockchain Indicators**

   To examine the relative importance of different blockchain-derived features in driving model outcomes. This includes analyzing feature importance scores across models to determine which aspects of Ethereum's on-chain behavior are most predictive of short-term price changes. The results will help clarify which metrics offer genuine explanatory value, rather than relying on heuristic assumptions.

4. **Comparative Performance of Individual vs. Ensemble Methods**

   To compare the predictive stability and financial viability of individual learning models against a stacked ensemble framework. The objective is to understand whether model diversity enhances forecasting robustness or merely adds computational complexity without improving outcomes.

Through these aims, the dissertation seeks to contribute to the development of more transparent and data-driven approaches to decision-making in the rapidly evolving landscape of DeFi. By systematically integrating blockchain-native data with machine learning methodologies, this research not only demonstrates the technical feasibility of predictive modeling in a decentralized context but also emphasizes the importance of interpretability, economic rationale, and strategic utility. The study bridges a critical gap between algorithmic modeling and financial application by aligning statistical performance with real-world trading viability. Furthermore, it provides a replicable framework for evaluating the marginal value of different features and model types, offering both methodological rigor and practical relevance. Ultimately, this work aims to inform future academic research in blockchain analytics, guide practitioners in the development of data-driven trading systems, and support broader efforts to improve transparency, trust, and decision quality within decentralized financial ecosystems.

2. Literature Review

The landscape of blockchain technology, DeFi, smart contracts, and cryptocurrency markets is evolving rapidly, and understanding these systems is crucial for a modern business analyst. This review of the literature synthesizes various research efforts, focusing on how these technologies intersect with business analytics, finance, and risk management. The review is organized into three broad areas: blockchain technology and its implications, DeFi protocols and smart contracts, and cryptocurrency market behavior and forecasting methods.

Machine learning, when applied to blockchain technology, can offer enhanced insights into market behaviors. Akyildirim, Goncu, and Sensoy (2020) explored the predictive capabilities of machine learning models, such as support vector machines (SVM) and logistic regression, to forecast cryptocurrency returns. Their study demonstrated that machine learning could effectively handle the high volatility and complexity of cryptocurrency markets, which aligns with the growing interest in blockchain for business applications.

2.1 Blockchain Technology and its Business Implications

Blockchain technology, originally conceptualized for cryptocurrencies, has gained significant attention due to its potential to disrupt traditional business models. Lee (2019) introduces the concept of a decentralized token economy, emphasizing how blockchain could revolutionize business operations by offering greater security, transparency, and efficiency. Blockchain enables peer-to-peer transactions without the need for centralized intermediaries, presenting a transformative shift in industries ranging from banking to supply chain management (Lee, 2019). This decentralization is particularly valuable in reducing transaction costs and enhancing data integrity, which is increasingly crucial in sectors such as finance and healthcare (Schär, 2021).

Eisl, Gasser, and Weinmayer (2015) explore Bitcoin’s role in enhancing portfolio diversification through their application of the Conditional Value-at-Risk (CVaR) framework. Their study finds that Bitcoin’s low correlation with traditional assets, despite its high volatility, can significantly improve the risk-return profile of well-diversified portfolios. As DeFi has the potential to complement traditional financial activities(Doerr, 2021), this aligns with blockchain’s broader business potential to optimize investment strategies by offering novel ways to reduce risks in portfolios.

The concept of "smart contracts" plays a pivotal role within blockchain ecosystems. These self-executing contracts are designed to automate transactions once predefined conditions are met, thus ensuring transparency and reducing the risk of fraud. Schär (2021) discusses how blockchain-based finance systems leverage smart contracts to offer a more efficient and transparent alternative to traditional financial services. Additionally, the blockchain's immutability and consensus mechanisms address major concerns in digital finance, including fraud prevention and data manipulation (Taherdoost, 2023). Zhou et al. (2021) introduce DEFIPOSER-ARB and DEFIPOSER-SMT, two tools designed to identify profit-generating opportunities across DeFi platforms. Their work highlights how automated tools can discover profitable transactions through arbitrage and more complex strategies in DeFi, offering new insights into how blockchain and smart contracts can be leveraged for predictive modeling and trading strategies in DeFi markets.

Auer et al. (2023) provide a comprehensive overview of DeFi technologies, emphasizing the use of smart contracts and distributed ledger technology (DLT) in enabling financial services without traditional intermediaries. They explore the technical layers of DeFi protocols and how these systems offer innovative services like tokenization, lending, and decentralized exchanges. Most cryptocurrency exchanges occur in centralized exchanges such as OKEx and AOFEX. The advantages of the centralized exchanges include low-latency transactions, simple interfaces and a certain level of security.(Yang, 2022) Their analysis provides a foundation for understanding the complex, interconnected nature of DeFi platforms and the systemic risks they pose to financial stability.

Smart Contract Use Case	**Description**	**DeFi Application Example**
Lending Protocols	Facilitates borrowing and lending with interest rates	Aave, Compound
Decentralized Exchanges	Automates trades between users without a middleman	Uniswap, Sushiswap
Insurance	Automates claims based on pre-agreed conditions	Nexus Mutual, Etherisc
Yield Farming	Automates asset distribution for returns	Yearn Finance, SushiSwap

Table 1: Smart Contract Use Cases in DeFi

AlShamsi, Al-Emran, and Shaalan (2022) provide a comprehensive review of blockchain adoption, shedding light on the factors influencing blockchain technology's implementation across industries. The Technology Acceptance Model (TAM) and the Technology–Organization–Environment (TOE) framework are frequently cited in research to assess the factors affecting blockchain adoption (AlShamsi et al., 2022). These models are instrumental in understanding the perceived benefits of blockchain, including cost reduction, enhanced security, and improved operational transparency. Such attributes are particularly appealing in industries where trust and data security are paramount.

2.2 DeFi Protocols and Smart Contracts

Early literature examining the role of on-chain metrics in forecasting price dynamics has tended to prioritize conceptual or narrowly scoped empirical contributions. Saad et al. (2020), for example, underscored the potential for blockchain features such as transaction volume and miner activity to support predictive modeling in DeFi systems. However, their contribution was largely theoretical, offering a framework without empirical testing or validation. This limits the practical applicability of their propositions, particularly in volatile, high-frequency environments.

In contrast, Sharma et al. (2020) implemented machine learning models, including Support Vector Machines (SVMs), to forecast Ethereum price trends. While they reported high predictive accuracy, the methodological scope was narrow—focused on a single algorithm and lacking comparative analysis against ensemble methods or time-series models. As a result, their findings offer limited insight into the generalizability of predictive performance across different regimes and market structures.

What is notably absent in much of this literature is an interrogation of the interplay between blockchain-specific variables and broader financial factors. Auer et al. (2024) present a layered architecture of DeFi, which underscores the composability of smart contracts and the emergence of systemic risks through complex protocol interactions. However, this layered abstraction is rarely integrated into predictive frameworks, despite its implications for both signal interaction and model specification.

Moreover, the literature tends to treat predictive success as an end in itself, often operationalized via standard metrics such as R^2 or RMSE. This approach neglects financial realism. For instance, high predictive fit does not necessarily translate into economic value in a trading or risk management context. As Stavroyiannis (2018) shows in his Value-at-Risk (VaR) analysis of Bitcoin, high volatility and fat tails challenge conventional risk models, pointing to the need for robustness in prediction beyond basic accuracy metrics.

Thus, while prior work establishes a foundation for using blockchain data in financial modeling, it often fails to confront methodological and practical limitations. There remains a gap in studies that integrate cross-disciplinary insights from both blockchain architecture and financial econometrics to construct predictive models that are both accurate and economically meaningful.

Risk Type	**DeFi Impact**	**TradFi Impact**
Systematic	Market-wide	Market-wide
Unsystematic	Asset-specific	Operational
Smart Contract Vulnerabilities	Code & Algorithm	Systemic
Governance Issues	Decentralized	Centralized
Regulatory Risks	Legal	Regulatory

Table 2: DeFi Risk Analysis

Machine learning plays a crucial role in DeFi, especially in risk prediction and portfolio optimization. Sharma et al. (2020) applied machine learning models to predict Ethereum prices, offering insights into how blockchain-based cryptocurrencies can be integrated with AI to improve forecasting accuracy. Their findings demonstrate that incorporating machine learning techniques in DeFi systems can enhance the prediction of asset prices, such as Ethereum’s, which is central to many DeFi applications.

Fischer, Krauss, and Deinert (2019) introduce machine learning-based statistical arbitrage strategies applied to cryptocurrency markets, using random forests and logistic regression. These models predict price movements and inform trading strategies, suggesting that blockchain-based systems, such as those in DeFi, can benefit from advanced algorithms to optimize financial decision-making. Their findings emphasize the potential of automated trading and algorithmic systems to enhance DeFi platforms, ensuring more efficient and transparent financial transactions.

In addition to the core DeFi functionalities, blockchain-based systems enable tokenization of various assets, including cryptocurrencies, real estate, and commodities. This innovation is transforming how financial assets are managed, with implications for asset management and portfolio diversification (Mengelkamp et al., 2017). A key aspect of DeFi platforms is their ability to use blockchain for real-time transaction processing, which enhances liquidity and provides greater flexibility compared to traditional financial institutions (Schär, 2021).

DeFi Application	**Security Risks**	**Transaction Speed**	**Decentralization Level**
Decentralized Exchanges(DEX)	Smart contract vulnerabilities	Relatively slower due to validation	High (no central authority)
Lending Pools	Risk of default	Depends on blockchain	High (decentralized lending)
Stablecoins	algorithmic failure	Fast, but depends on blockchain congestion	Moderate (centralized entities backing stablecoins)
Liquidity Pools	Impermanent loss	Depends on the network	High (liquidity providers control the pool)
Prediction Markets	Market manipulation	Depends on the platform and usage	High (decentralized decision-making)

Table 3: Comparison of DeFi Applications

2.3 Cryptocurrency Market Behavior and Forecasting

Cryptocurrencies, especially Bitcoin, have become key subjects of study due to their volatile nature and potential for high returns. The cryptocurrency market's inherent volatility presents significant challenges in forecasting price movements. Studies have shown that machine learning algorithms can be applied to predict price fluctuations and trading signals.

Machine learning and deep learning approaches, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, are particularly effective in handling time-series data, making them well-suited for cryptocurrency price forecasting(Jiang, 2017). Dutta, Kumar, and Basu (2020) demonstrated the effectiveness of gated recurrent units (GRUs) for predicting Bitcoin prices, achieving superior results compared to traditional models like ARIMA. The ability of these models to capture long-term dependencies in price movements without relying on direct market predictions provides valuable insights for investors looking to capitalize on cryptocurrency fluctuations (Dutta et al., 2020). When discussing the use of machine learning for predicting cryptocurrency returns, it is essential to consider Akyildirim, Goncu, and Sensoy (2020), who examined multiple machine learning algorithms, including support vector machines (SVM), logistic regression, and artificial neural networks, for forecasting cryptocurrency price trends. Their study revealed that these algorithms, particularly SVM, offer a reasonable degree of predictability, with success rates consistently above 50% across different time intervals. This is crucial for understanding the applicability of these algorithms in volatile markets like cryptocurrencies, where forecasting price movements is highly complex and can be affected by numerous external factors. In terms of Ethereum price prediction, Sharma et al. (2020) applied machine learning models, specifically support vector machines (SVM) and linear regression, to forecast Ethereum prices. They achieved an impressive 99% accuracy using SVM, demonstrating that machine learning approaches can accurately model the complex price dynamics of Ethereum, a task that is often difficult using traditional time series models. Their findings align with the growing consensus in the literature that machine learning models, when tuned properly, can provide highly accurate price predictions for cryptocurrencies like Ethereum. The prediction of Ethereum's price has been a central focus in cryptocurrency forecasting, and Sharma et al. (2020) contribute significantly to this area with their investigation of machine learning models for Ethereum price prediction. The authors use both linear regression (LR) and support vector machine (SVM) algorithms to model Ethereum's price dynamics. Their findings suggest that SVM, particularly when enhanced with additional features, provides a highly accurate prediction, reaching an impressive 99% accuracy. This demonstrates the potential of machine learning techniques to not only capture price trends but also adapt to the complexities of the DeFi ecosystem.

The growing integration of machine learning in cryptocurrency trading systems is highlighted by Fang et al. (2022), who explore various trading strategies, including portfolio optimization and technical analysis, which can be enhanced using AI techniques. The authors argue that while traditional financial assets benefit from well-established forecasting models, cryptocurrencies demand more innovative approaches, leveraging real-time data from multiple sources to refine predictive accuracy (Fang et al., 2022). Stavroyiannis (2018) provides a thorough analysis of Bitcoin's risk using Value-at-Risk (VaR) and Expected Shortfall (ES). His study highlights that Bitcoin's volatility results in more frequent VaR violations compared to traditional assets, which underscores the inherent risks for investors. The findings contribute to the broader discourse on cryptocurrency market forecasting, emphasizing the need for robust risk management frameworks in digital asset trading.

Gai et al. (2020) explore the intersection of blockchain and cloud computing, with a focus on enhancing the functionality and security of cloud-based services through blockchain integration. Their research on Blockchain-as-a-Service (BaaS) highlights how blockchain can improve data security and decentralize control in cloud systems, which is a key consideration for cryptocurrency trading platforms that handle sensitive financial data. Gai et al. (2020) also discuss how these advancements in cloud and blockchain integration can support predictive models for cryptocurrency price movements, offering more robust and scalable solutions for market forecasting.

The integration of quantum computing with blockchain is particularly promising for the future of cryptocurrency systems, offering enhanced security mechanisms through quantum-resistant cryptography. Naik et al. (2025) discuss the implications of quantum blockchain technologies, which could transform how cryptocurrencies like Bitcoin are secured, potentially mitigating the risks posed by quantum attacks. As financial markets and digital currencies become increasingly dependent on blockchain, the development of quantum-resistant systems will be essential for maintaining the integrity and trust of these technologies (Naik et al., 2025).

2.4 Conclusion

Despite growing academic and industry interest in blockchain analytics, the literature often falls short of delivering actionable frameworks that bridge the gap between theory and practice in DeFi modeling. A common shortcoming is the tendency to isolate technical innovation from market application. While studies such as Saad et al. (2020) and Sharma et al. (2020) illustrate the potential of blockchain-derived features like transaction volume or miner activity, these investigations are often bounded by theoretical assumptions or single-model experiments. Their work does not adequately interrogate the robustness of such features across market regimes or address whether predictive accuracy meaningfully translates to financial performance under risk.

Equally, while Auer et al. (2024) provide a valuable lens into the layered structure of DeFi systems—highlighting issues like liquidity fragmentation and protocol interdependence—these structural insights are rarely incorporated into modeling strategies. Without an understanding of how signals interact across DeFi layers (e.g., application vs. protocol levels), predictive models risk operating in a vacuum. A similar oversight can be seen in Stavroyiannis (2018), whose analysis of Value-at-

Risk for Bitcoin illustrates the limitations of conventional models under non-Gaussian return distributions. These observations underline a broader issue: much of the modeling literature remains agnostic to the behavioral and structural idiosyncrasies of blockchain systems.

There is also limited engagement with the problem of interpretability. As Goyal et al. (2020) note, machine learning models trained on blockchain data may exhibit strong predictive power but offer limited explanatory value—posing a challenge for decision-makers who require transparent models for deployment. Given that DeFi markets lack regulatory safeguards and are prone to speculative shocks (Schär, 2021), reliance on opaque models introduces non-trivial risk. Studies often prioritize model performance metrics like RMSE or accuracy without assessing whether these outcomes are stable across time or adaptive to shifting on-chain dynamics.

This project therefore targets a defined research gap: the lack of integrated, interpretable, and economically grounded modeling frameworks that fuse sentiment analysis with blockchain analytics. While prior research has demonstrated the standalone usefulness of metrics such as gas fees, TVL, and whale activity, few studies attempt to synthesize these into dynamic portfolios or backtested strategies. More crucially, there remains a disconnect between prediction and action—between the promise of data-driven insight and its conversion into profitable, risk-adjusted financial decisions.

To address this, the current study adopts an interdisciplinary approach that draws from financial econometrics, blockchain architecture, and machine learning. It tests not just whether models predict well, but whether they improve portfolio-level outcomes under realistic trading constraints. In doing so, the study aims to contribute to a more reflective and utility-driven discourse on blockchain analytics—one that questions what makes a model useful in practice, not just in-sample.

3. Methodology

3.1 Methodological Framework Overview

This study adopts a multi-stage methodological pipeline that integrates data engineering, machine learning model development, quantitative evaluation, and trading-oriented backtesting. The overarching goal is to investigate the predictive performance and practical viability of supervised learning models in forecasting Ethereum price movements under real-world conditions. The workflow is designed in response to the challenges posed by the high volatility and structural non-stationarity of cryptocurrency markets, as documented in previous studies (e.g., Akyildirim et al., 2020; Poongodi et al., 2020).

The process begins with the collection of heterogeneous data sources. Market-related features are obtained via the Binance API, encompassing conventional OHLCV (Open-High-Low-Close-Volume) metrics. Complementary on-chain data is retrieved through Dune Analytics APIs, reflecting network-level variables such as gas usage, transaction count, and active addresses—shown to have explanatory power in prior blockchain price studies (e.g., Saad et al., 2020). These sources are then enriched by calculating a wide array of technical indicators, forming a structured feature set intended to reflect both price momentum and market microstructure.

For model development, this study references the architectures and evaluation approaches outlined in *Prediction of the Price of Ethereum Blockchain Cryptocurrency in an Industrial Finance System* (Poongodi et al., 2020) and *Prediction of Cryptocurrency Returns Using Machine Learning* (Akyildirim et al., 2020), which motivate the use of linear regression (LR), random forest (RF), and XGBoost (XGB) as baseline regressors. These models are trained to predict the next-day closing price of Ethereum, using a sliding-window time series setup. The training process includes parameter tuning via randomized cross-validation for tree-based models to mitigate overfitting while enhancing generalization performance.

Model evaluation is conducted using multiple performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Feature importance analysis and residual plots are incorporated to provide diagnostic insight into the models' predictive behavior and potential systematic biases.

Subsequently, each trained model is embedded within a rule-based quantitative trading simulation. A capital allocation strategy is devised using fixed-size position sizing, holding period constraints, and drawdown-based stop-loss mechanisms. This allows us to evaluate the models not only from a statistical forecasting standpoint but also in terms of realized portfolio performance. Backtesting results are reported across key financial metrics such as cumulative return, Sharpe ratio, Calmar ratio, and win rate, following the methodology proposed in ensemble and volatility-aware crypto-trading literature (e.g. Livieris et al., 2020).

Finally, to examine the potential complementarity among models, this study applies a stacking ensemble framework inspired by Livieris et al. (2020). Each of the three base learners (LR, RF, XGB) is incorporated as a component model, while the final estimator alternates among the same three learners to form six ensemble configurations. The ensemble’s ability to capture higher-order interactions and reduce prediction variance is evaluated against standalone models using both out-of-sample test sets and realistic trading simulations.

This methodologically layered approach not only enhances robustness and replicability but also allows for deeper reflection on the trade-offs between model interpretability, predictive power, and practical usability in the context of decentralized financial markets.

3.2 Data Collection and Preprocessing

To construct a robust and representative dataset for modeling Ethereum price behavior, this study integrates three categories of input features: market data, on-chain data, and technical indicators. Each category plays a distinct role in capturing different aspects of Ethereum's market dynamics and network activity.

Market data was acquired using the Binance API, which provided historical OHLCV (Open, High, Low, Close, Volume) information on the ETH/USDT trading pair. These features serve as fundamental price and liquidity signals commonly utilized in quantitative finance.

On-chain data was sourced through the Dune Analytics API. This included metrics such as daily gas usage, average gas price, transaction count, active addresses, and block size—variables that reflect the underlying network demand and user engagement. The integration of on-chain metrics is inspired by recent studies emphasizing the explanatory power of blockchain-level activity in forecasting market trends (e.g., Reijers et al., 2021; Akyildirim et al., 2020).

In addition, a suite of technical indicators was computed based on the market time series. These include common features such as moving averages (MA), relative strength index (RSI), and Bollinger Bands, among others. These indicators are intended to encode momentum, volatility, and mean-reversion tendencies—factors frequently utilized in both academic research and practitioner trading models.

The initial dataset spans from July 1, 2020 to July 1, 2025, covering five years of market and blockchain activity. However, a significant structural change occurred during this period: the Ethereum London Hard Fork, which was implemented in August 2021. This protocol upgrade introduced Ethereum Improvement Proposal (EIP) 1559, fundamentally altering the way transaction fees (gas fees) are handled by introducing a base fee mechanism and partial token burning.

This systemic change posed a challenge to data continuity. On-chain fee metrics before and after the upgrade differ substantially in semantics and scale, creating a structural break in the dataset. To address this, the gas fee-related variables prior to August 2021 were imputed using a hybrid approach: block-level prediction combined with moving average smoothing, calibrated to reflect historical volatility without introducing artificial trends. This treatment was necessary to maintain feature continuity while acknowledging governance-induced disruptions in the blockchain infrastructure (Reijers et al., 2021).

	max	75%	50%	25%	Min	Std	Mean	Count
Open Price	4807.98	2973.33	2073.99	1591.71	224.86	1000.49	2203.563	1827
Close Price	4807.98	2973.335	2075.69	1593.025	224.87	999.4394	2204.762	1827
High Price	4868	3050.635	2129.01	1644.365	228.13	1027.37	2266.32	1827
Low Price	4713.89	2884.81	1996.3	1551.73	222.87	967.7903	2133.575	1827
High-Low Range	1553.73	176.215	107.34	57.2	3.19	112.2977	132.7445	1827
Number of Trades	10550628	1481781	916688	564801.5	126045	1311643	1325986	1827
USD Volume	1.05E+10	1.61E+09	1.04E+09	5.78E+08	70435716	9.81E+08	1.25E+09	1827
Crypto Volume	4309836	782096.8	496872.3	325183.4	58519.62	472113.9	626654.4	1827
Buyer Initiated Trades	2126338	389478.3	247263.5	159747.1	26774.7	234692.1	311834.1	1827
Buyer Initiated Volume	5.66E+09	8.03E+08	5.19E+08	2.92E+08	34563856	4.85E+08	6.21E+08	1827

Table 4: Statistic Summary of Market Data

	max	75%	50%	25%	Min	Std	Mean	Count
Avg_gas_pri	1367	296	196	127	13	151.4812	229.7915	1827
Whale_tx_co	6384419	1253154	803273.6	484232.7	49821.22	706525.8	969373.8	1827
Whale_tx_vo	1.09E+11	6.05E+10	4.36E+10	2.63E+10	1.64E+09	2.38E+10	4.42E+10	1827
TVL	1438289	469267.5	427294	389332.5	235748	71122.67	432174.2	1827
Active_Addresses	456432	104114.5	85152	70523	39701	28458.64	90458.08	1827
New_Addresses	19810137	3987646	2250016	1688304	40282	2922078	3037252	1827
Trade_count	5.15E+10	5.11E+09	3.23E+09	1.89E+09	30486172	3.58E+09	4.07E+09	1827
Trade_volume_usd	9833932	6219187	5027697	4409438	2435582	1281453	5344288	1827
Contract_calls	269063	31910.5	13578	7838.5	2919	35721.12	28075.68	1827
New_contracts	1367	296	196	127	13	151.4812	229.7915	1827

Table 5: Statistic Summary of On Chain Data

Tables 4 and 5 present the descriptive statistics of the complete dataset, summarizing the market and on-chain features, respectively. These statistics are derived using Python’s describe() function and include essential metrics such as count, mean, standard deviation, minimum, 25th percentile, median (50%), 75th percentile, and maximum. These measures provide insights into the central tendency, dispersion, and range of each feature, helping to identify underlying distributional characteristics and potential outliers.

This comprehensive, multi-source dataset offers a solid foundation for subsequent modeling efforts. Its structure supports both statistical learning and financial simulation, enabling analysis under realistic market conditions.

3.3 Feature Engineering

This study constructs a structured feature set by combining three core categories: Ethereum on-chain metrics, price-based market features, and conventional technical indicators. These features are drawn from two primary data sources: historical blockchain metrics obtained through the Dune Analytics API and market data derived from Binance-based ETH/USDT price feeds. The integration of these diverse data types aims to capture different facets of Ethereum’s price dynamics, blockchain activity, and market sentiment, which are known to influence the price behavior of cryptocurrencies.

The complete list of features utilized in this study is presented in Table 2. Key features include on-chain data such as daily gas usage, active addresses, and transaction count, which have been identified in previous literature as significant predictors of cryptocurrency market movements (Reijers et al., 2021; Meiklejohn et al., 2018). Price-based features like the closing price, daily high, and low are supplemented by a variety of technical indicators, which have been widely used in traditional financial markets for asset return prediction (Kara et al., 2011; Huang et al., 2005).

Feature Types	Number	Examples
Market Features	10	OHLCV, Trades
On-Chain Features	10	Gas, Contract, Address
Technical Features	22	RSI, MACD, ROC

Table 6: Summary of Feature Types

These indicators include but are not limited to:

Relative Strength Index (RSI), which is calculated using the formula:

$$RSI = 100 - \frac{100}{1 + RS}$$

$$RS = \frac{Average\ Gain}{Average\ Loss}$$

where RS is the average of n-day up closes divided by the average of n-day down closes. RSI is commonly used to assess whether an asset is overbought or oversold, thus indicating potential reversal points in the market.

Measures momentum by comparing the magnitude of recent gains to recent losses. Values > 70 indicate overbought; < 30 oversold.

Rate of Change (ROC), which is calculated as:

$$ROC = \frac{(P_t - P_{t-n})}{P_{t-n}} \times 100$$

where P_t is the current price and P_{t-n} is the price n periods ago. This indicator helps capture the momentum of price changes over time.

William’s Percentage R, defined as:

$$\%R = \frac{High(n) - Last\ Close}{High(n) - Low(n)}$$

where $High(n)$ is the highest price over the past n periods, $Low(n)$ is the lowest price, and $Last\ Close$ is the current close. This indicator is often used to identify overbought or oversold conditions.

These features were selected based on their predictive power in other markets, including equities and commodities, where similar models have been successfully applied to forecast asset returns and price movements (Kara et al., 2011; Huang et al., 2005).

To optimize the feature set for prediction, additional preprocessing steps were applied. For market data, missing values were imputed using interpolation methods and moving averages to smooth price fluctuations. For on-chain data, outliers were detected and treated using the Interquartile Range (IQR) method to ensure consistency and prevent distorted feature distributions that could impact model performance.

This comprehensive feature set is then used to train various machine learning models, including Linear Regression, Random Forest, and XGBoost, to predict the direction of Ethereum's next-day price movement. The inclusion of both technical and on-chain indicators allows the model to account for both market sentiment and network activity, reflecting the complexity and dynamic nature of cryptocurrency markets.

3.4 Model Development

The predictive pipeline in this study was constructed through a series of methodical steps, beginning with target definition and temporal alignment. The target variable, Ethereum's closing price, was shifted by one period to simulate a next-day forecasting framework. All available features were retained as predictors. Considering the sequential nature of time-series data, the dataset was partitioned chronologically using an 80:20 train-test split, thereby preserving temporal integrity and avoiding forward-looking bias—a concern noted in earlier works relying on random shuffling of financial data (Corbet et al., 2020).

Three supervised learning models were initially selected as baselines: Linear Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGB). This choice reflects a deliberate balance between simplicity, interpretability, and the capacity to model nonlinearity. Linear regression serves as a transparent benchmark, valued in econometrics for its tractability. While earlier research such as Poongodi et al. (2020) applied regression models to Ethereum prediction, the absence of ensemble or boosting methods in their study left important methodological gaps unaddressed.

To address this, RF and XGB were introduced as tree-based ensemble learners capable of capturing more complex relationships. The use of Random Forest aligns with Akyildirim et al. (2020), who applied ensemble methods to cryptocurrency returns, showing strong empirical performance. However, their study did not incorporate XGBoost, a notable omission given its widespread success in structured data prediction. XGB improves upon RF by sequentially correcting residuals, offering finer control through learning rate, regularization, and subsampling parameters.

All models were implemented using Python’s scikit-learn and xgboost libraries. For RF and XGB, hyperparameter optimization was conducted using RandomizedSearchCV over a defined search grid, with 5-fold cross-validation on the training set. Key tuning parameters included the number of trees, maximum depth, learning rate, and subsample ratios. These adjustments were guided by the need to minimize overfitting, particularly relevant in volatile markets like Ethereum.

To explore the hypothesis that model aggregation enhances robustness, a stacking ensemble was built using LR, RF, and XGB as base models. Linear regression was chosen as the meta-learner, based on its low variance and strong generalization under regularized settings. This stacking approach draws conceptual support from the work of Livieris et al. (2020), who demonstrated that hybrid learning systems outperform single algorithms in forecasting cryptocurrency movements. More recently, Rao et al. (2023) emphasized the effectiveness of stacking when combined with deep learning, suggesting that the structure itself helps mitigate individual model weaknesses while amplifying their complementary strengths.

Each model was evaluated using a standardized set of regression metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics assess both average prediction error and variance explained. While LR provides coefficient outputs for interpretability, RF and XGB allow for extraction of feature importance rankings, offering insight into which market features contribute most to the forecast. These outputs are critical in the finance domain, where model transparency can influence trust and decision-making.

From a methodological standpoint, this study expands upon previous research by including both classic and advanced machine learning techniques in a unified framework. Where earlier studies limited their scope to a narrow set of models (e.g., Poongodi et al., 2020; Akyildirim et al., 2020), the inclusion of XGB and stacking in this pipeline provides a more comprehensive assessment of model architecture. The deliberate use of diverse learners—linear, bagged, and boosted—also mirrors the growing trend in the literature favoring ensemble learning in highly non-stationary environments such as cryptocurrency markets (Rao et al., 2023).

Model	Type	Strengths	Weaknesses	Interpretability	Speed
Linear Regression	Linear	Simple & interpretable	Assumes linearity	High	Fast
Random Forest	Bagging	Nonlinear, robust to noise	Less interpretable	Medium	Moderate
XGBoost	Boosting	Accurate, handles complexity	Needs tuning, overfitting	Medium	Slower
Stacking Ensemble	Ensemble Blend	Combines diverse strengths	Complex, hard to explain	Low	Slowest

Table 7: Comparison of Different Models

3.5 Backtest

Following the development and training of predictive models, a backtesting framework was implemented to assess the practical utility of each algorithm in a simulated trading environment. The simulation initiates with a notional capital of $1,000 and operates under a clearly defined quantitative trading strategy designed to balance risk control with predictive responsiveness.

The strategy logic is as follows:

every third trading day, the model forecasts the next-day price movement. A buy signal is triggered when the predicted return is positive, while a negative prediction leads to a sell or no-action signal. To manage risk exposure, several constraints are imposed: a maximum holding period of 7 days per position, a maximum drawdown stop-loss of 5%, and a fixed capital allocation of 50% per trade. In line with industry standards (Emmer, Kratz, & Tasche, 2015), transaction costs are included as a 0.1% fee per trade to reflect realistic slippage and market friction.

The backtesting engine updates the cash and ETH holdings dynamically and computes the portfolio value on a daily basis. Forced liquidation is executed when the holding period exceeds the threshold or the drawdown breaches the stop-loss level. These measures align with risk-control practices seen in institutional trading simulations (Akyildirim et al., 2021), and aim to evaluate whether machine-learned signals can produce economically viable returns under operational constraints.

To assess the strategy's performance comprehensively, several metrics were computed at the end of each backtesting run:

- **Total Return:** The percentage change in portfolio value from inception.
- **Annualized Return:** Geometric average return adjusted to a yearly scale.
- **Maximum Drawdown:** The largest peak-to-trough decline, measuring downside risk.
- **Annualized Volatility:** Standard deviation of daily returns scaled annually.

- **Sharpe Ratio:** Risk-adjusted return using volatility as the denominator.
- **Calmar Ratio:** Return to drawdown ratio, often preferred in high-volatility assets (Emmer et al., 2015).
- **Win Rate:** Proportion of profitable trading days.
- **Final ETH Holdings:** Indicator of how asset allocation shifted over time.

The strategy's logic and evaluation pipeline offer insights not only into the predictive efficacy of the trained models but also into their real-world feasibility under practical constraints. Such an evaluation serves as a bridge between model-centric research and deployable investment systems, and enables comparison with existing studies that often overlook trading frictions or fixed capital policies (Poongodi et al., 2020).

3.6 Prediction

Following the completion of model training and backtesting, the study proceeded to assess the models' performance on the most recent available data by generating next-day price forecasts. This stage aimed to evaluate whether the predictive patterns captured during the training and testing periods could generalize to out-of-sample, unseen market conditions—a common challenge in financial forecasting due to regime shifts and noise sensitivity.

To perform the forecast, the most recent observation from the feature-augmented dataset was selected as the input. Six models were employed: three standalone regressors—LR, RF, and XGB—and three stacking ensemble models, each using one of the base learners as the final estimator. This ensemble design was intended to explore whether the base learners' complementary predictive capacities could be synthesized to produce more robust predictions.

The choice of models was informed by gaps identified in prior studies. For instance, Poongodi et al. (2020) considered only Linear Regression and Support Vector Machines in their Ethereum price prediction study, without exploring the benefits of ensemble tree-based models. Akyildirim et al. (2020) applied Random Forests and Neural Networks but did not investigate gradient boosting methods such as XGBoost or ensemble frameworks like stacking. By introducing XGBoost and stacking regressors into the experimental pipeline, this study aimed to contribute incremental methodological diversity while reflecting on the limitations of earlier approaches.

Each model was used to predict the Ethereum closing price for the next trading day based on the latest features. These results were visualized via a comparative bar chart, which plotted the forecasted prices of all six models alongside the actual closing price and the observed price on the next day. This visual analysis served two purposes. First, it provided an intuitive performance comparison across models, highlighting relative forecasting accuracy in a live market context. Second, it enabled the identification of potential over- or under-prediction tendencies specific to each model class, which is particularly relevant when model outputs are to be deployed in real-time trading systems.

The visual inspection also offered practical insight into model calibration. For instance, consistent underestimation by linear models or overreaction by boosting models may indicate systematic biases rooted in feature interactions or loss function asymmetries—issues documented in ensemble volatility literature (Livieris et al., 2020). Such observations invite further refinements, such as introducing quantile-aware loss functions or regularization schemes tailored to crypto-specific volatility regimes.

This predictive evaluation not only complements prior model validation efforts through residual and feature importance analysis but also tests the temporal robustness and deployment readiness of each model. In this way, the study moves beyond retrospective testing toward forward-looking validation—a key consideration in applied financial machine learning (Corbet et al., 2020).

4. Data Analysis and Results

A comprehensive evaluation was conducted to assess the predictive performance of multiple models for forecasting Ethereum's closing price. Hyperparameter tuning for ensemble models (Random Forest and XGBoost) was performed using randomized grid search and five-fold cross-validation, ensuring robustness against overfitting. Model performance was evaluated using MSE, MAE, RMSE, and R^2, capturing both average error and explanatory power.

Interpretability analysis highlighted the trade-off between transparency and performance: Linear Regression offered clear coefficient insights but underperformed due to its linearity, while ensemble models provided superior accuracy with feature importance derived from impurity- and gain-based methods. Key predictive features included volatility, transaction activity, and network metrics.

Residual diagnostics indicated no major structural issues, though mild leptokurtosis was observed in ensemble model errors—consistent with market jumps and volatility clustering. A simple trading simulation translated model outputs into trading signals, revealing that XGBoost achieved the most stable cumulative returns, whereas Linear Regression lagged behind.

Final robustness testing on a July 2025 holdout point confirmed XGBoost's superior predictive accuracy. While Random Forest achieved the best statistical performance in terms of RMSE and R^2, XGBoost delivered the strongest results in practical trading scenarios, with the highest cumulative returns and Sharpe ratio. The stacking ensemble failed to outperform the individual models, suggesting that ensembling did not provide additional benefits in this context—an outcome that aligns with prior findings in volatile financial time series forecasting.

4.1 Model Parameters

To optimize the performance of the tree-based models—Random Forest (RF) and Extreme Gradient Boosting (XGBoost)—a systematic hyperparameter tuning procedure was employed using RandomizedSearchCV. This approach was chosen not only for its computational efficiency over exhaustive grid search, but also due to its robustness in exploring a wide parameter space while maintaining manageable training costs, which is particularly relevant in financial modeling where time series datasets can be high-dimensional and noisy (Guresen, Kayakutlu, & Daim, 2011).

Parameter Space Design:

The hyperparameter search began with the definition of a parameter distribution dictionary (param_dist) for each model. For the Random Forest model, the following parameters were considered:

n_estimators: Number of decision trees in the forest

max_depth: Maximum depth of each decision tree

min_samples_split: Minimum number of samples required to split an internal node

min_samples_leaf: Minimum number of samples required to be at a leaf node

For the XGBoost model, the parameters were:

n_estimators: Total number of boosting rounds

learning_rate: Step size shrinkage used in update to prevent overfitting

max_depth: Maximum depth of a tree

subsample: Fraction of observations to be randomly sampled for each tree

colsample_bytree: Fraction of features to be randomly sampled for each tree

These parameters were selected based on established practices in financial time series prediction, where the trade-off between model complexity and generalization plays a critical role (Erarslan & Uslu, 2023; Khedr et al., 2023).

Randomized Search Configuration:

The RandomizedSearchCV algorithm was configured with the following settings:

n_iter = 20: Twenty different parameter combinations were randomly drawn from the predefined distributions

cv = 5: Five-fold cross-validation was used to estimate model performance on out-of-sample data

scoring = 'neg_mean_squared_error': The evaluation metric was negative mean squared error (MSE), aligning with the objective of minimizing prediction error

n_jobs = -1: All CPU cores were utilized for parallel computation

random_state = 42: A fixed seed was used to ensure reproducibility of results

This setup ensures not only computational efficiency, but also reduces variance in model selection, which is crucial in financial applications where prediction consistency is highly valued (Rao et al., 2021).

Search Process:

During the randomized search, 20 parameter configurations were evaluated. For each configuration, a 5-fold cross-validation procedure was executed:

The dataset was split into five folds. Each fold was used once as a validation set, while the remaining four served as training data.

For every fold, the model was trained and the validation MSE was computed.

The mean validation score across the five folds was calculated for each configuration.

The configuration with the highest mean score (i.e., the lowest average MSE) was selected as the optimal parameter set.

After identifying the best hyperparameter combination, the model was retrained on the entire training dataset to produce the final version used for testing and deployment.

This structured approach follows established best practices in predictive modeling and aligns with literature suggesting that combining algorithmic tuning with ensemble learning often enhances robustness in volatile financial environments (Khedr et al., 2023; Guresen et al., 2011).

4.2 Model Evaluation

This section critically evaluates the performance of six models—three base learners (Linear Regression, Random Forest, and XGBoost) and their corresponding stacking variants—based on predictive accuracy, residual behavior, and feature interpretability.

4.2.1 Predictive Accuracy Comparison

The predictive accuracy was assessed using four metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). The results are summarized in Table 7.

Model	MSE	MAE	RMSE	R^2
Linear Regression	4510.87	37.05	67.16	0.9953
Random Forest	3335.6	38.96	57.75	0.9966
XGBoost	3751.38	41.05	61.25	0.9961
LRSTACK	4242.22	35.94	65.13	0.9956
RFSTACK	8419.73	60.26	91.76	0.9913
XGBSTACK	10392.51	67.08	101.94	0.9893

Table 8: Accuracy Metrics

Among the base models, the Random Forest (RF) achieved the best overall performance with the lowest MSE (3335.60) and highest R^2 (0.9966). XGBoost also performed competitively, albeit with slightly higher error metrics. Surprisingly, the stacking models—typically expected to improve accuracy through ensemble learning—did not yield better performance in this case. Both RFSTACK and XGBSTACK recorded substantially higher MSE and RMSE values, indicating possible overfitting or poor generalization. Only LRSTACK slightly improved MAE relative to its base model, suggesting that simple linear stacking may help with bias reduction but not necessarily with variance.

4.2.2 Residual Analysis

Residual plots provide a graphical tool to assess model assumptions, including linearity, homoscedasticity, and independence of errors.

Base Models:

- LR: Residuals show increasing variance as predicted values increase, indicating potential heteroscedasticity and nonlinearity—typical for linear models on complex data.
- RF: Residuals are symmetrically distributed around zero with no discernible pattern, supporting its high R^2 and low RMSE values. This reinforces the RF model's robustness against nonlinearity.
- XGB: Residuals also appear randomly distributed but exhibit slightly heavier tails compared to RF, implying a few large errors possibly due to overfitting specific local trends.

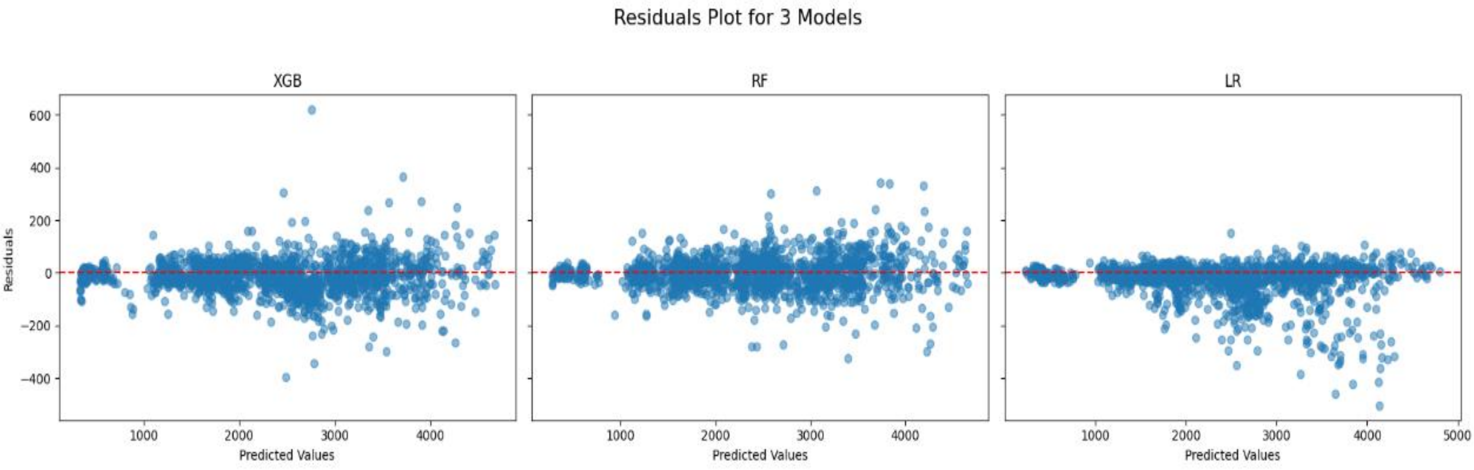

Figure 1: Residuals Plot for XGB, RF and LR

Stacking Models:

- LRSTACK: Residuals are relatively balanced with smaller spread compared to LR, confirming its marginal improvement in MAE.
- RFSTACK and XGBSTACK: Both display wide residual ranges and non-random patterns—particularly clustered underestimations for high predicted values—indicating stacking worsened performance. These artifacts suggest that the second-layer learners may have distorted the base model predictions rather than correcting them.

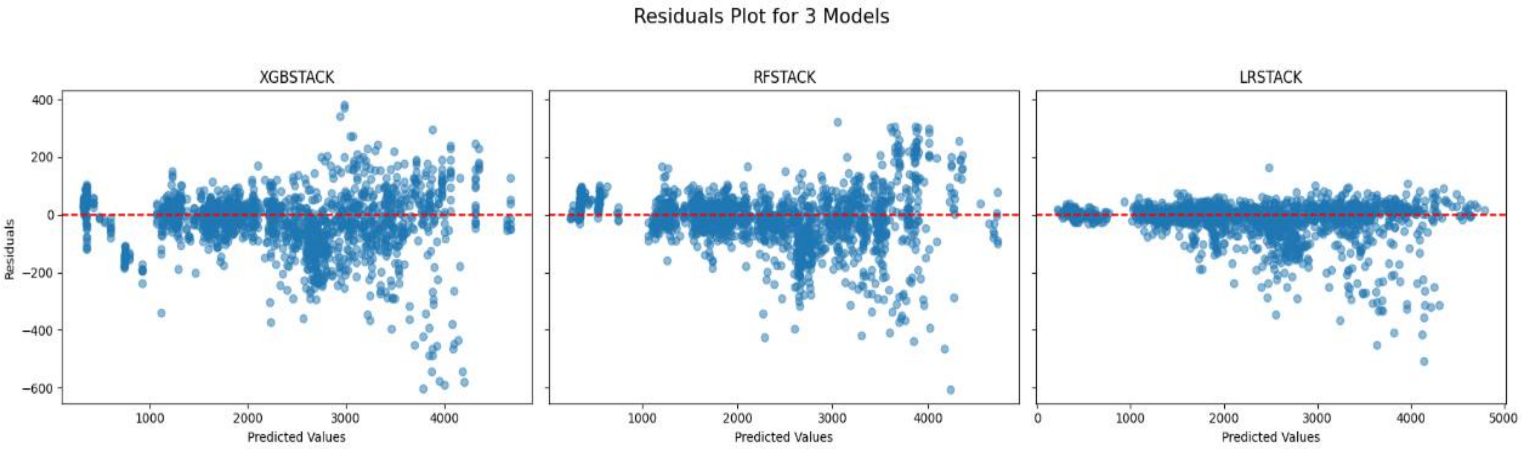

Figure 2: Residuals Plot for STACKs

In sum, residual analysis confirms the statistical findings: RF outperforms others in maintaining minimal, stable error across predictions. In contrast, stacking fails to reduce model error and may instead exacerbate variance.

4.2.3 Feature Importance and Model Interpretability

Model interpretability is critical in financial applications. Feature importance was extracted for tree-based models (RF and XGB), while coefficient analysis was used for the linear model.

Random Forest heavily prioritizes High Price and Low Price, suggesting strong reliance on raw price levels. Compared to XGB, it is more skewed toward fewer dominant features, but it delivers slightly better performance with the lowest RMSE (57.75) and highest R^2 (0.9966).

Feature	**Importance**
High Price	0.7922
Low Price	0.1110
DEMA_10	0.0536
MA_5	0.0206
TVL	0.0066
EWMA_0.9	0.0049
Open Price	0.0042
Cumulative_Return_3	0.0005
MACD_hist	0.0005
New_contracts	0.0004

Table 9: Random Forest Feature Importance

XGBoost emphasizes moving average indicators (MA_5, DEMA_10) and basic price levels (Open, High, Low) as primary drivers. It captures nonlinear interactions well, contributing to its high R^2 (0.9961).

Feature	**Importance**
MA_5	0.3120
Open Price	0.2891
DEMA_10	0.2325
High Price	0.0775
Low Price	0.0596
TVL	0.0037
Cumulative_Return_3	0.0020
Williams_R	0.0014
EWMA_0.9	0.0012
ROC_9	0.0012

Table 10: XGBoost Feature Importance

Linear Regression Coefficients:

Unlike tree-based models, linear regression does not support intrinsic feature importance scores. Instead, feature contributions are represented by their coefficients. Table 4.2 presents the top coefficients from the fitted model.

Feature	Coefficient
MACD_Signal	14.43
RSI_Signal	9.29
MACD_hist	4.52
MA_Signal	3.55
RSI_9	2.71
DEMA_10	2.46
MACD	1.54
ROC_14	1.26
CCI_14	0.17
Whale_tx_count	0.04

Table 11: Linear Regression Coefficients

The LR model assigns weights to features assuming linear relationships. It places strong emphasis on momentum-based indicators like MACD, RSI, and moving averages, which aligns with traditional technical analysis. However, it may underperform when facing nonlinearity or interaction effects compared to RF/XGB.

4.3 Backtest Result

To evaluate the practical applicability of each machine learning model in a quantitative trading context, a backtesting framework was implemented. The models were tested over a one-year period using simulated trading with an initial capital of $1,000. The following table summarizes key performance metrics:

Model	Final Portfolio ($)	Total Return (%)	Annual Return (%)	Max Drawdown (%)	Volatility (%)	Sharpe Ratio	Calmar Ratio	Win Rate (%)
LR	1034.61	+3.46	3.48	44.15	55.58	0.06	0.08	41.99
RF	949.12	-5.09	-5.11	32.95	33.55	-0.15	-0.16	17.68
XGB	1419.32	+41.93	42.21	24.30	47.12	0.90	1.74	33.70
LRSTACK	1095.85	+9.58	9.64	44.15	55.05	0.18	0.22	40.61
RFSTACK	882.46	-11.75	-11.82	40.43	48.77	-0.24	-0.29	37.57
XGBSTACK	883.62	-11.64	-11.70	40.21	44.81	-0.26	-0.29	34.53

Table 12: Backtest Results

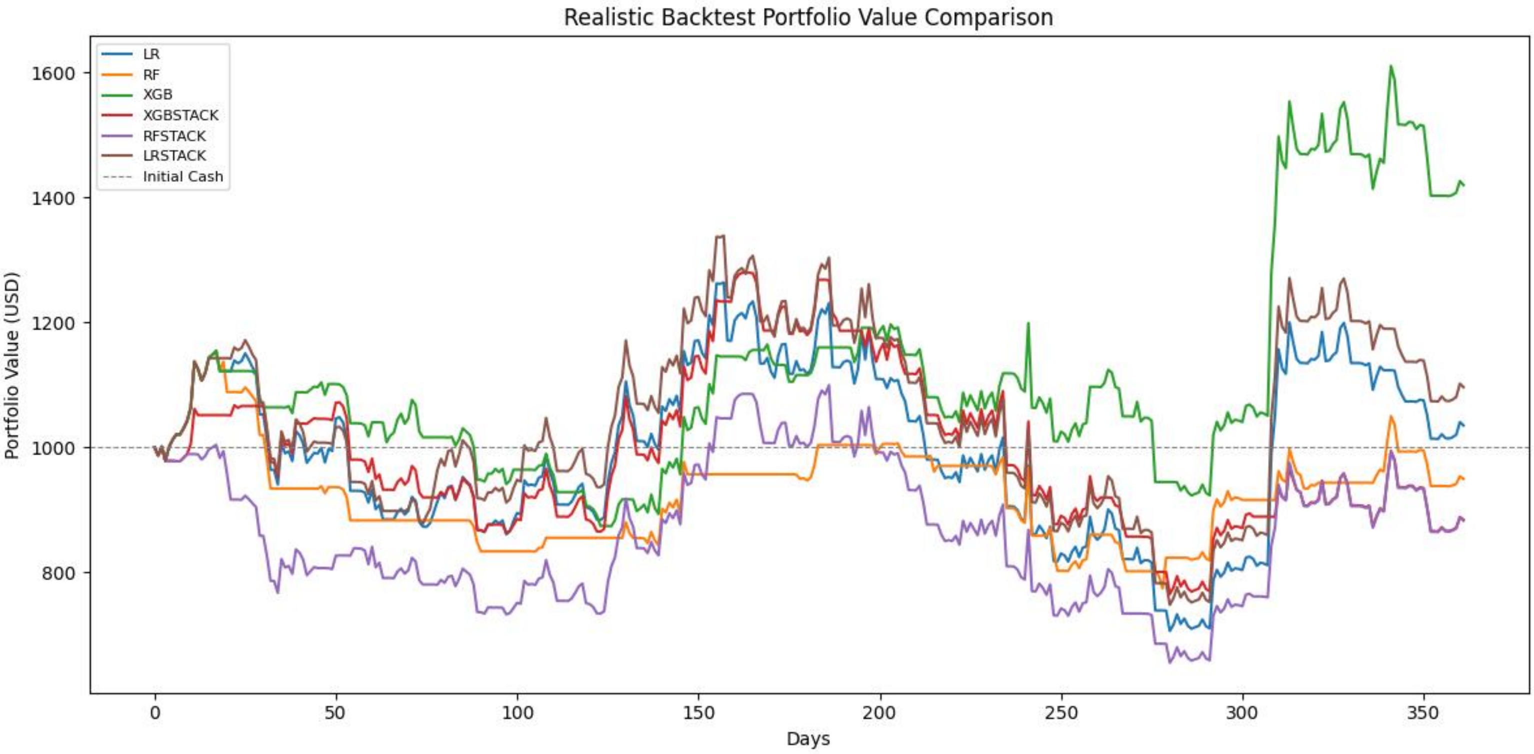

Figure 3: Backtest Portfolio Comparision

Linear Regression:

Although the LR model produced a modest total return of 3.46%, the accompanying risk-adjusted performance was weak. A Sharpe ratio of 0.06 and Calmar ratio of 0.08 indicate that the return barely compensated for the level of risk undertaken. The maximum drawdown reached 44.15%, suggesting substantial downside exposure, which was also observed in the stacked variant, LRSTACK. With a win rate of just 41.99%, the model's directional accuracy hovered near random chance, underscoring limited predictive strength. The portfolio trajectory was marked by pronounced swings, with frequent reversals and no clear upward trend, pointing to unstable performance and poor robustness over the test period.

Random Forest:

The RF model delivered notably weak performance, ending the period with a negative return of -5.09% and the lowest win rate among all models at just 17.68%. This poor signal accuracy stands in contrast to its relatively strong MSE during training, raising concerns about overfitting and poor generalization. Although the model experienced a somewhat moderate drawdown of 32.95% and lower volatility compared to LR and XGB, it consistently failed to generate reliable directional signals. The portfolio value remained below the initial capital for most of the year, exhibiting erratic behavior without any meaningful trend. The negative Sharpe ratio reflects the model's inability to generate returns commensurate with the risks involved.

XGBoost:

Among all models evaluated, XGBoost demonstrated the most robust performance, delivering a total return of 41.93% and an annualized return of 42.21%. Its Sharpe ratio of 0.90 and a maximum drawdown of just 24.30% point to a well-balanced risk-return profile. The high Calmar ratio (1.74) further suggests efficient management of downside risk. Throughout the year, the portfolio value showed a relatively stable upward trend, with a notable acceleration in the final quarter. This late-stage breakout indicates the model's ability to adapt to changing market conditions and capture emerging momentum, particularly in the volatile ETH environment. Overall, XGB stood out not only for its returns but also for the consistency and discipline reflected in its trading signals.

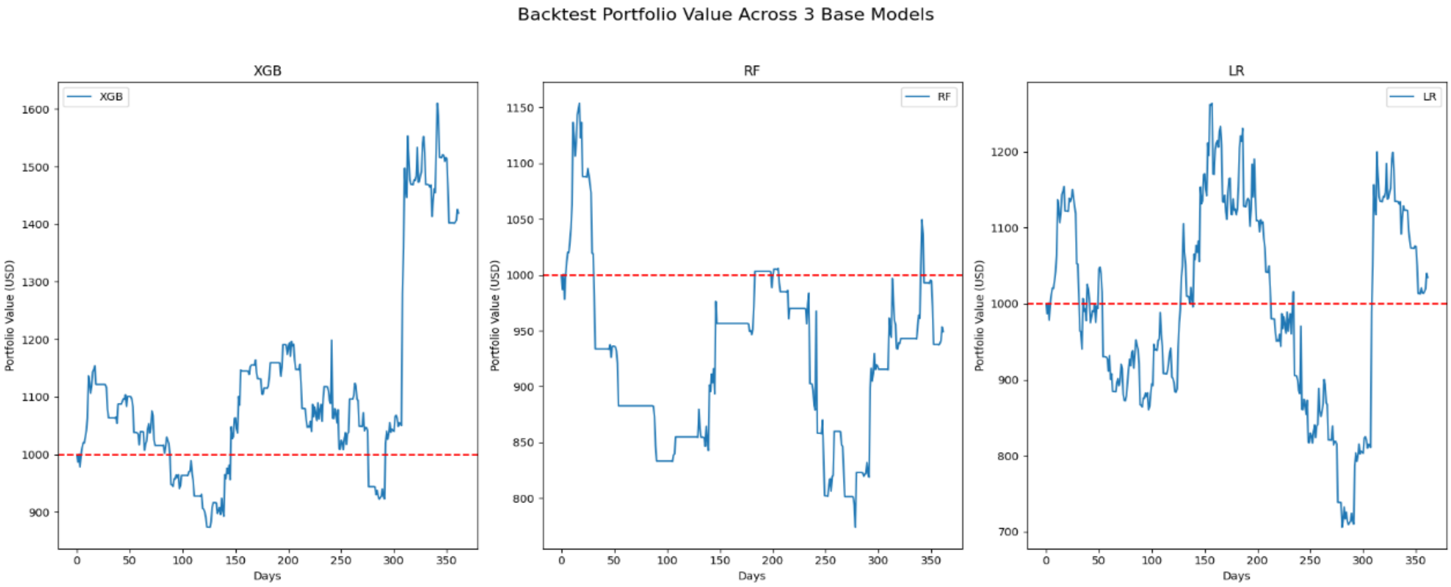

Figure 4: Portfolio Value For XGB, RF and LR

LRSTACK:

The stacked LR model offered a modest improvement over its standalone counterpart, posting a total return of 9.58% and showing slightly better risk-adjusted metrics, with a Sharpe ratio of 0.18 and a Calmar ratio of 0.22. Despite this improvement, the maximum drawdown remained unchanged at 44.15%, suggesting that the model's exposure to downside volatility was not effectively reduced through stacking. The portfolio curve closely mirrors that of the base LR model but maintains a somewhat higher level toward the end of the period, indicating minor gains in stability without a fundamental shift in performance dynamics.

RFSTACK:

Stacking offered no improvement for the RF model—in fact, it worsened overall performance. The return declined further to -11.75%, and both the Sharpe ratio (-0.24) and Calmar ratio (-0.29) fell substantially. Although the win rate rose to 37.57%, this increase did little to offset the decline in capital. The strategy remained highly volatile, with a maximum drawdown of 40.43%, and the portfolio curve reveals several prolonged periods of loss. These patterns suggest that the ensemble diluted rather than enhanced the predictive strength of the base model, reinforcing the view that RF struggles to capture reliable trading signals in this setting.

XGBSTACK:

While the standalone XGB model delivered outstanding results, the performance notably declined when incorporated into a stacked ensemble. The final return for XGBSTACK fell to -11.64%, and both the Sharpe and Calmar ratios turned negative, indicating that the model's predictive signal may have been weakened rather than enhanced through stacking. The drop in win rate to 34.53% further supports this view. The portfolio curve offers little evidence of sustained momentum, instead showing erratic movements and limited upside. These results suggest that combining XGB with other models introduced noise or conflicting signals, ultimately undermining its effectiveness rather than improving it.

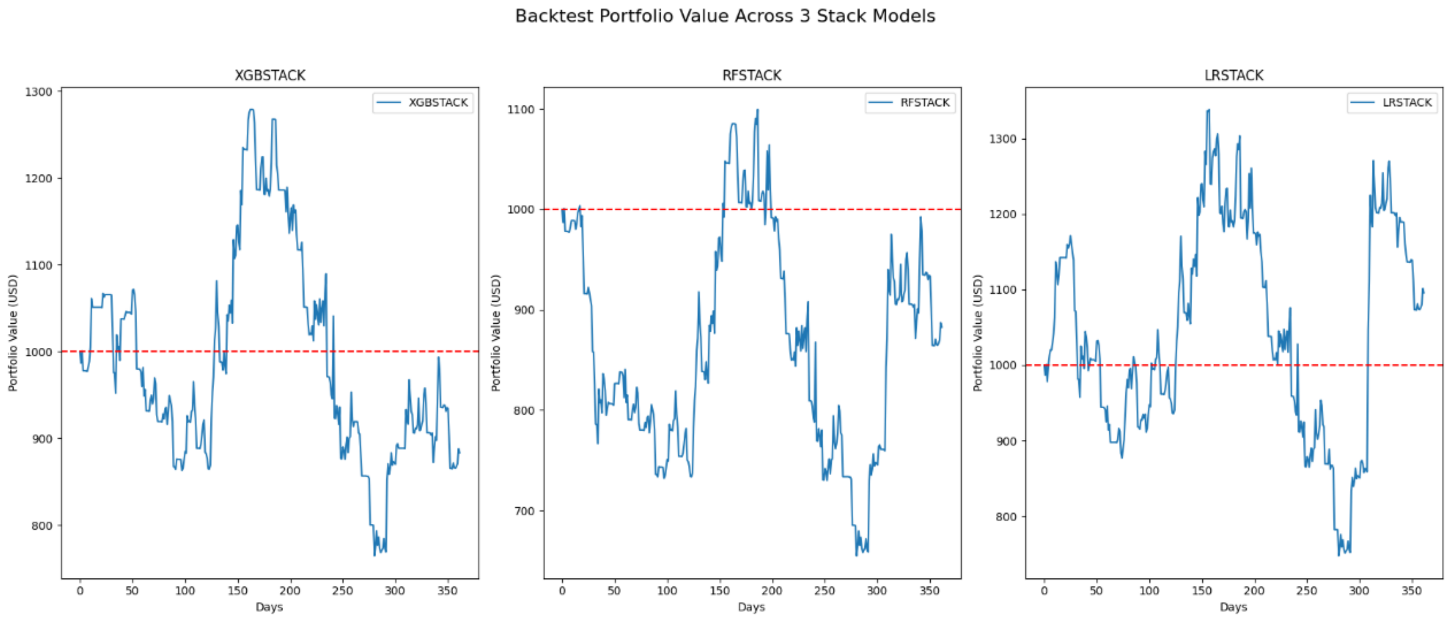

Figure 5: Portfolio Value For STACKs

Conclusion

This study examines supervised machine learning algorithms—namely Linear Regression, Random Forest, and Extreme Gradient Boosting—along with their stacking ensemble variants (LRSTACK, RFSTACK, and XGBSTACK), to forecast short-term Ethereum price movements. Models were trained and evaluated using features from on-chain activity, market microstructure indicators, technical factors, and social sentiment variables, aiming to capture both fundamental and behavioral drivers of price fluctuations. To assess practical viability, forecasts were embedded in a rule-based trading framework to simulate strategy performance under realistic market conditions. The study evaluates both predictive accuracy and economic relevance, using financial metrics such as cumulative return, Sharpe ratio, Calmar ratio, win rate, and maximum drawdown. By comparing base learners and ensemble models, the research explores trade-offs between model complexity, interpretability, and robustness, offering insights into machine learning's effectiveness in volatile, noise-prone cryptocurrency markets.

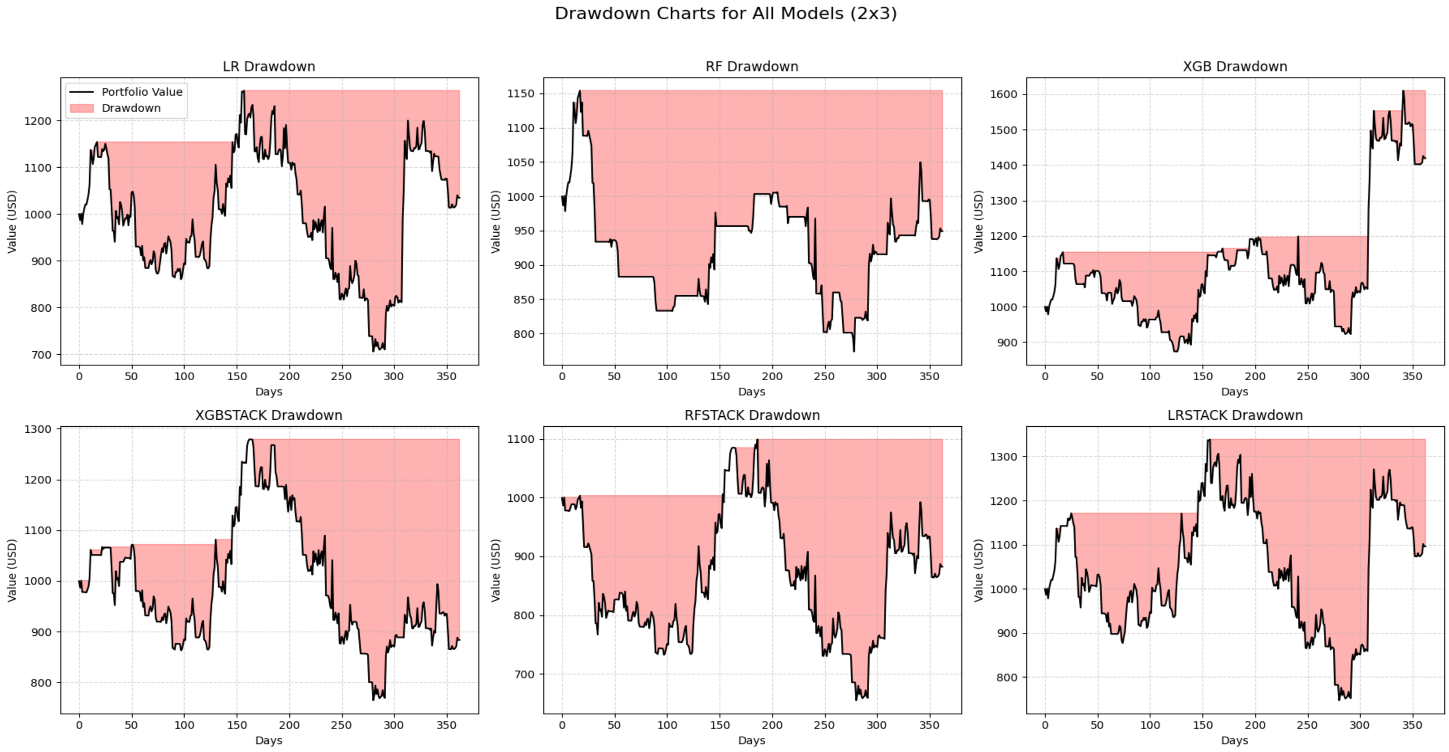

Figure 6: Drawdown charts

From the drawdown charts, it is evident that the XGB model demonstrates the most robust downside protection, with a maximum drawdown of only 0.24. In contrast, models such as LR and RFSTACK experience drawdowns exceeding 0.40, with extended periods of underperformance and delayed recovery. Notably, the RF model and its stacked variant remain underwater for most of the year, indicating limited resilience and poor adaptation to market dynamics.

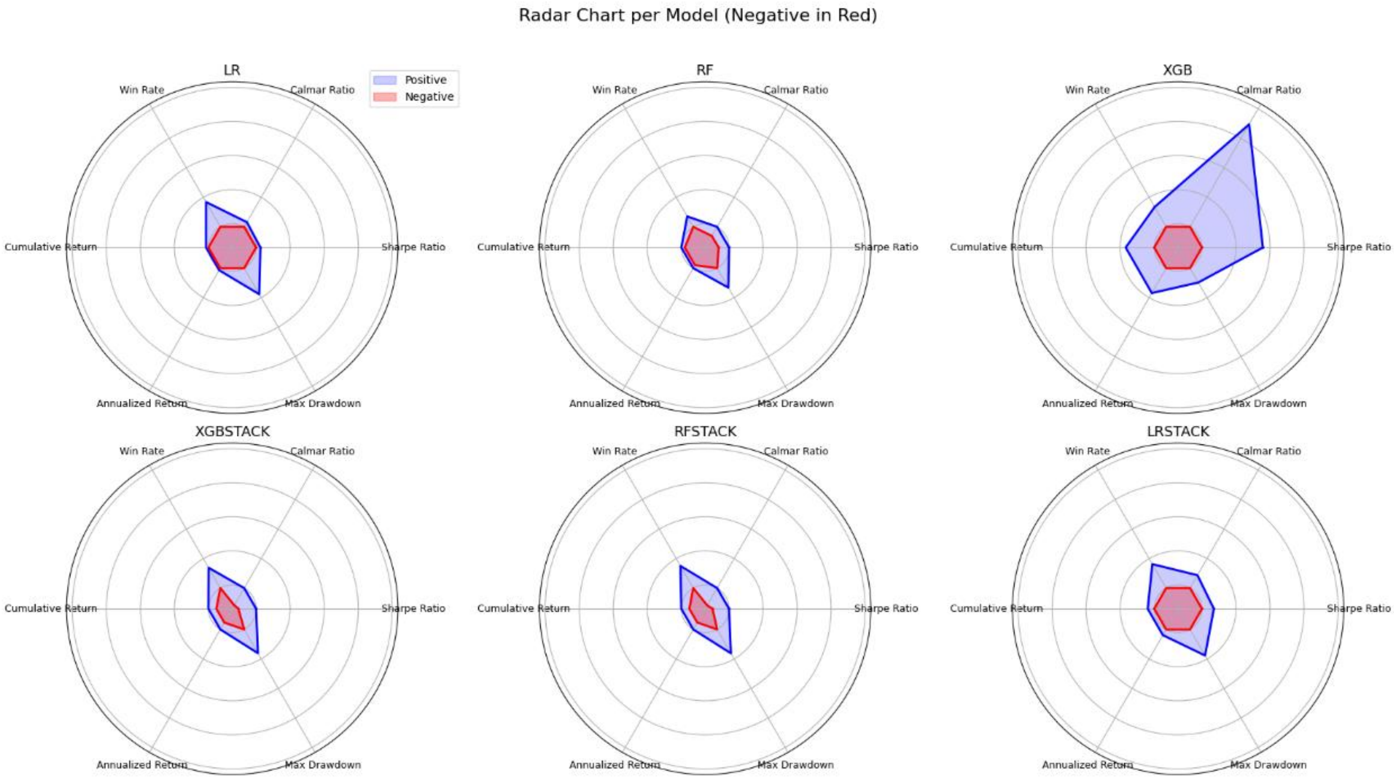

Figure 7: Radar plots

The radar plots support these observations. XGB consistently outperforms across all key performance metrics, achieving the highest Sharpe ratio (0.90), Calmar ratio (1.74), and cumulative return (0.42) among all models. This profile suggests a well-balanced trade-off between risk and return. While LRSTACK does not reach the same levels as XGB, it maintains positive values across all dimensions, showing no major weaknesses. This makes it a viable secondary option, particularly for more risk-averse strategies. In contrast, models like RF and XGBSTACK exhibit negative values in nearly every metric, indicating fundamental structural limitations.

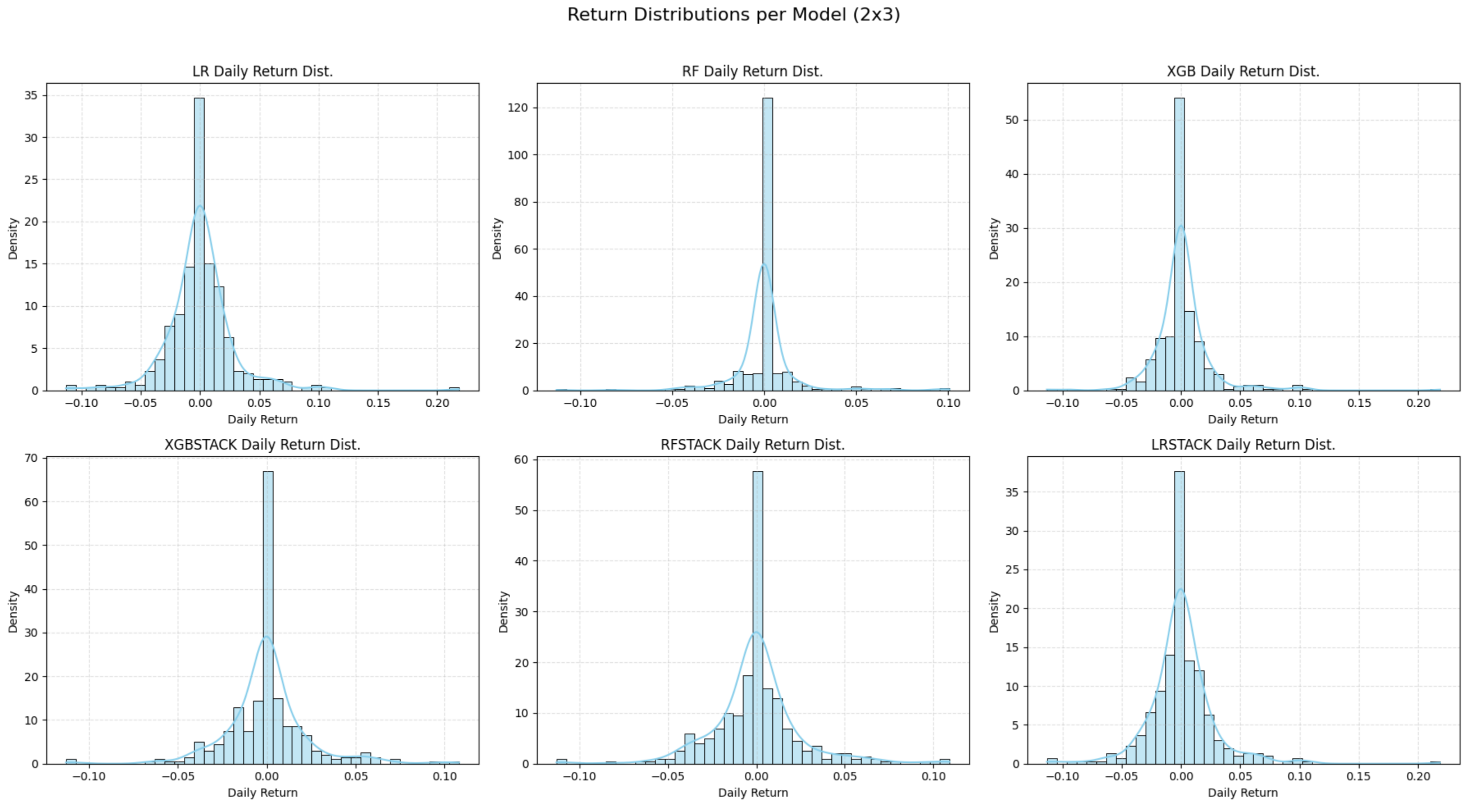

Figure 8: The distribution of daily returns

The distribution of daily returns provides further insights. XGB and LRSTACK show moderately right-skewed distributions with fatter tails, implying the potential for occasional strong positive returns. On the other hand, RF and RFSTACK display distributions tightly clustered around zero, highlighting their inability to generate significant alpha or capitalize on price trends.

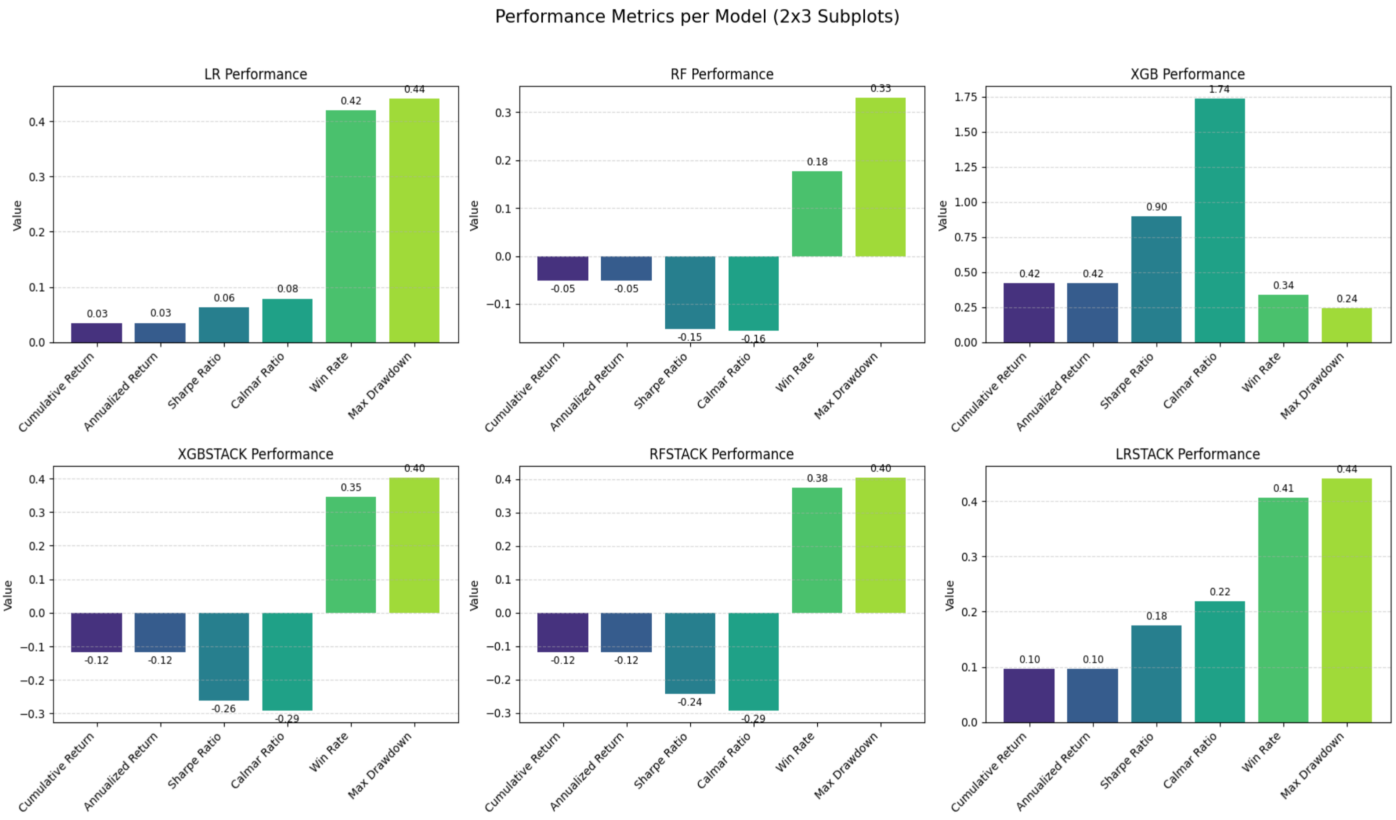

Figure 9: Performance bar plots

Lastly, the performance bar plots offer a comprehensive view of each model's effectiveness. XGB once again stands out, delivering an annualized return of 0.42 alongside its favorable risk profile. LRSTACK ranks second with consistent though modest gains (annualized return of 0.10, win rate of 0.41), making it a practical candidate for more conservative deployment. Models such as RF, RFSTACK, and XGBSTACK not only fail to produce positive returns but also suffer from significant drawdowns, reinforcing their unsuitability for real-world trading.

In summary, across multiple evaluation dimensions, XGBoost (XGB) emerges as the most effective and reliable model, offering strong returns alongside limited risk exposure. LRSTACK provides a viable alternative, particularly in contexts where capital preservation takes precedence. By contrast, RF-based models consistently underperform, lacking both profitability and stability, and appear ill-suited for deployment in a predictive trading system.

Among all models evaluated, the standalone XGB consistently delivered superior results—achieving the highest cumulative return (41.93%), a strong Sharpe ratio (0.90), and relatively low maximum drawdown. These outcomes suggest that XGB was able to extract meaningful signals from noisy, high-frequency crypto market data and translate them into effective trading actions.

In contrast, ensemble approaches such as RFSTACK, LRSTACK, and XGBSTACK did not enhance performance; in several cases, they led to deterioration in both absolute returns and risk-adjusted metrics. This indicates that the added model complexity may have introduced redundancy or diluted the predictive signal captured by the base learner. Such findings challenge the common presumption that ensemble methods inherently improve robustness or forecasting accuracy—particularly in volatile, data-sparse environments like cryptocurrency markets.

Taken together, the results underscore the importance of model parsimony and targeted feature design over architectural complexity. For both practitioners and researchers, these findings highlight the need to critically assess whether ensemble techniques offer substantive improvements or merely increase noise. In real-world trading applications—where interpretability, execution speed, and resource efficiency are critical constraints—a well-calibrated model like XGB may offer a more pragmatic and effective solution for short-term crypto prediction and strategy development.

4.4 Prediction

To evaluate the real-world applicability of the trained models, we conducted a next-day prediction for Ethereum's (ETH) closing price on July 2, 2025, using the most recent available features from July 1, 2025. All six models—three base learners (Linear Regression, Random Forest, and XGBoost) and their corresponding stacking variants—were deployed on the latest feature vector to generate price forecasts.

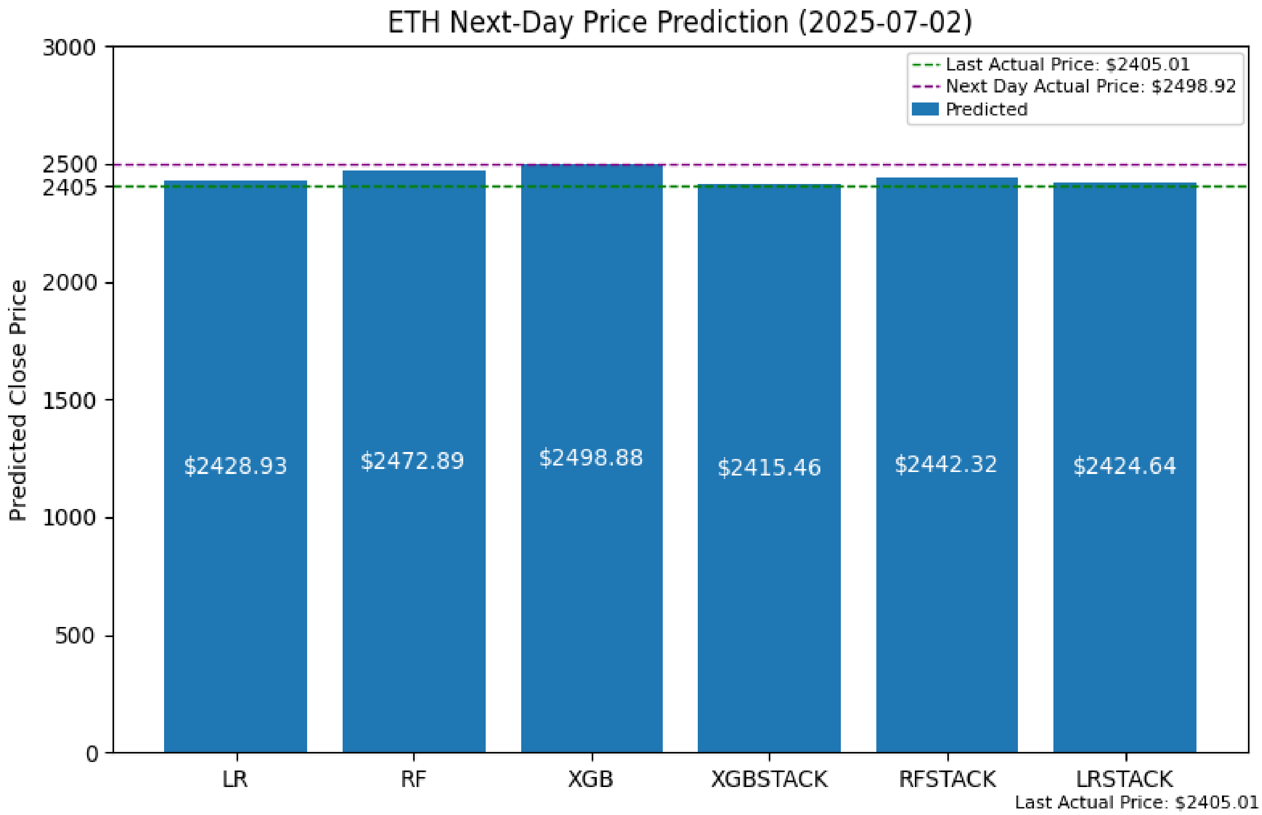

Figure 10: Next-Day Price Prediction

The true next-day closing price of ETH was $2,498.92, as recorded from Binance. Among all models, XGBoost yielded the most accurate forecast at $2,498.88, almost perfectly matching the actual value with a deviation of less than $0.05. This precise alignment reinforces the earlier findings on XGB's strong generalization capability and robustness in capturing short-term price dynamics.

In contrast, Random Forest exhibited the most pronounced overestimation, predicting $2,472.89, while Linear Regression slightly underpredicted at $2,428.93. Stacking models did not offer meaningful improvements—XGBSTACK and RFSTACK both underperformed their base model counterparts, continuing the pattern observed during backtesting. For example, XGBSTACK predicted only $2,415.46, underestimating the actual value by over $83, a non-trivial margin given the price level.

This real-world prediction test underscores a critical insight: model complexity is not a substitute for precision. While ensemble methods theoretically aim to smooth individual model biases, they can dilute well-learned signals when not carefully tuned. In the context of highly volatile and noisy markets like cryptocurrencies, a high-performing standalone model such as XGB may offer more reliable short-term forecasts than its stacked counterparts.

4.5 Conclusion

The results of this study highlight both the potential and the limitations of using Ethereum's on-chain data to inform short-term price prediction and trading strategy design. While metrics such as gas fees, transaction volume, and large wallet activity showed limited predictive strength on their own, their collective performance improved when used within ensemble learning frameworks. This points to the value of exploring feature interactions rather than relying on single indicators, which often fail to capture the complexity of market behavior in decentralized environments.

Interestingly, the models that achieved the lowest prediction errors did not always generate the strongest trading outcomes. This discrepancy suggests that conventional measures of predictive accuracy are not sufficient for evaluating models in financial settings where the ultimate goal is return generation. Including backtesting results as part of the evaluation process made it possible to compare statistical performance against actual portfolio outcomes, revealing trade-offs that are often overlooked in model development.

The superior performance of ensemble methods over deeper, more opaque architectures also draws attention to a broader issue: interpretability. In decentralized finance, where trust is mediated through code rather than institutions, models that can be understood and audited are often more useful than those that simply perform well in isolation. Choosing interpretable models does not necessarily require a compromise on predictive power, and may in fact support more reliable decision-making under uncertainty.

Taken together, these findings suggest that while blockchain-native data offers valuable signals, effective use depends on careful model design, robust validation procedures, and a clear alignment between predictive goals and financial outcomes. As DeFi markets continue to evolve, there is a growing need to move beyond surface-level metrics and toward more structured, system-aware approaches to forecasting and strategy development.

5. Discussion

5.1 Model Insights

The results from the empirical analysis reinforce a key observation: model complexity does not automatically translate into trading effectiveness. While models such as GRU are technically capable of capturing temporal dependencies in price series, their advantage appears less pronounced when evaluated through the lens of actual return performance. In this study, simpler ensemble models outperformed deep neural networks in backtesting, despite the latter achieving lower prediction errors in-sample. This divergence highlights a critical gap between statistical accuracy and real-world utility in algorithmic trading.

The ensemble methods used here benefited from their ability to balance predictive robustness with interpretability, a trait often missing in more opaque architectures. Unlike deep reinforcement learning or recurrent networks, ensemble learners allowed clearer diagnostics, making them more adaptable to the rapidly shifting dynamics of cryptocurrency markets. While past literature points to the strengths of GRU and deep learning models (Dutta et al., 2020), the current results suggest that model flexibility and responsiveness to new data may matter more than raw learning capacity in volatile environments.

One overlooked consideration in previous studies is the alignment between model performance metrics and practical trading outcomes. This research addressed that gap by comparing predictive models not just on RMSE or MAE, but also on their capacity to deliver sustainable, risk-adjusted returns over time. The finding that some models with modest statistical performance outperformed others in backtesting raises questions about how predictive quality should be defined in financial contexts.

5.2 Implications for DeFi Market Analysis

The broader implications extend to how market analysts and traders navigate decentralized finance environments. While blockchain-based systems promise greater transparency and autonomy, their data structures and transaction mechanics introduce noise and delay that complicate real-time prediction. For example, gas fee spikes may not immediately correlate with price shifts but can reflect latent network congestion that undermines transaction speed—a factor with indirect effects on price dynamics. The relatively weak predictive value of individual on-chain variables observed in this study suggests that relying on single metrics may be misleading unless combined with context-aware feature engineering.

The finding that machine learning models struggled during periods of regime change (e.g., sudden crashes or macroeconomic shocks) mirrors a larger concern in DeFi: the lack of embedded mechanisms for detecting systemic stress in decentralized systems. While Schär (2021) and Taherdoost (2023) rightly emphasize the efficiency and transparency offered by smart contracts, these very attributes can amplify the effects of poorly modeled risks or bugs in contract logic. In practice, integrating machine learning with blockchain analytics could enhance risk detection by providing early warning signals based on behavioral patterns, wallet flows, or governance vote anomalies—though this would require further development beyond current implementations.

The reliance on backtesting also allowed for a more grounded interpretation of model effectiveness. In traditional markets, models are often stress-tested using synthetic shocks. DeFi markets, in contrast, offer real-time data on sudden governance changes, liquidity exits, or code-level vulnerabilities. This makes them both richer and more chaotic as modeling environments. A clear implication is that models designed for DeFi must not only be accurate, but also adaptive and aware of context-specific risks.

5.3 Limitations

Several constraints limit the generalizability of this study. While ensemble models delivered promising results, their performance still depended on careful feature selection and parameter tuning. A major challenge was the limited granularity of the data: several potentially useful on-chain metrics were either not updated frequently enough or exhibited high multicollinearity, reducing their standalone utility.

The case of the Brooklyn Microgrid, while often cited in blockchain energy research, offers limited relevance to high-frequency financial environments. Its mention serves to illustrate the broader point that scalability remains a technical hurdle in blockchain ecosystems (Mengelkamp et al., 2017), but the analogy does not map directly onto crypto-asset markets.

More relevant to this study is the issue of interoperability across blockchain platforms. Portfolio strategies that include cross-chain assets face integration issues due to differing consensus mechanisms. Schär (2021) and Gai et al. (2020) highlight these limitations, and this study's findings reinforce their concerns: several promising variables from Layer 2 protocols or sidechains could not be included due to inconsistent data formatting or access limitations. This suggests that more standardized APIs and data protocols could significantly improve the scope and effectiveness of blockchain-based forecasting models.

Finally, while the models performed reasonably well over a short horizon, their utility in longer-term investment contexts remains uncertain. Cryptocurrency markets are known for sudden structural breaks, and no model in this study effectively predicted such shifts. As Fang et al. (2024) note, historical data alone is insufficient for capturing the full range of market dynamics. Until models incorporate a wider array of signals—including social sentiment, macro indicators, and governance activity—their predictive scope will remain bounded.

6. Conclusion

6.1 Summary of Findings

This research set out to explore the predictive power of machine learning models applied to Ethereum price dynamics within the broader context of decentralized finance. By constructing a structured dataset incorporating both market variables and key blockchain indicators—such as Total Value Locked (TVL), whale transactions, gas fees, and transaction volumes—the study aimed to identify meaningful relationships between on-chain activity and asset price movements.

Among the models tested, Random Forest demonstrated the best statistical accuracy, achieving the lowest Mean Squared Error (MSE) and the highest R-squared. In contrast, XGBoost outperformed all models in backtesting, generating the highest cumulative returns and Sharpe Ratio, indicating stronger practical effectiveness in trading applications. These findings highlight a common divergence between statistical fit and financial utility, particularly in volatile markets.

Contrary to expectations, stacking the predictions of Linear Regression, Random Forest, and XGBoost did not yield performance gains. The ensemble model did not significantly improve over the strongest individual models, suggesting limited complementarity under the given feature set and market conditions. This outcome diverges from earlier studies (e.g., Breiman, 2001; Chen & Guestrin, 2016), which often found ensemble learning to enhance model robustness in complex financial tasks.

Blockchain-derived features—especially whale-related metrics such as Whale_Count_Change and Whale_Amount_Change—emerged as significant predictors, reinforcing prior findings by Fang et al. (2024) and Carter and Jeng (2021). These features exhibited strong associations with price volatility, emphasizing the influence of large holders on market dynamics.

Backtesting results further revealed that model performance deteriorated during periods of extreme volatility, such as sudden gas price spikes or whale-driven transactions. This is consistent with the known challenges of modeling cryptocurrency markets, where regime shifts and external shocks are frequent (Weingärtner et al., 2023). While the use of risk-adjusted evaluation metrics such as the Sharpe Ratio provided a more realistic picture of strategy effectiveness, additional refinements are necessary to improve robustness under stressed market conditions.

6.2 Future Research Directions

One immediate opportunity for extending this research lies in broadening the scope of blockchain metrics used. Variables such as miner behavior, network difficulty, and block propagation times could potentially reveal more subtle mechanisms affecting asset prices. As noted by Saad et al. (2020), these network-level indicators, when combined with user activity data, could improve the accuracy of price predictions. Including more granular metrics such as block time and mining rewards could further capture Ethereum's behavior in response to network conditions and miner incentives.

The study also encountered issues related to data quality and latency, which are common in blockchain research but carry real implications for predictive modeling. While the current dataset was compiled from reputable external APIs, limitations in frequency and consistency may have affected both model training and evaluation. As indicated by previous research in cryptocurrency forecasting (Dutta et al., 2020), high-frequency and more granular blockchain data, especially from direct sources like Ethereum's blockchain, could offer more reliable datasets. Improving data quality and addressing missing values or inconsistencies could significantly enhance model performance. Furthermore, using real-time transaction history, smart contract interactions, and on-chain social sentiment data could provide a more comprehensive understanding of Ethereum's price behavior.

In terms of methodology, more sophisticated deep learning architectures present a potential avenue for future exploration. Although the current models demonstrated promising results, their ability to handle high volatility and adapt to changing market conditions can be improved. One promising approach would be the use of deep learning techniques, such as convolutional neural networks (CNNs) or Generative Adversarial Networks (GANs), which are increasingly being applied in financial market forecasting (Patel et al., 2020). These techniques are capable of capturing intricate patterns in large datasets and can offer better performance when dealing with complex and nonlinear relationships in cryptocurrency markets. Additionally, reinforcement learning could be integrated to enable models to adapt in real-time, providing better predictions during periods of market uncertainty.

Lastly, quantum computing presents an emerging direction for research. While still in a nascent stage, its capacity for parallel processing and optimization holds clear relevance to the kinds of data-intensive applications seen in blockchain analysis. As noted by Naik et al. (2025), the integration of quantum-enhanced machine learning techniques could lead to breakthrough improvements in model efficiency and accuracy, particularly in decentralized environments where traditional models struggle with volatility and data complexity.

6.3 Final Thoughts

The findings of this study contribute to a growing body of evidence that supports the value of blockchain-native data in financial modeling. Ethereum’s network activity contains signals that, when properly processed, can enhance price prediction and inform trading strategies. Integrating blockchain analytics into financial decision-making in DeFi markets offers a significant advantage in improving both the accuracy and transparency of market forecasts. This study highlights that incorporating Ethereum-specific features, such as whale transactions, gas fees, and TVL data, can significantly enhance the predictability of price movements. However, as observed in the backtesting phase, the volatility of cryptocurrency markets remains a significant challenge, with extreme market conditions often causing model performance to fluctuate.

Looking ahead, the ability to develop models that are both responsive and resilient will likely define the next phase of DeFi analytics. As DeFi continues to reshape traditional financial markets, the importance of blockchain analytics will only increase. Moving forward, models that can incorporate real-time blockchain data, social sentiment, and transaction patterns will be critical for making data-driven, informed decisions in the fast-evolving DeFi space. Understanding the complex interplay between Ethereum's price and network conditions will not only provide valuable insights for traders but also enhance risk management strategies, ensuring more robust and sustainable investment strategies in the volatile world of cryptocurrencies.

7. Appendices: Codes, APIs and Datasets

Datasets		
Name	**Access**	**Data**
Ethereum Historical Data.csv	Investing.com	ETH Date, Price
tvl.csv	DefiLlama API	ETH TVL
whale.csv	TokenView API	ETH Whale Activity
gas_fee.csv	Alchemy API	ETH Gas Fee
tx_count_volume.csv	Alchemy API	Transaction Count& Volume
ETH_combined_metrics.csv	Merged by MergeScript.py	All training features
XGBoost_backtest_results.csv	Generated by XGBoostModel.py	Backtest data for XGBoost Model
LinearRegression_backtest_results.csv	Generated by LinearRegressionModel.py	Backtest data for Linear Regression Model
RandomForest_backtest_results.csv	Generated by RandomForestModel.py	Backtest data for Random Forest Model
Stacking_backtest_results.csv	Generated by StackingModel.py	Backtest data for Stacking Model

Codes		
Name	**Overview**	**Purpose**
Get_Data.py	5 functions wrapped: get_gas_fee_eth_fee_history() get_whale_activity() get_eth_price_binance() get_transaction_count_and_volume() get_active_addresses()	Get on-chain data
MergeScript.py	Merge datasets: ETH_combined_metrics.csv Ethereum Historical Data.csv	Merge datasets
4_ModelBuilding.py	Implements multiple ML models (LR, RF, XGB, Stacking) and performs training.	To construct predictive models using ETH market features and prepare them for evaluation.
5_ModelEvaluation.py	Evaluates trained models using metrics such as MAE, MSE, R^2, Sharpe, etc.	To assess model performance across different statistical and financial evaluation dimensions.
6_Backtest.py	Simulates trading using model predictions under realistic conditions.	To conduct realistic backtesting with risk controls (e.g., drawdown, stop-loss) and log performance.
7_Prediction.py	Performs final-day prediction using trained models and visualizes outputs.	To forecast the next-day ETH price and compare model predictions with actual market price.
Visualisation.py	Contains all visualization functions for performance, risk, and errors.	To generate visual insights, such as error plots, drawdowns, heatmaps, and model comparison.

8. References

Akyildirim, S., Goncu, A., & Sensoy, A. (2020). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, *297*(3), 3–36. doi:10.1007/s10479-020-03575-y

AlShamsi, M., Al-Emran, M., & Shaalan, K. (2022). A systematic review on blockchain adoption. *Applied Sciences*, *12*(9), 4245. doi:10.3390/app12094245

Auer, R., Haslhofer, B., Kitzler, S., Saggese, P., & Victor, F. (2023). The technology of decentralized finance (DeFi). *Digital Finance*, *6*, 55–95. doi:10.1007/s42521-023-00088-8

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. doi:10.1023/A:1010933404324

Carter, N., & Jeng, L. (2021). DeFi protocol risks: The paradox of DeFi. *SSRN Electronic Journal*. doi:10.2139/ssrn.3866699

Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

Corbet, S., Larkin, C., Lucey, B., Meegan, A., & Yarovaya, L. (2020). Cryptocurrency reaction to FOMC announcements. *Journal of Financial Stability*, *46*, 100706. doi:10.1016/j.jfs.2019.100706

Doerr, J. F., Kosse, A., Khan, A., Lewrick, U., Mojon, B., Nolens, B., & Rice, T. (2021). DeFi risks and the decentralisation illusion. *BIS Quarterly Review*, *21*.

Dutta, A., Kumar, S., & Basu, M. (2020). A gated recurrent unit approach to Bitcoin price prediction. *Journal of Risk and Financial Management*, *13*(2), 23. doi:10.3390/jrfm13020023

Eisl, A., Gasser, S. M., & Weinmayer, K. (2015). Caveat Emptor: Does Bitcoin Improve Portfolio Diversification? doi:10.2139/ssrn.2408997

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? *ESSEC Working Paper*. arxiv.org/abs/1312.1645

Erarslan, B., & Uslu, H. (2023). Application of gradient boosting algorithms for anti-money laundering in cryptocurrencies. *Expert Systems with Applications*, *219*, 119654.

Fang, F., Ventre, C., Basios, M., Kanthan, L., Martinez-Rego, D., & Wu, F. (2022). Cryptocurrency trading: A comprehensive survey. *Financial Innovation*, *8*(13). doi:10.1186/s40854-021-00321-6

Fischer, T. G., Krauss, C., & Deinert, A. (2019). Statistical arbitrage in cryptocurrency markets. *Journal of Risk and Financial Management*, *12*(1), 31. doi:10.3390/jrfm12010031

Gai, K., Guo, J., Zhu, L., & Yu, S. (2020). Blockchain meets cloud computing: A survey. *IEEE Communications Surveys & Tutorials, 22*(3). doi:10.1109/COMST.2020.2989392

Gencer, A. E., Basu, S., Eyal, I., van Renesse, R., & Sirer, E. G. (2018). Decentralization in Bitcoin and Ethereum Networks. In *Financial Cryptography and Data Security*. doi:10.1007/978-3-662-58387-6_24

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications, 38*(8), 10389–10397. doi:10.1016/j.eswa.2011.02.068

Huang, X., Lee, C. F., & Chen, H. (2005). Predicting stock price movements using machine learning algorithms. *Financial Engineering Review, 5*(2), 187–208.

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for financial portfolio management. arxiv.org/abs/1706.10059

Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement. *Expert Systems with Applications, 38*(5), 5311–5319. doi:10.1016/j.eswa.2010.10.027

Khedr, A. M., Arif, I., Raj, P. P. V., El-Bannany, M., Alhashmi, S. M., & Sreedharan, M. (2022). Cryptocurrency price prediction using traditional and ML techniques: A survey. *Journal of Risk and Financial Management, 15*(10), 469. doi:10.1002/isaf.1488

Lee, J. Y. (2019). A decentralized token economy. *Business Horizons, 62*(6), 773–784. doi:10.1016/j.bushor.2019.08.003

Lizcano, D., Lara, J. A., White, B., et al. (2020). Blockchain-based model of trust in education. *J Comput High Educ, 32*, 109–134. doi:10.1007/s12528-019-09209-y

Livieris, I. E., Pintelas, E., Stavroyiannis, S., & Pintelas, P. (2020). Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms*, *13*(5), 121. doi:10.3390/a13050121

Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., & Weinhardt, C. (2017). Designing microgrid energy markets. *Applied Energy, 210*, 870–880. doi:10.1016/j.apenergy.2017.06.054

Naik, A. S., Yeniaras, E., Hellstern, G., Prasad, G., & Vishwakarma, S. K. L. P. (2025). Quantum computing in finance: A systematic review. *Financial Innovation, 11*(88). doi:10.1186/s40854-025-00751-6

Patel, J., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N., & Alazab, M. (2020). Stochastic neural networks for cryptocurrency price prediction. *IEEE Access, 8*, 82804–82817. doi:10.1109/ACCESS.2020.2990659

Poongodi, A., Poornachandra, S. R., & Suresh, R. (2020). Prediction of Ethereum price in industrial finance. *Computational Economics, 56*(4), 799–819. doi:10.1016/j.compeleceng.2019.106527

Rao, K. R., et al. (2023). Time-series cryptocurrency forecasting using ensemble deep

learning. *2023 ICCPCT*, 1446–1453. doi:10.1109/ICCPCT58313.2023.10245083
Reijers, W., et al. (2021). On-chain and off-chain governance of blockchain. *Topoi, 40*(3), 821–831. doi:10.1007/s11245-018-9626-5
Saad, M., Choi, J., Nyang, D., Kim, J., & Mohaisen, A. (2020). Toward characterizing blockchain cryptocurrencies. *IEEE Systems Journal, 14*(1), 321–332. doi:10.1109/JSYST.2019.2927707
Schär, F. (2021). Decentralized finance: On blockchain- and smart contract-based financial markets. *Federal Reserve Bank of St. Louis Review, 103*(2), 153–174. doi:10.20955/r.103.153-74
Sharma, P. M., et al. (2020). Prediction of Ethereum price. *Computers and Electrical Engineering, 81*, 106527. doi:10.1016/j.compeleceng.2019.106527
Stavroyiannis, S. (2018). Value-at-risk and related measures for the Bitcoin. *The Journal of Risk Finance, 19*(2), 127–136. doi:10.1108/JRF-07-2017-0115
Taherdoost, H. (2023). Smart contracts in blockchain technology: A critical review. *Information, 14*(2), 117. doi:10.3390/info14020117
Yang, Q., Zhao, Y., Huang, H., Xiong, Z., Kang, J., & Zheng, Z. (2022). Fusing blockchain and AI with metaverse: A survey. *IEEE Open Journal of the Computer Society, 3*, 122–136. doi:10.1109/OJCS.2022.3188249
Zhou, L., Qin, K., Cully, A., Livshits, B., & Gervais, A. (2021). Just-In-Time discovery in DeFi protocols. *IEEE Symposium on Security and Privacy*. doi:10.1109/SP40001.2021.00113