Journal of Posthumanism
2025
Volume: 5, No: 7, pp. 802-826
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i7.2843

Disclosure Determinants of Blockchain Crowdfunding Performance for Sustainable Smart City Financing: An Explainable Optuna-Optimized Machine Learning Approach

Gihan Ali[1], Zakaria Yahia[2]

Abstract

Smart cities present a transformative paradigm for urban development, yet securing sustainable financing remains a critical challenge. While traditional funding mechanisms struggle with scalability limitations, FinTech innovations like Initial Coin Offerings (ICOs) have emerged as a viable alternative. Leveraging blockchain technology, ICOs enable decentralized capital raising through token sales, offering transparency and global investor access. However, their effectiveness is compromised by market volatility, information asymmetry, and the absence of reliable predictive frameworks. This study addresses these limitations by developing an explainable hybrid machine learning model that combines: (1) Light Gradient Boosting Machine (LGBM) for efficient feature selection through histogram-based learning, (2) Optuna-optimized Extremely Randomized Trees regression that mitigates overfitting via enhanced randomization while excelling with noisy financial data, and (3) interpretability tools including SHAP values and feature importance analysis. Optuna's automated hyperparameter optimization further enhances computational efficiency, enabling robust predictions of post-ICO returns. The proposed model demonstrates superior predictive performance (R^2=0.814, MSE=0.005, MAE=0.051), significantly outperforming both linear regression and state-of-the-art ML models. Key findings identify token supply (63% predictive power) as negatively correlated with returns - reflecting dilution effects and investor perceptions of scarcity- while fundraising success (15%) and Bitcoin returns (8%) show positive influences. These results provide practical guidance for investors and regulators, while establishing ICOs as a potential sustainable financing mechanism for smart city initiatives. The study contributes both methodologically through its optimized hybrid architecture and practically by enhancing decision-making in blockchain-based urban development financing.

Keywords: *Initial Coin Offering Performance Prediction, Disclosure Determinants, Light Gradient Boosting Machine (LGBM), Extremely Randomized Trees regression (Extra Trees), Optuna Optimization, Explainable Machine Learning (ML), Sustainable Finance, Smart Cities.*

Introduction

Smart cities are rapidly emerging as a solution to urban challenges, driven by advances in technology, sustainability, and digital innovation. However, one of the critical challenges in their development is securing sustainable funding sources. Traditional funding methods such as government budgets and public-private partnerships (PPPs), while still relevant, may not always be sufficient to support the long-term development of smart cities, especially in an era of rapid technological change. FinTech tools have recently emerged as innovative and sustainable finance methods. Initial Coin Offerings (ICOs) recently gained attention as a novel funding

[1] Department of Accounting, College of Business Administration in Hawtat Bani Tamim, Prince Sattam bin Abdulaziz University, Saudi Arabia, Email: g.ali@psau.edu.sa, (Corresponding Author).
[2] Industrial and Manufacturing Engineering Department, Innovative Design Engineering School, Egypt-Japan University of Science and Technology, Egypt, Email: zakaria.yahia@ejust.edu.eg, Mechanical Engineering Department, Faculty of Engineering, Fayoum University, Egypt, Email: zakaria.yahia@fayoum.edu.eg

mechanism. ICOs are blockchain-based smart contracts designed to raise capital by issuing digital coins or tokens. A blockchain serves as a decentralized, transparent ledger that records transactions securely and immutably (Iansiti & Lakhani, 2017). Within ICOs, smart contracts function as self-executing protocols that automate transactions, eliminating intermediaries and reducing fundraising costs for firms (Momtaz, 2020). While structurally akin to crowdfunding, ICOs uniquely involve the sale of tokens—cryptographically secured digital assets that provide utility, currency, or security benefits to investors (Akiba et al., 2019). These tokens can later be traded on secondary markets post-ICO. Therefore, ICOs offer a transparent, efficient alternative to traditional financing, particularly for smart city projects. By facilitating direct token sales to global investors, they bypass conventional funding barriers while enabling rapid scaling of urban innovations. The token economy promotes sustainable resource management and democratized investment, reducing transaction costs and enhancing transparency in urban development. However, risks like regulatory uncertainty, volatility, and information asymmetry persist (Catalini & Gans, 2020).

Predicting ICO success is challenging due to information asymmetry and market volatility. Early research relied on statistical models like OLS regression, but these approaches struggled with limitations such as multicollinearity and low predictive accuracy (Alaka et al., 2018; Fahlenbrach & Frattaroli, 2021; Lyandres et al., 2022). More recent studies have turned to machine learning (ML), achieving better results. For instance, Meoli & Vismara (2022) demonstrated ML's superior predictive power, while Wang et al. (2022) proposed a deep regression model for document analysis, showing how whitepaper structure influences ICO outcomes. Additional factors like team expertise and social media sentiment have also been incorporated through advanced techniques like attention-based deep learning (Xu et al., 2021) and hybrid ML models addressing class imbalance (Ali et al., 2022). The integration of ML models has enhanced prediction accuracy while incorporating multiple factors. Nonetheless, most current research predominantly emphasizes forecasting ICO success rather than evaluating post-ICO performance.

This study addresses these gaps by proposing an intelligent decision model that integrates: (1) Light Gradient Boosting Machine (LGBM) for feature selection, (2) Optuna-optimized Extremely Randomized Trees regression (Extra Trees) for predicting ICO returns, (3) feature importance analysis, and (4) Shapley Additive Explanations for interpretability.

LGBM, originally developed by Microsoft researchers in 2017 (Ke et al., 2017), was used to identify and analyze both financial and non-financial signals that influence ICO performance. LGBM is a gradient boosting framework optimized for handling large datasets. Its speed, efficiency, and accuracy enable it to manage extensive datasets with distributed processing, often outperforming other boosting methods. The framework's low memory consumption is due to its conversion of continuous values into discrete bins (McCarty et al., 2020). Additionally, LGBM's built-in feature importance function, which measures the frequency of each feature's use in tree construction, plays a crucial role in feature selection (Sarıkaya et al., 2022). By setting a threshold, features with scores above it are selected, leading to more accurate and interpretable models (Banga et al., 2023; Chandrashekar & Sahin, 2014).

Extra Trees represent a powerful ensemble learning technique, particularly suited for handling complex and noisy datasets. Developed by Geurts et al. (2006), Extra Trees are an ensemble learning technique that builds upon Random Forest by introducing additional layer of randomness to the tree construction process. This added randomness helps reduce overfitting and variance compared to other ensemble methods. Extra Trees also tend to be more computationally efficient than Random Forest due to their strategy of randomly selecting split points. Additionally, they demonstrate superior performance when dealing with datasets containing noisy data points, a common characteristic of financial data. However, the random split point selection can potentially introduce some bias. Using the full training dataset helps mitigate this bias (Geurts et al., 2006; González et al., 2020; Majidi et al., 2022).

Nonetheless, optimal ML model performance requires careful hyperparameter tuning, as these parameters critically determine model outcomes. While traditional methods like Grid Search and gradient-based optimizers (SGD, Adam) have been widely used, they face computational limitations with complex models in high-dimensional spaces. To address these limitations, modern optimization frameworks like Optuna have emerged. Optuna automates hyperparameter optimization through adaptive sampling and efficient search algorithms, enhancing computational efficiency while optimizing model performance. The framework rapidly converges to optimal parameter configurations (Akiba et al., 2019). Its dynamic parameter space exploration makes it highly effective for optimizing tree-based ensembles like Extra Trees and XGBoost (Lai et al., 2024).

Identifying the disclosure determinants that significantly influence Initial Coin Offering (ICO) performance is critical for stakeholders, including investors, regulators, and project managers. To derive actionable insights, feature importance analysis is employed to quantify the relative contribution of each predictor variable in forecasting ICO returns. To enhance the interpretability of the proposed model, a tree-based machine learning approach is utilized, leveraging TreeExplainer to ensure both transparency and a robust theoretical foundation. This methodology not only identifies the most influential predictors but also elucidates their marginal effects on ICO performance. TreeExplainer operationalizes Shapley values to aggregate local explanations into a global model interpretation, thereby facilitating a comprehensive assessment of feature interactions and predictive dynamics (Lundberg et al., 2020). Furthermore, Partial Dependence Plots (PDPs) are employed to visualize the conditional relationships between the top predictors and post-ICO returns, offering additional clarity on their individual effects.

This study makes several important contributions to the existing literature. First, it introduces a novel hybrid optimized ML model, specifically designed to address the complexities of ICO performance prediction, demonstrating significant improvements over traditional approaches. Second, our deployment of Optuna for hyperparameter tuning reduces computational cost compared to traditional Grid Search (Akiba et al., 2019), while maintaining robustness in high-dimensional spaces—a critical advantage for noisy datasets. The optimized Extra Trees ensemble further mitigates overfitting risks inherent in financial data (Geurts et al., 2006; Majidi et al., 2022). Third, the model's performance is rigorously evaluated using a range of statistical metrics and compared against statistical models, individual ML algorithms, and other ensemble methods, including bagging, and boosting ensembles. The findings consistently highlight the superior predictive accuracy and robustness of the proposed model. Fourth, by employing the fine-tuned extra trees for feature importance analysis, the study identifies the most influential variables driving ICO returns, offering valuable insights for investors and other stakeholders aiming to make well-informed decisions. Finally, to enhance interpretability, SHapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDP) are used to explain the global structure of the model and the effect of each variable, ensuring transparency without compromising performance. Collectively, these contributions advance the literature on ICO performance prediction, while offering practical guidance for investors, Regulators, and project managers operating in the complex and highly volatile FinTech market.

Empirical results demonstrate that our proposed model—which combines LGBM for feature selection and Optuna-optimized Extra Trees regression for ICO return prediction—significantly outperforms traditional linear regression, individual ML models, and other ensemble approaches. The proposed model achieves superior performance across all evaluation metrics, with an R^2 of 0.814, MSE of 0.005, and MAE of 0.051. This significantly outperforms the best individual benchmark model (Random Forest), which yielded an R^2 of 0.715, MSE of 0.008, and MAE of 0.068. Feature importance analysis reveals token supply (0.63), total funds raised (0.15), and Bitcoin returns (0.08) as the primary determinants of ICO returns, ranked by their relative predictive power. SHAP and partial dependence analyses show token supply as the strongest negative predictor of ICO returns, likely due to dilution effects and investor perceptions of scarcity. Both total funds raised and Bitcoin returns positively influence ICO returns. These findings suggest that robust fundraising serves as a project viability signal that boosts investor confidence, while simultaneously demonstrating the interdependence between ICO performance and broader cryptocurrency market trends.

The paper is organized as follows: Section 2 reviews relevant literature, Section 3 details our methodology, Section 4 presents experimental results comparing our approach with alternative methods, followed by feature importance analysis and model interpretation, and Section 5 concludes.

Literature Review

This review examines blockchain-based ICOs as a sustainable funding mechanism for smart city initiatives. Additionally, it discusses recent studies that employ various predictive models for forecasting ICO performance.

ICOs as a Financing Mechanism for Smart Cities

The financial landscape for smart cities is evolving rapidly, with various funding sources being explored to support their development. Blanck & Ribeiro (2021) describe a Smart City as enabling more efficient services, improved infrastructure monitoring, enhanced collaboration among economic actors, and innovative business models in public and private sectors. Financing smart cities differs from traditional urban projects due to their reliance on emerging technologies like IoT, cloud computing, and green tech, which require specialized financial models (Catalini & Gans, 2018; Momtaz, 2020). Unlike conventional infrastructure funded through loans or bonds, these high-risk, tech-driven initiatives demand innovative funding mechanisms. The prevalence of startups and tech ventures in smart cities further necessitates tailored financing solutions beyond traditional approaches, calling for diverse funding sources aligned with technological and entrepreneurial needs.

Given the complexity and novelty of financing smart cities, a variety of funding sources have been identified, each with its respective advantages and challenges. Public funding (municipal, national, or supranational) supports innovation via policies, grants, or investment banks but faces fiscal limits. Private capital (Venture Capital, corporate investments) drives R&D and startups, while PPPs enable large-scale projects by sharing risks. Crowdfunding aids small community initiatives, and regional instruments (EU funds, bonds) target infrastructure needs (Blanck & Ribeiro, 2021; Hedegaard, 2024; Mirzaee & Sardroud, 2022).

According to Kavta and Yadav (2017), smart city development can be financed through a diverse set of sources (Figure 1), each with its strengths and limitations. Traditionally, debt financing—including bank loans and bonds—offers stable and predictable capital but is often constrained by interest rate uncertainties, high interest costs, and the need for collateral. Equity investments, such as venture capital, provide substantial funding opportunities for innovative projects; however, they typically involve shared risk, profit-sharing, and may be sensitive to market fluctuations. Multilateral financial institutions (e.g., the World Bank) supply grants, loans, and capacity-building support, but their approval processes can be lengthy and are often tied to policy conditions. Crowd financing, including public contributions and mini-bonds, offers access to broad communities for raising capital quickly, yet it is limited by regulatory hurdles, investor trust, and scalability constraints. Lastly, innovative models like green bonds and payment enablers leverage financial innovation to fund sustainability and mobility, but they may face market volatility and policy risks.

However, traditional funding methods, such as PPPs and Venture Capital (VC), often fall short in addressing the unique demands of urban innovation projects due to bureaucratic hurdles and limited access to capital (Hollands, 2020; Kitchin, 2014). As cities increasingly embrace digital transformation, there is a growing need for alternative funding mechanisms that are scalable and flexible. Therefore, ICOs have emerged as a transparent and efficient funding mechanism for smart city projects, particularly for innovative but high-risk components such as IoT infrastructure, green technologies, and urban startups (Narayanan et al., 2016). By facilitating direct token sales to global investors, this approach circumvents traditional financing barriers while democratizing access to capital and enabling rapid scaling of urban innovations. The tokenized economy inherent in ICOs also promotes sustainable resource management while simultaneously reducing transaction costs and enhancing transparency in urban development (Catalini & Gans, 2020).

Specifically, ICOs offer three significant advantages compared to traditional funding methods. First, they achieve near-zero transaction costs through blockchain-based smart contracts that eliminate intermediaries and redistribute value across the network. Second, the ability to quickly list tokens on exchanges provides immediate liquidity - a critical feature for investors who typically avoid long lock-up periods. Research indicates that ICOs often employ strategic token undervaluation to attract broader investor participation, thereby enhancing market depth and platform viability. Third, ICO participants frequently serve dual roles as both investors and end-users, enabling projects to secure early market validation while fostering long-term engagement through potential token appreciation (Catalini & Gans, 2018; Momtaz, 2020). By combining cost efficiency, liquidity advantages, and innovative investor engagement models, ICOs represent a transformative development in project financing, complementing traditional methods by mobilizing global capital for next-generation smart city initiatives.

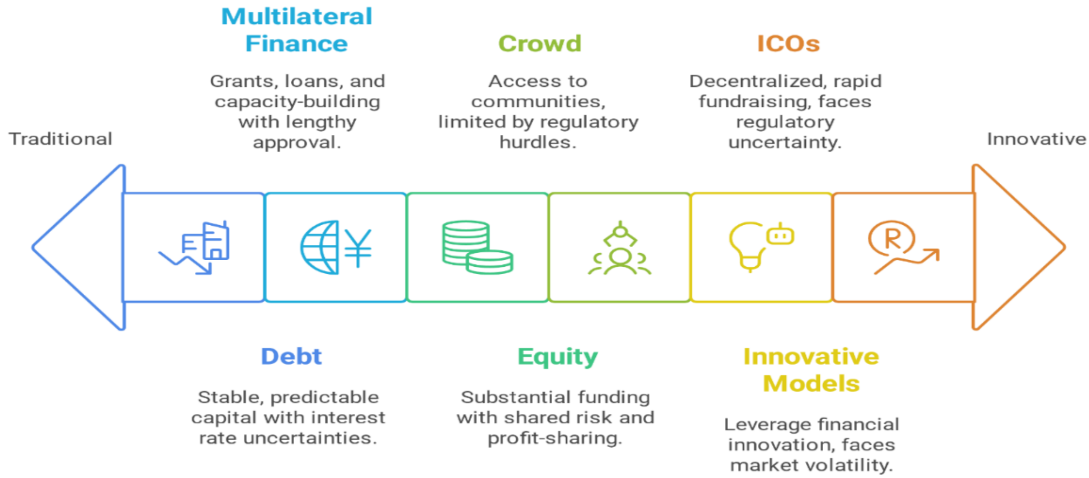


Figure 1. Smart City Financing Sources from Traditional to Innovative Approaches

However, significant challenges persist, including regulatory uncertainty, price volatility, and information asymmetries that can affect ICO market stability (Catalini & Gans, 2020). Building on Spence (1973) work, signaling theory suggests that high-quality ventures can communicate their value to potential investors by sending signals that reduce information asymmetry. These signals help investors differentiate between ventures of varying quality, allowing higher-quality ventures to attract more funding. Importantly, ICOs play a dual role in the context of information asymmetry, as they can both mitigate and worsen this challenge. On one hand, ICOs leverage blockchain technology to enhance transparency by recording transactions publicly and utilizing smart contracts to automate processes, reducing the risk of fraud. Moreover, they democratize access to funding, enabling direct communication between startups and investors while bypassing intermediaries. Furthermore, quality signals such as well-crafted whitepapers, team credentials, and technical roadmaps can help investors make informed decisions. On the other hand, ICOs also raise information asymmetry due to their unregulated nature, which allows ventures to exaggerate or omit critical details, and their reliance on technically complex blockchain technology, which many investors struggle to understand. The early-stage nature of most ICO-funded projects and the anonymity often associated with ventures further complicate due diligence, increasing uncertainty and the risk of fraud (Fisch, 2019).

This information asymmetry leads to market inefficiencies. Momtaz (2021) investigated the occurrence of moral hazard in the context of ICOs. His research reveals that in the absence of regulatory oversight, entrepreneurs often exaggerate information in whitepapers to attract more investments quickly. While this strategy helps raise funds faster, it backfires once tokens are publicly traded, as investors adjust their valuations based on collective market knowledge, leading to lower returns and increased likelihood of venture failure. The study supported two key hypotheses: the "asymmetric information hypothesis," where exaggerated signals are initially profitable, and the "crowd-learning hypothesis," where these signals are penalized after trading begins. Additionally, he examined how information asymmetry in the ICO market can lead to agency conflicts, where investors struggle to differentiate between high- and low-quality ventures due to limited transparency.

Collectively, the risks associated with information asymmetry inside the unregulated markets of ICOs highlight the critical role of signals that attract investment and ensure long-term success for ventures. By identifying key predictors using LGBM and analyzing critical signals through feature importance, SHAP values, and dependency plots, this research offers actionable insights for fintech investors and developers while improving transparency and trust in the ICO ecosystem.

Evolution of Predictive Models for ICO Performance

The prediction of ICO performance has evolved from traditional statistical methods to advanced ML techniques. Early research relied on traditional statistical methods, primarily ordinary least squares (OLS) regression. Lyandres et al. (2022) used OLS to demonstrate that ICO returns are positively correlated with platform adoption (measured by wallet activity and transaction volume), suggesting that an active investor base enhances long-term performance. Additionally, their findings indicate a positive correlation between token returns and the performance of both Bitcoin and Ethereum. Similarly, Fahlenbrach & Frattaroli (2021) applied OLS and found that longer lockup periods and source code disclosure improve ICO returns by reducing fraud risk, while presale discounts negatively impact performance. However, OLS models face limitations, including sensitivity to multicollinearity and outliers (Alaka et al., 2018). To address endogeneity, Domingo et al. (2020) employed a two-step system Generalized Method of Moments (GMM) estimator, revealing that ICO returns are positively influenced by Bitcoin spot/futures returns and social media sentiment but negatively affected by presale periods. Expanding on market dynamics, Aslan et al. (2023) found that offer price and investor sentiment are critical drivers of post-ICO performance, though their effects vary between bull and bear markets.

Textual analysis has also been used to assess ICO whitepapers. Florysiak & Schandlbauer (2022) showed that high-quality issuers provide more informative whitepapers. They also found that while investors initially rely on expert ratings, once tokens become publicly traded, OLS analysis shows informative white papers correlate with key market indicators including underpricing levels, investment returns, and market liquidity. Supporting this, (Hsieh & Oppermann, 2021) and Zhang et al. (2019) used regression analysis to establish that concise whitepapers, shorter fundraising periods, and readable disclosures significantly improve initial returns. Alternative approaches include generalized linear models (GLM). Shrestha et al. (2021) applied GLM and found that ICOs from countries with stronger institutions raise more funds, achieve higher liquidity, and exhibit lower volatility. Meanwhile, Fuchs & Momtaz (2024) examined governance mechanisms, showing that token retention boosts fundraising, whereas resale restrictions harm post-ICO performance—effects moderated by the founding team's human capital. While these methods advanced understanding, they often face limitations including assumptions of variable independence, restrictive dataset requirements, multicollinearity issues, and sensitivity to outliers (Alaka et al., 2018).

To overcome these challenges, researchers have increasingly adopted ML-based approaches to improve the accuracy of ICO predictions. Meoli & Vismara (2022) applied ML techniques to analyze ICOs launched between 2014 and 2019, achieving substantially higher forecasting accuracy compared to standard Logit models. Wang et al. (2022) created a deep regression model for document analysis, designed to extract features from whitepapers, and demonstrated that structure and presentation of these documents are key factors in forecasting ICO success. Xu et al. (2021) used an attention-based deep learning model to study online comments, emphasizing the significance of social indicators like expert evaluations and team expertise in determining ICO success. Similarly, Ali et al. (2022) introduced a hybrid intelligent decision-making model that combines Information Gain Directed Feature Selection with a Fuzzy Support Vector Machine to address class imbalance in predicting ICO success. Their model surpassed traditional classifiers, proving the effectiveness of hybrid machine learning models in tackling the complexities of ICO markets. However, those studies focus primarily on predicting ICO success rather than assessing post-ICO performance. This study addresses these gaps by introducing a hybrid intelligent decision model, which integrates Light Gradient Boosting Machine (LGBM) for feature selection with fine-tuned Extra Trees regression to assess post-ICO performance.

Methodology

This section begins by outlining the dataset used for analysis before introducing our novel hybrid model for ICO investment decisions. As shown in Figure 2, our framework integrates three key components: LGBM for efficient feature selection, Optuna-optimized Extra Trees for return prediction, and interpretability tools to extract actionable insights for stakeholders. The system operates through distinct training and testing phases, each designed to address the complexities of ICO performance prediction. The following subsections provide detailed explanations of each component, including their specific objectives and implementation within our analytical framework.

Dataset Description

This study leverages publicly available data on ICOs, compiled by Fahlenbrach & Frattaroli (2021). The dataset construction involves compiling a list of completed ICOs from secondary sources and filtering for those with total funding exceeding $1 million. The data, initially collected manually, originates from primary sources such as whitepapers, issuer documents, archived websites, social media updates, Github repositories, bitcointalk.org announcements, and national commercial registries. Secondary market prices in USD are sourced from coinmarketcap.com. The dependent variable is the logarithmic return, calculated 270 calendar days post-ICO completion using the average crowdsale price. If the average price is unavailable, the return is computed using the midpoint between the maximum and minimum crowdsale prices. For delisted ICOs, the return is based on the last available price before delisting.

Data Preprocessing

Since real-world datasets frequently contain missing values, errors, inaccuracies, and inconsistencies that may significantly degrade model performance, data preparation is an essential stage in ML (Ramírez-Gallego et al., 2017). To ensure the reliability and accuracy of the model, the dataset is carefully processed using a number of preprocessing methods, including dataset cleaning, normalization, and instance selection (Nguyen et al., 2021). Missing crowdsale price data are excluded, while a median filter is applied to handle missing values in the remaining features. The Winsorization strategy, capping extreme numbers at the 5th and 95th percentiles, is utilized to minimize the influence of outliers (Kwak & Kim, 2017). Additionally, to address data skewness, all continuous features undergo log transformation. Before using the log transformation to account for zero values, a constant value of one is added for every feature. The final preprocessed dataset consists of 292 ICOs.

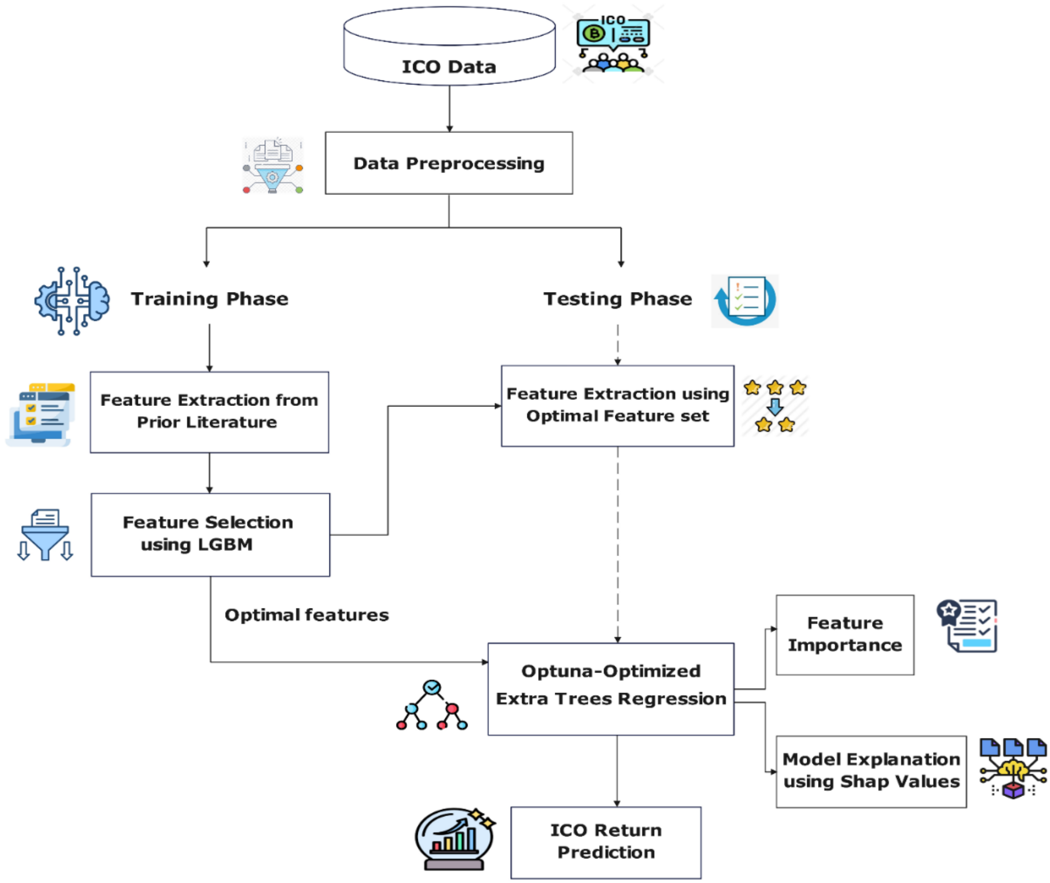


Figure 2. Block Diagram of the Proposed Intelligent Decision Model for ICO Performance Prediction

Feature Extraction

Feature extraction aims to identify key variables that could influence post-ICO performance, ensuring that the most relevant factors are captured for accurate prediction. Appendix Table 1 presents the comprehensive set of features extracted for analysis, which are drawn from prior studies on ICO performance determinants (e.g., Ahmad et al., 2023; Amsden & Schweizer, 2018; Benedetti & Kostovetsky, 2021; Blaseg, 2018; Boreiko & Risteski, 2021; Bourveau et al., 2022; Davydiuk et al., 2023; Domingo et al., 2020; Fahlenbrach & Frattaroli, 2021; Fisch & Momtaz, 2020; Howell et al., 2020; Momtaz, 2020, 2021; Roosenboom et al., 2020; Shrestha et al., 2021; Yen et al., 2021). These features encompass a comprehensive range of 60 independent variables categorized into eight groups: Team Characteristics, Token Characteristics, ICO Characteristics, Governance Characteristics, Financial Details, Documentation, Technical Characteristics, and Crypto Dynamics.

Feature Selection Using LGBM

The proposed model adopts LGBM for feature selection, utilizing its efficient gradient boosting framework to balance computational efficiency with high performance. LGBM integrates GBDT techniques alongside innovations like Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS) to enhance training speed, reduce memory usage, and improve model accuracy (Mienye & Sun, 2022; Sarıkaya et al., 2022). GOSS reduces computational costs by prioritizing samples with high gradient values and excluding those with minimal contribution to the learning process, allowing for a more efficient estimation of information gain (González et al., 2020; Sarıkaya et al., 2022). EFB further optimizes efficiency by bundling nearly exclusive features in sparse feature spaces, reducing data dimensionality without compromising information (Sarıkaya et al., 2022).

LGBM is combined with the "SelectFromModel" technique to identify critical features for predicting post-ICO performance. This method serves as a meta estimator, supporting any model that provides feature importance metrics or coefficients. Features with importance scores exceeding a defined threshold are retained, while less significant features are discarded (Banga et al., 2023; Chandrashekar & Sahin, 2014). Further explanation and mathematical bases of the LGBM algorithm can be found in (Ke et al., 2017; Mienye & Sun, 2022; Wang & Wang, 2020).

The LGBM algorithm processes preprocessed ICO data as input. Through iterative training and validation subset partitioning, the model refines feature importance rankings to identify the most predictive variables. The final selection yields six key features demonstrating the highest predictive power for ICO outcomes: (1) Token Supply, (2) Total Amount Raised, (3) Crowdsale Amount Raised, (4) Team Token Allocation, (5) Investor Token Allocation, and (6) Bitcoin Return.

Prediction Using Extra Trees Regression

The proposed model employs a fine-tuned Extra-Trees regression to predict post-ICO performance. Extra-Trees algorithm is an ensemble technique that extends Random Forest by introducing additional randomness, enhancing model diversity and reducing overfitting (González et al., 2020). The algorithm builds multiple decision trees using random feature subsets and split thresholds, aggregating predictions via averaging (regression) or majority voting (classification) (González et al., 2020; Schmid et al., 2023). Key inputs include the training dataset, number of trees, maximum depth, random split points, minimum leaf samples, and features per split. Outputs include predictions and feature importance scores. Extra-Trees ability to reduce variance and overfitting while maintaining high predictive accuracy makes it an excellent choice for tasks requiring reliable forecasting, such as predicting post-ICO performance. However, the basic Extra-Trees algorithm lacks systematic hyperparameter tuning, which is addressed by the proposed Fine-Tuned Extra Trees algorithm using Optuna Optimization.

Optuna Optimization for Extra Trees

The performance of machine learning models heavily depends on the configuration of their hyperparameters. Hyperparameter optimization is a critical process that can significantly enhance the predictive capability of models. Researchers have traditionally employed methods such as Grid Search and optimization algorithms like Stochastic Gradient Descent (SGD) and Adam for tuning hyperparameters. However, as model complexity increases, these methods often become computationally inefficient, especially in high-dimensional parameter spaces. To address these challenges, advanced frameworks like Optuna have been developed. Optuna is a state-of-the-art hyperparameter optimization framework that employs adaptive sampling strategies and efficient optimization algorithms to quickly identify the best hyperparameter combinations (Akiba et al., 2019). Its ability to dynamically navigate the hyperparameter space makes it particularly suitable for optimizing tree-based models such as Extra Trees and XGBoost (Lai et al., 2024), which are critical in financial applications such as post-ICO performance forecasting.

The advantages of Optuna over conventional hyperparameter tuning methods are manifold. Unlike Grid Search or Random Search, which require exhaustive evaluation of hyperparameter combinations, Optuna uses techniques like Tree-structured Parzen Estimator (TPE) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to focus on promising regions of the search space (Akiba et al., 2019; Hadianti & Kodri, 2023). This adaptive sampling approach not only reduces computational time but also improves the likelihood of finding optimal hyperparameters. Additionally, Optuna incorporates an efficient pruning mechanism that terminates unpromising trials early, further conserving resources (Srinivas & Katarya, 2022). Indeed, Optuna has shown superior performance in various applications, including network intrusion detection systems (Parekh et al., 2024) and geophysical deep learning problems (Almarzooq & bin Waheed, 2024). Compared to traditional methods like grid search, Optuna provides quantifiable results to justify hyperparameter choices and achieves better model performance (Almarzooq & bin Waheed, 2024). In a comparative study of hyperparameter optimization tools, Optuna demonstrated better performance for combined algorithm selection and hyperparameter optimization (CASH) problems, while HyperOpt excelled in multilayer perceptron architecture selection (Shekhar et al., 2021). These studies highlight the importance of automated hyperparameter optimization in improving model accuracy, efficiency, and adaptability across various domains, including cybersecurity and geophysics (Almarzooq & bin Waheed, 2024; Parekh et al., 2024). The obtained optimized parameters with Optuna are shown in Table 2.

HYPERPARAMETER	VALUE
1- NUMBER OF TREES	100
2- MAXIMUM DEPTH OF THE TREE	12
3- MINIMAL NUMBER OF SAMPLES NEEDED TO SPLIT AN INTERNAL NODE	0.04725
4- MINIMUM NUMBER OF SAMPLES REQUIRED TO BE AT A LEAF NODE	3
5- CRITERION (THE FUNCTION TO MEASURE THE QUALITY OF A SPLIT)	absolute_error
6- MINIMUM IMPURITY DECREASE THRESHOLD	0.000158

Table 2. Tuned Parameters of Extra Trees

Experimental Results

The analysis begins by benchmarking the proposed model's predictive performance against various regression approaches using multiple evaluation metrics. Following this comparative assessment, we validate the importance of the selected features to enhance their practical utility for investment decisions. The section concludes with model interpretation through SHapley Additive exPlanations (SHAP) and partial dependence plots, offering more profound insights of the model’s results.

Experimental Setup

To demonstrate the efficiency of the suggested hybrid model in forecasting post-ICO performance, we compare it with its components, statistical models, and several state-of-the-art ML algorithms. Using test data, we calculated four metrics to assess each model's performance: 1) R-squared, 2) Mean Squared Error (MSE), 3) Mean Absolute Error (MAE), and 4) Mean Absolute Percentage Error (MAPE), as shown in (Equations 1-4) (Erdebilli & Devrim-İçtenbaş, 2022; Nguyen et al., 2021).

$$R^2 = 1 - \frac{\sum_{m=1}^{\mathcal{M}} (y_m - \hat{y}_m)^2}{\sum_{m=1}^{\mathcal{M}} (y_m - \bar{\hat{y}}_m)^2} \quad (1)$$

$$MSE = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} (y_m - \hat{y}_m)^2 \quad (2)$$

$$MAE = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} |y_m - \hat{y}_m| \quad (3)$$

$$MAPE = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \frac{|y_m - \hat{y}_m|}{y_m} \times 100 \quad (4)$$

Where, y_m is the actual observed value for observation m, $\hat{y}_m$ is the estimated value for observation m, $\bar{\hat{y}}_m$ is the mean of estimated values, $\mathcal{M}$ is the total number of observations. Higher R-squared models are more effective in terms of prediction power as they show a stronger capacity to account for the variability in the observed data (Erdebilli & Devrim-İçtenbaş, 2022). For training and evaluation, the dataset is divided into 70% training and 30% testing.

Analysis of Experimental Results

To evaluate the suggested hybrid model's effectiveness in forecasting post-ICO performance, its performance is compared against several state-of-the-art algorithms. These algorithms are categorized into two groups: (1) Linear Regression Models, and Individual ML Models, and (2) Ensemble Models.

Performance Comparison Against Linear and Individual ML Models

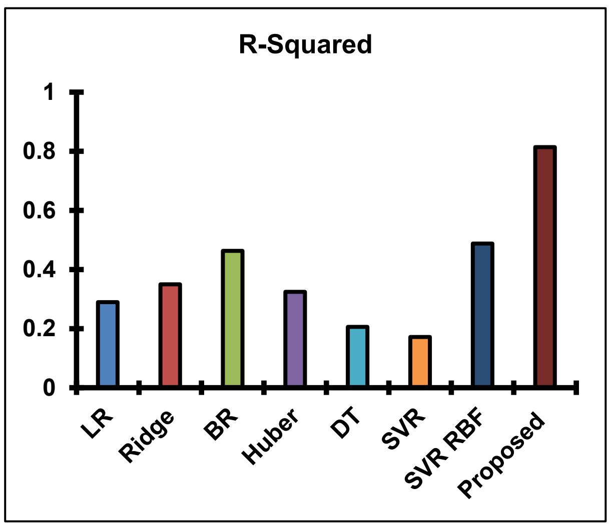
This subsection evaluates the proposed hybrid model, which combines LGBM for feature selection and fine-tuned Extra Trees for regression, against linear regression techniques, and Individual ML Models. The comparison is based on predicting post-ICO performance over 270 days, with performance measured using R^2, MSE, MAE, and MAPE.

As shown in Table 3 and Figure 3, the proposed hybrid model demonstrates superior predictive accuracy compared to both traditional regression methods and individual ML algorithms. The model achieves an R^2 of 0.814, significantly outperforming the best linear model (Bayesian Ridge at 0.463) and the top standalone ML approach (SVR RBF at 0.4877). Error metrics further reinforce this advantage: the hybrid model yields the lowest MSE (0.005 vs. 0.01459 for BR and 0.0139 for SVR RBF), MAE (0.051 vs. 0.1074 for LR and 0.10247 for DT), and MAPE (1.15

vs. 2.1486 for BR and 2.1784 for SVR RBF). These results consistently position the hybrid approach as the most effective solution for ICO performance forecasting, with Bayesian Ridge and SVR RBF emerging as the top-performing alternatives among linear and ML models, respectively.

Model	R-squared	MSE	MAE	MAPE
LR	0.2895	0.0193	0.1074	3.1429
Ridge	0.3497	0.0177	0.1329	2.6977
BR	0.463	0.01459	0.121	2.1486
Huber	0.3243	0.0184	0.1355	2.8742
DT	0.2058	0.0216	0.10247	2.69827
SVR Linear	0.1716	0.0225	0.15	3.063
SVR RBF	0.4877	0.0139	0.118	2.1784
Proposed Model	0.814	0.005	0.051	1.15

Table 3. Results of the Proposed Model in Comparison with Linear Regression-based Models



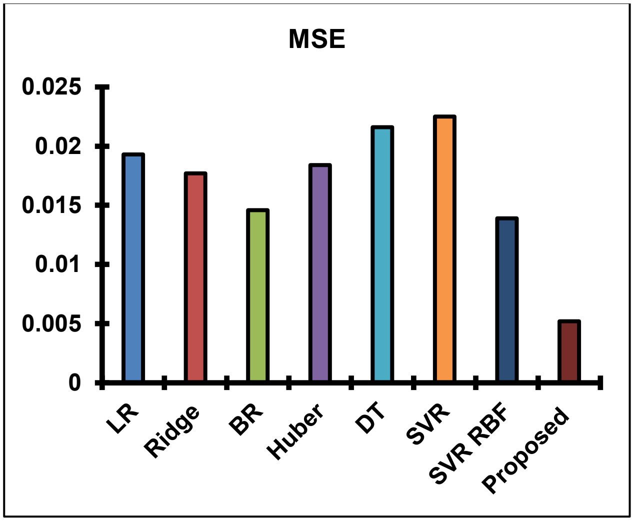


Figure 3. Bar Plots Showing the R-Squared, And MSE Results for Linear Regression Models and Individual ML Algorithms Against the Proposed Model for Predicting Performance 270 Days after ICO

Performance Comparison Against Ensemble Models

This subsection evaluates the proposed hybrid model, which combines LGBM for feature selection and fine-tuned Extra Trees for regression, against several ensemble models. These models include Random Forest (RF), Gradient Boosting Regressor (GBR), LGBM, Adaptive Boosting (ADA), Extreme Gradient Boosting (XGB), Categorical boosting (CAT), and Extra Trees (ET). The comparison is based on predicting post-ICO performance over 270 days, with performance measured using R^2, MSE, MAE, and MAPE.

As depicted in Table 4 and Figure 4, the proposed model reveals exceptional predictive power in comparison with various ensemble models in forecasting ICO returns. Although some ensemble models perform well, the hybrid model consistently outperforms them across all metrics. With an R^2 of 0.814, the model demonstrates superior explanatory power compared to alternative ensemble approaches, exceeding RF (0.7146) and ET (0.6942) by significant margins. Similarly, the hybrid model achieves the lowest MSE of 0.005, outperforming RF, the best-performing ensemble in this metric, which records an MSE of 0.0078.

The proposed model further establishes its superiority in MAE, achieving a value of 0.051, outperforming RF and CAT, which both achieve an MAE of 0.068. Most notably, the hybrid model records the lowest MAPE of 1.15, exceeding the performance of LGBM, which achieves the lowest MAPE among ensembles at 1.7. Overall, while ensemble models, particularly RF, ET, and CAT, deliver strong performance in predicting ICO returns, the proposed hybrid model surpasses all considered ensembles across the evaluated metrics, solidifying its position as the most effective model in this comparison.

Model	**R-squared**	**MSE**	**MAE**	**MAPE**
RF	0.7146	0.0078	0.068	1.7945
GBR	0.6634	0.0091	0.0713	1.8914
LGBM	0.5576	0.012	0.0823	1.7
ADA	0.5297	0.0128	0.1131	3.959

XGB	0.6371	0.00986	0.0993	1.7175
CAT	0.6919	0.0084	0.068	1.7579
ET	0.6942	0.0083	0.0691	1.849
Proposed Model	0.814	0.005	0.051	1.15

Table 4. Results of the Proposed Model in Comparison with Ensemble Models

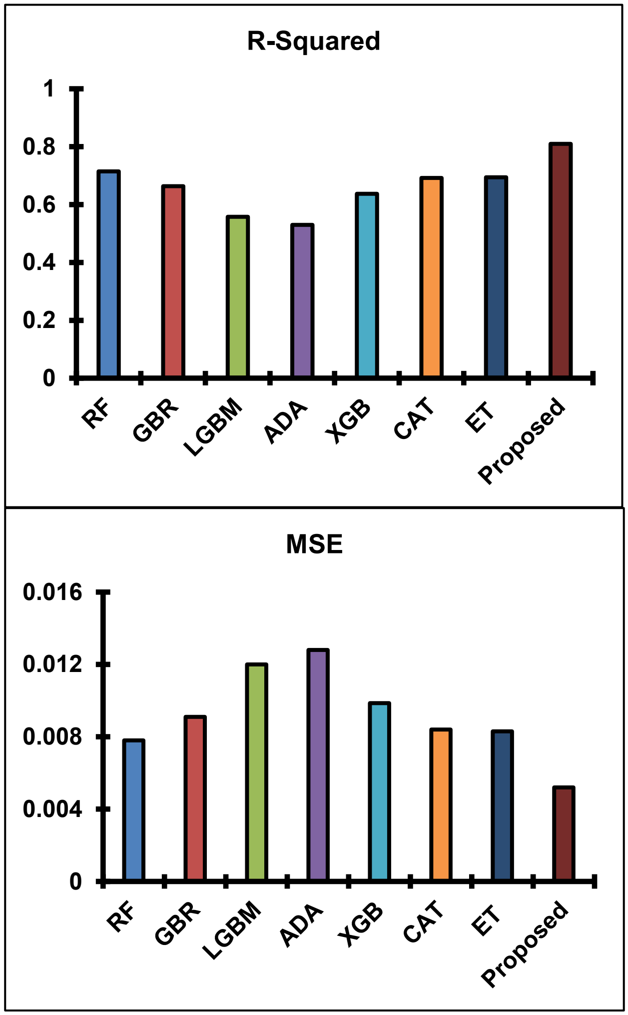


Figure 4. Bar Plots Showing the R^2, and MSE Results for Ensemble Models and the Proposed Model in Forecasting Returns 270 Days after ICO

In summary, the experimental findings reveal that the suggested hybrid model achieves higher R^2 values and smaller error metrics in comparison with all other models considered in this study. This confirms its superior performance in forecasting ICO returns, offering a robust and accurate framework for post-ICO performance analysis.

Feature Importance Analysis

Making sound investment decisions in the ICO marketplace requires an understanding of the critical signals for forecasting ICO returns. In contemporary applications, tree-based models represent the most popular class of nonlinear methods (Ozturkkal & Wahlstrøm, 2024). The fine-tuned Extra Trees algorithm was employed in this study to evaluate the importance of chosen variables.

As illustrated in Figure 5, the most significant factor is the supply of tokens, or the total number of tokens, which exhibits the highest predictive power at 63%. This is followed by the total amount raised in US dollars, contributing 15% to the model's predictive power. Other noteworthy variables, such as Bitcoin return and team token share, account for 8% and 5%, respectively. These findings highlight the importance of token characteristics, ICO financial details, team skin in the game, as well as the dynamics of the broader cryptocurrency market. This comprehensive analysis offers insightful information for investors aiming to take well-informed decisions in the ICO market. However, the feature importance plot may yield inconsistent results because it is derived from subsamples rather than the entire dataset (Altmann et al., 2010). To improve interpretation reliability and address this limitation, we employ Shapley values to analyze the same predictions while revealing the directionality of relationships.

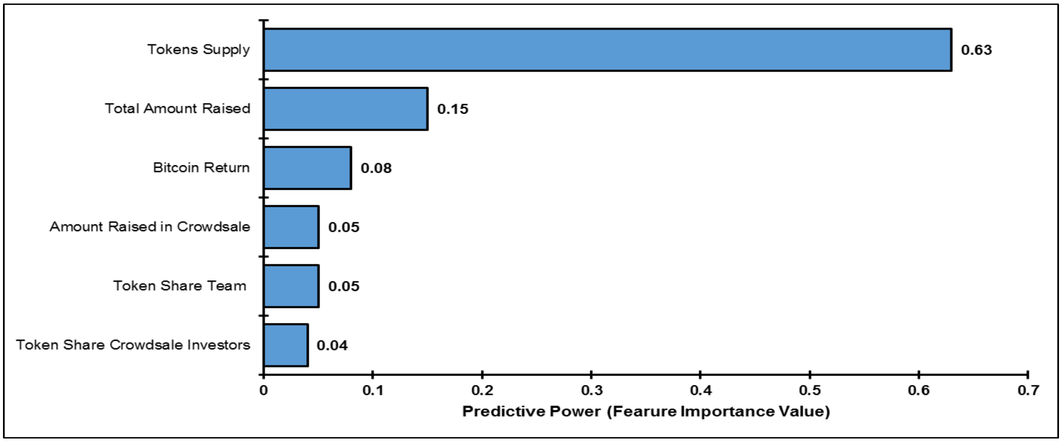


Figure 5. Relative Feature Importance using the Fine-tuned Extra Trees

Model Interpretation

SHapley Additive exPlanations (SHAP) is an explainability technique that quantifies feature contributions to machine learning predictions using Shapley values from cooperative game theory. The method employs a linear combination of binary variables to represent feature-prediction relationships, calculating each feature's impact by evaluating all possible subsets. Particularly effective for structured data with limited features, SHAP provides both global (dataset-level) and local (instance-level) interpretability, with global importance determined by averaging absolute SHAP values per feature (Guijarro et al., 2022). For this study, we implement TreeSHAP—a computationally efficient, exact interpretation method specialized for tree-based models. Unlike traditional approaches (e.g., permutation importance) that measure feature significance through the effect of feature removal on model performance, TreeSHAP directly quantifies attribution magnitudes, offering a more nuanced interpretation.

Figure 6 demonstrates that the key predictive features are consistent across both methods: Token Supply, Total Amount Raised, and Bitcoin Return. A comparison of the top five variables in Figure 5 (Extra Trees feature importance) and Figure 6 reveals that they are nearly identical, with only slight differences in ranking.

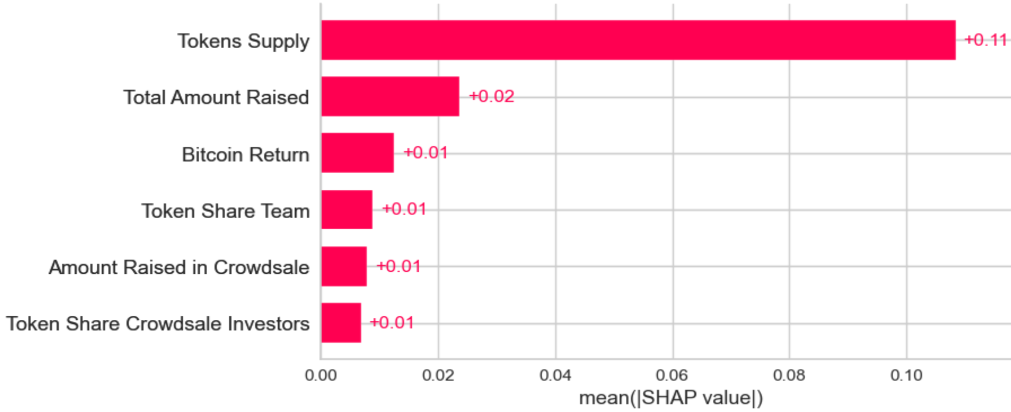


Figure 6. The Feature Importance of the Proposed Model using Shap values

The SHAP values can be further visualized in a summary plot (Figure 7), which ranks variables by their impact on ICO return predictions—from most influential (top) to least (bottom). Each dot represents an individual observation, where right-side positioning indicates a positive effect on returns and left-side a negative effect, while color reflects magnitude—blue for low values and red for high values of the variable. Figures 7 demonstrate that token supply emerges as the strongest negative predictor of ICO returns. Figure 8 further demonstrates the partial dependency plots for the top explanatory variables. Specifically, the first dependency plot in figure 8 shows a clear negative relationship between the logarithm of token supply and expected ICO returns. As log-transformed token supply increases (from approximately 6.5 to 10), expected returns decline sharply, with higher returns concentrated at the lower end of the supply range. This suggests that projects with lower effective token supply benefit from scarcity-driven investor interest. Beyond a token supply of about 8.5, returns flatten near zero, indicating diminishing returns as supply increases. The histogram reveals that most observations fall between log supply values of 7.5 and 9.5—an area associated with lower returns—highlighting the predictive value of lower token supply for stronger post-ICO performance.

This finding aligns with established tokenomics research (Howell et al., 2020; Lyandres et al., 2022), where larger token supplies correlate with diminished returns. The observed relationship likely stems from two key mechanisms: (1) dilution effects from excessive token issuance, and (2) investor valuation of scarcity (Catalini & Gans, 2016). Specifically, Howell et al. (2020) demonstrate that projects with larger token supplies experience lower returns due to investor concerns about future dilution and inflationary pressures. Further, Catalini & Gans (2020) explain how token supply affects valuation through scarcity as larger supplies dilute perceived value, aligning with traditional economic principles of supply-demand equilibrium.

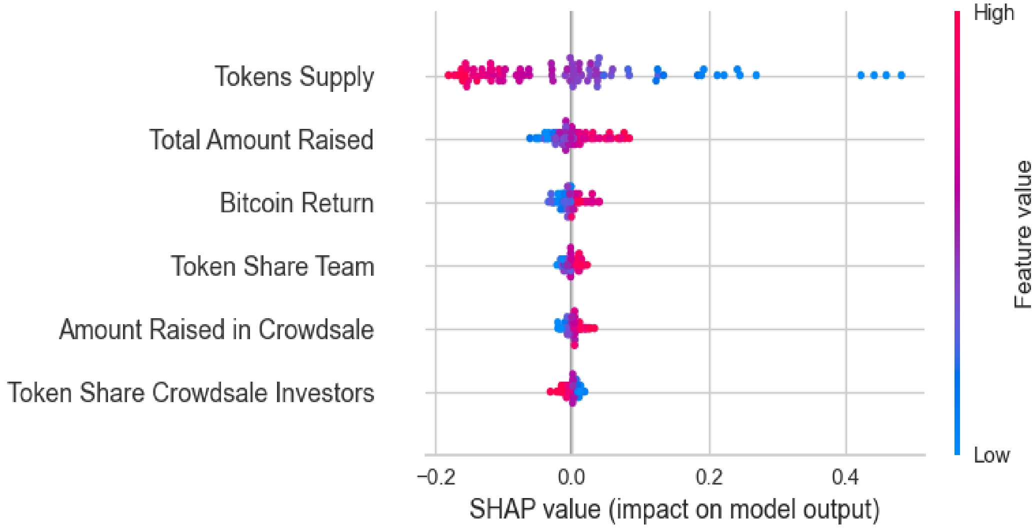
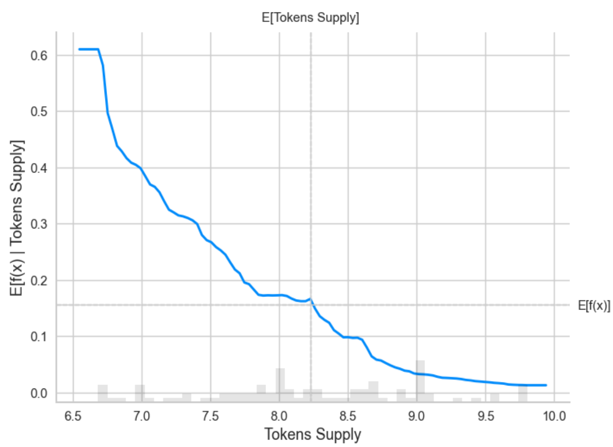
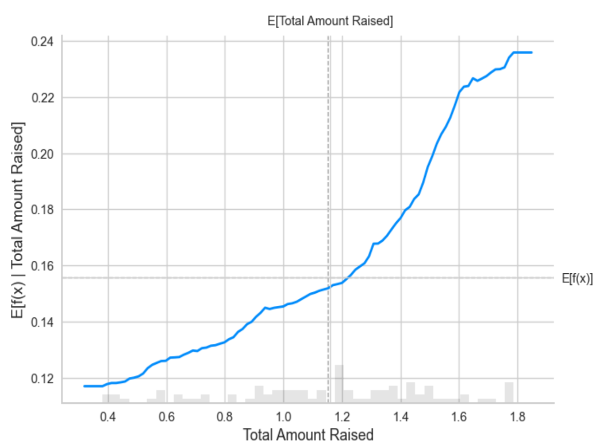


Figure 7. Global Interpretability Plots of the Proposed Model





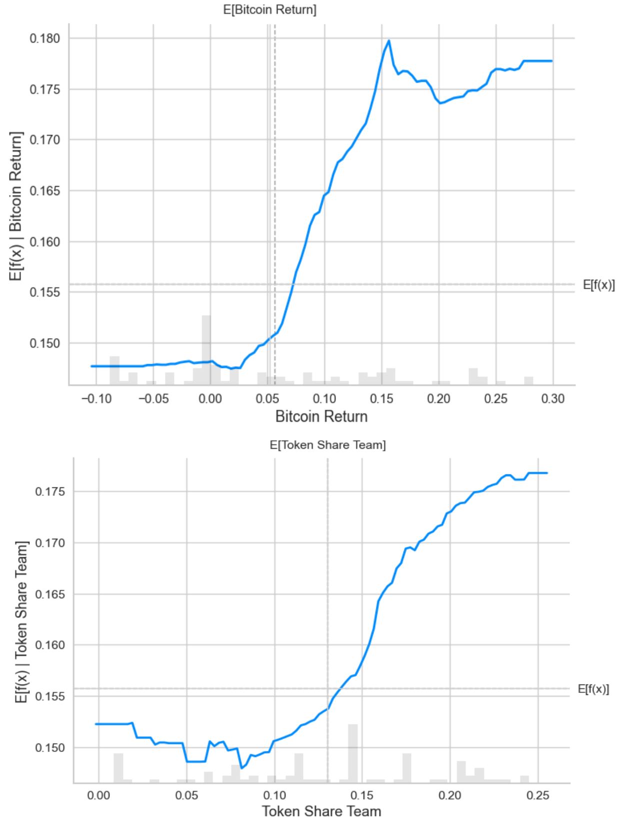


Figure 8. Partial Dependency Plots for the Top Four Predictive Variables

The second plot in figure 8 shows a positive relationship between the logarithm of the total amount raised and ICO returns. As the total amount raised increases, expected returns rise steadily. This suggests that projects raising more capital tend to deliver higher predicted returns. Reference lines mark the mean log amount raised and overall expected return, highlighting that returns exceed average levels when fundraising is above the mean. The histogram shows most projects cluster between log values of 1.0 and 1.4, aligning with moderately higher returns. This result is consistent with findings of Fahlenbrach & Frattaroli (2021), supporting the idea that strong fundraising signals project viability and boosts investor confidence.

Furthermore, figure 7 highlights the positive influence of Bitcoin returns on post-ICO financial success, underscoring the interconnectedness of ICOs with broader cryptocurrency market trends. Specifically, the third plot in figure 8 shows a nonlinear, positive relationship between Bitcoin returns and predicted ICO returns, with all continuous variables log-transformed. As Bitcoin returns increase from negative values to around 0.10, expected ICO returns remain relatively flat near 0.15. Beyond this threshold, returns rise sharply, peaking around a Bitcoin return of 0.15, then stabilizing just below the peak. This suggests that strong Bitcoin performance significantly boosts expected ICO returns, likely driven by enhanced investor sentiment and market optimism (Domingo et al., 2020; Lyandres et al., 2022). The histogram shows that most Bitcoin returns cluster around zero, emphasizing that substantial increases in ICO returns tend to occur during bullish Bitcoin periods.

The last plot in figure 8 illustrates the relationship between the percentage of token supply held by team members (Token Share Team) and predicted ICO returns. When the team holds a very low share (below ~10%), predicted returns remain flat or slightly below average, suggesting limited commitment or investor confidence. As the share increases beyond this threshold, predicted returns rise sharply, indicating that moderate ownership (10%–20%) aligns incentives and boosts trust in the project. Beyond ~20%, the curve plateaus, implying diminishing returns or concerns about centralization. Overall, the plot suggests that a balanced level of team token ownership is perceived most favorably in predicting ICO returns. These findings align with the notion that "skin in the game" (proxied by retained tokens) is crucial for post-ICO success, as entrepreneurs with greater ownership stakes are less likely to shirk responsibilities. This supports the conclusions of Gan et al. (2020) and Lyandres et al. (2022), who link larger retained ownership to higher post-ICO outcomes. However, the results are inconsistent with Fahlenbrach & Frattaroli (2021) who find higher team token shares correlating negatively with ICO returns.

Conclusion

This study presents a novel hybrid machine learning framework that significantly advances the predictive modeling of Initial Coin Offering (ICO) performance, with important implications for smart city financing. By integrating Light Gradient Boosting Machine (LGBM) for feature selection and Optuna-optimized Extremely Randomized Trees (Extra Trees) for return prediction, our model achieves superior performance ($R^2 = 0.814$, MSE = 0.005, MAE = 0.051) compared to both traditional regression methods and state-of-the-art machine learning alternatives. The model's robust predictive capability stems from its ability to handle noisy financial data while mitigating overfitting through enhanced randomization and automated hyperparameter optimization.

Three key findings emerge from our analysis: First, token supply demonstrates the strongest predictive power (63%) but exhibits a negative relationship with returns, reflecting investor sensitivity to token dilution and scarcity effects. Second, successful fundraising (15% importance) serves as a positive signal of project viability, while Bitcoin returns (8%) highlight the interdependence between ICO performance and broader cryptocurrency market trends. Third, our interpretability analysis through SHAP values and partial dependence plots provides transparent insights into these relationships, addressing critical information asymmetries in the ICO market.

This study makes significant dual contributions to both academic research and real-world applications. Methodologically, it advances FinTech scholarship by demonstrating how optimized hybrid machine learning architectures can overcome the predictive limitations of traditional statistical and ML approaches in blockchain-based investments. The framework establishes new standards for analyzing decentralized financing instruments while maintaining computational efficiency and interpretability.

Practically, the research delivers actionable value across multiple stakeholder groups: (1) Investors gain a robust decision-support tool for evaluating ICO opportunities, (2) Regulators acquire evidence-based insights to develop policies that standardize disclosures and mitigate fraud risks, and (3) Urban developers receive empirical validation of ICOs as a viable smart city financing mechanism that combines blockchain transparency with decentralized capital formation efficiency. These applications align with broader smart city objectives by enabling data-driven, sustainable financing strategies for urban innovation.

This research, while advancing the understanding of ICO performance prediction, has some limitations that merit discussion. Specifically, the dataset comprises 292 ICOs, which, while robust, may not fully represent the diversity of the ICO market. Additionally, ICO performance can also be sensitive to investor sentiment, regulatory developments, and macroeconomic conditions. The model may not fully account for these factors. By addressing these aspects, future studies can build upon this research and extend its applicability to a broader range of ICOs. Nevertheless, this study establishes an important foundation for data-driven approaches to sustainable urban development financing in the digital age, bridging the gap between cutting-edge machine learning techniques and real-world financial decision-making in smart city ecosystems.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University, Saudi Arabia for funding this research work through the project number (PSAU/2024/02/31560).

Data Availability

Data is available from the corresponding author upon request.

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

Ahmad, M. F., Kowalewski, O., & Pisany, P. (2023). What determines initial coin offering success: a cross-country study. Economics of Innovation and New Technology, 32(5), 622–645. https://doi.org/10.1080/10438599.2021.1982712

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. Expert Systems with Applications, 94, 164–184. https://doi.org/10.1016/J.ESWA.2017.10.040

Ali, G. M., Gomaa, I. I., & Darwish, S. M. (2022). An Intelligent Model for Success Prediction of Initial Coin Offerings. IEEE Access, 10, 58589–58602. https://doi.org/10.1109/ACCESS.2022.3178369

Almarzooq, H., & bin Waheed, U. (2024). Automating hyperparameter optimization in geophysics with Optuna: A comparative study. Geophysical Prospecting, 72(5), 1778–1788. https://doi.org/10.1111/1365-2478.13484

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340–1347.

Amsden, R., & Schweizer, D. (2018). Are Blockchain Crowdsales the New "Gold Rush"? Success Determinants of Initial Coin Offerings. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3163849

Aslan, A., Şensoy, A., & Akdeniz, L. (2023). Determinants of ICO success and post-ICO performance. Borsa Istanbul Review, 23(1), 217–239. https://doi.org/https://doi.org/10.1016/j.bir.2022.10.005

Banga, A., Ahuja, R., & Sharma, S. C. (2023). Performance analysis of regression algorithms and feature selection techniques to predict PM2.5 in smart cities. International Journal of System Assurance Engineering and Management, 14(3), 732–745. https://doi.org/10.1007/S13198-020-01049-9/METRICS

Benedetti, H., & Kostovetsky, L. (2021). Digital Tulips? Returns to investors in initial coin offerings. Journal of Corporate Finance, 66, 101786. https://doi.org/10.1016/J.JCORPFIN.2020.101786
Blanck, M., & Ribeiro, J. L. D. (2021). Smart cities financing system: An empirical modelling from the European context. Cities, 116, 103268. https://doi.org/10.1016/J.CITIES.2021.103268
Blaseg, D. (2018). Dynamics of Voluntary Disclosure in the Unregulated Market for Initial Coin Offerings. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3207641
Boreiko, D., & Risteski, D. (2021). Serial and large investors in initial coin offerings. Small Business Economics, 57(2), 1053–1071. https://doi.org/https://doi.org/10.1007/s11187-020-00338-8
Bourveau, T., De George, E. T., Ellahie, A., & Macciocchi, D. (2022). The Role of Disclosure and Information Intermediaries in an Unregulated Capital Market: Evidence from Initial Coin Offerings. Journal of Accounting Research, 60(1), 129–167. https://doi.org/10.1111/1475-679X.12404
Burns, L., & Moro, A. (2018). What Makes an ICO Successful? An Investigation of the Role of ICO Characteristics, Team Quality and Market Sentiment. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3256512
Catalini, C., & Gans, J. S. (2018). Initial coin offerings and the value of crypto tokens. National Bureau of Economic Research.
Catalini, C., & Gans, J. S. (2020). Some simple economics of the blockchain. Communications of the ACM, 63(7), 80–90. https://doi.org/10.1145/3359552/SUPPL_FILE/P80-CATALINI-SUPP.PDF
Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16–28. https://doi.org/10.1016/J.COMPELECENG.2013.11.024
Chen, K. (2019). Information asymmetry in initial coin offerings (ICOs): Investigating the effects of multiple channel signals. Electronic Commerce Research and Applications, 36, 100858. https://doi.org/10.1016/J.ELERAP.2019.100858
Davydiuk, T., Gupta, D., & Rosen, S. (2023). De-Crypto-ing Signals in Initial Coin Offerings: Evidence of Rational Token Retention. Management Science, 69(11), 6584–6624. https://doi.org/10.1287/MNSC.2022.4631
Dittmar, R. F., & Wu, D. (Andrew). (2019). Initial Coin Offerings Hyped and Dehyped: An Empirical Examination. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.3259182
Domingo, R. S., Piñeiro-Chousa, J., & Ángeles López-Cabarcos, M. (2020). What factors drive returns on initial coin offerings? Technological Forecasting and Social Change, 153, 119915. https://doi.org/10.1016/J.TECHFORE.2020.119915
Erdebilli, B., & Devrim-İçtenbaş, B. (2022). Ensemble Voting Regression Based on Machine Learning for Predicting Medical Waste: A Case from Turkey. Mathematics 2022, Vol. 10, Page 2466, 10(14), 2466. https://doi.org/10.3390/MATH10142466
Fahlenbrach, R., & Frattaroli, M. (2021). ICO investors. Financial Markets and Portfolio Management, 35(1), 1–59. https://doi.org/10.1007/S11408-020-00366-0/TABLES/11
Felix, T. H., & von Eije, H. (2019). Underpricing in the cryptocurrency world: evidence from initial coin offerings. Managerial Finance, 45(4), 563–578. https://doi.org/10.1108/MF-06-2018-0281/FULL/PDF
Fisch, C. (2019). Initial coin offerings (ICOs) to finance new ventures. Journal of Business Venturing, 34(1), 1–22. https://doi.org/10.1016/J.JBUSVENT.2018.09.007
Fisch, C., & Momtaz, P. P. (2020). Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). Journal of Corporate Finance, 64, 101679. https://doi.org/10.1016/J.JCORPFIN.2020.101679
Florysiak, D., & Schandlbauer, A. (2022). Experts or charlatans? ICO analysts and white paper informativeness. Journal of Banking & Finance, 139, 106476.

https://doi.org/10.1016/J.JBANKFIN.2022.106476

Fuchs, J., & Momtaz, P. P. (2024). Token Governance in Initial Coin Offerings: Implications of Token Retention and Resale Restrictions for ICO Success. Small Business Economics, 1–39. https://doi.org/https://doi.org/10.1007/s11187-024-00945-9

Gan, J., Tsoukalas, G., & Netessine, S. (2020). Initial Coin Offerings, Speculation, and Asset Tokenization. Https://Doi.Org/10.1287/Mnsc.2020.3796, 67(2), 914–931. https://doi.org/10.1287/MNSC.2020.3796

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63(1), 3–42. https://doi.org/10.1007/S10994-006-6226-1/METRICS

Giudici, G., & Adhami, S. (2019). The impact of governance signals on ICO fundraising success. Journal of Industrial and Business Economics, 46(2), 283–312. https://doi.org/10.1007/S40812-019-00118-W/METRICS

González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Information Fusion, 64, 205–237. https://doi.org/10.1016/J.INFFUS.2020.07.007

Guijarro, F., Long Tran, K., Anh Le, H., Hien Nguyen, T., Trung Nguyen, D., Thai Binh Ward, N., & Chi Minh City, H. (2022). Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. Data 2022, Vol. 7, Page 160, 7(11), 160. https://doi.org/10.3390/DATA7110160

Hadianti, S., & Kodri, W. A. G. (2023). Optimization of the machine learning approach using Optuna in heart disease prediction. Journal Medical Informatics Technology, 59–64.

Hedegaard, M. (2024). ’Smart City Projects Financing. SocioEconomic Challenges, 8(1), 2024. https://doi.org/10.61093/sec.8(1).286-309.2024

Hollands, R. G. (2020). Will the real smart city please stand up? : Intelligent, progressive or entrepreneurial? The Routledge Companion to Smart Cities, 179–199. https://doi.org/10.4324/9781315178387-13

Howell, S. T., Niessner, M., & Yermack, D. (2020). Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. The Review of Financial Studies, 33(9), 3925–3974. https://doi.org/10.1093/RFS/HHZ131

Hsieh, H. C., & Oppermann, J. (2021). Initial coin offerings and their initial returns. Asia Pacific Management Review, 26(1), 1–10. https://doi.org/10.1016/J.APMRV.2020.05.003

Iansiti, M., & Lakhani, K. R. (2017). The truth about blockchain. Harvard Business Review, 95(1), 118–127.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (n.d.). Lightgbm: A highly efficient gradient boosting decision tree. Proceedings.Neurips.CcG Ke, Q Meng, T Finley, T Wang, W Chen, W Ma, Q Ye, TY LiuAdvances in Neural Information Processing Systems, 2017•proceedings.Neurips.Cc. Retrieved September 8, 2024, from https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, 30. https://github.com/Microsoft/LightGBM.

Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. https://doi.org/10.4135/9781473909472

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. Korean Journal of Anesthesiology, 70(4), 407. https://doi.org/10.4097/KJAE.2017.70.4.407

Lai, L. H., Lin, Y. L., Liu, Y. H., Lai, J. P., Yang, W. C., Hou, H. P., & Pai, P. F. (2024). The Use of Machine Learning Models with Optuna in Disease Prediction. Electronics (Switzerland), 13(23). https://doi.org/10.3390/electronics13234775

Lai, L.-H. ;, Lin, Y.-L. ;, Liu, Y.-H. ;, Lai, J.-P. ;, Yang, W.-C. ;, Hou, H.-P. ;, Pai, P.-F., Lai, L.-H., Lin, Y.-L., Liu, Y.-H., Lai, J.-P., Yang, W.-C., Hou, H.-P., & Pai, P.-F. (2024). The Use of Machine Learning Models with Optuna in Disease Prediction. Electronics 2024, Vol. 13, Page 4775, 13(23), 4775. https://doi.org/10.3390/ELECTRONICS13234775

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lyandres, E., Palazzo, B., & Rabetti, D. (2022). Initial Coin Offering (ICO) Success and Post-ICO Performance. Management Science, 68(12), 8658–8679. https://doi.org/10.1287/MNSC.2022.4312

Majidi, S. H., Hadayeghparast, S., & Karimipour, H. (2022). FDI attack detection using extra trees algorithm and deep learning algorithm-autoencoder in smart grid. International Journal of Critical Infrastructure Protection, 37, 100508. https://doi.org/10.1016/J.IJCIP.2022.100508

McCarty, D. A., Kim, H. W., & Lee, H. K. (2020). Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification. Environments 2020, Vol. 7, Page 84, 7(10), 84. https://doi.org/10.3390/ENVIRONMENTS7100084

Meoli, M., & Vismara, S. (2022). Machine-learning forecasting of successful ICOs. Journal of Economics and Business, 121, 106071. https://doi.org/10.1016/J.JECONBUS.2022.106071

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 10, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

Mirzaee, A. M., & Sardroud, J. M. (2022). Public–private-partnerships (PPP) enabled smart city funding and financing. Smart Cities Policies and Financing: Approaches and Solutions, 117–131. https://doi.org/10.1016/B978-0-12-819130-9.00011-5

Momtaz, P. P. (2020). Initial Coin Offerings. PLOS ONE, 15(5), e0233018. https://doi.org/10.1371/JOURNAL.PONE.0233018

Momtaz, P. P. (2021). Entrepreneurial Finance and Moral Hazard: Evidence from Token Offerings. Journal of Business Venturing, 36(5), 106001. https://doi.org/10.1016/J.JBUSVENT.2020.106001

Nguyen, X. C., Nguyen, T. T. H., La, D. D., Kumar, G., Rene, E. R., Nguyen, D. D., Chang, S. W., Chung, W. J., Nguyen, X. H., & Nguyen, V. K. (2021). Development of machine learning - based models to forecast solid waste generation in residential areas: A case study from Vietnam. Resources, Conservation and Recycling, 167, 105381. https://doi.org/10.1016/J.RESCONREC.2020.105381

Ozturkkal, B., & Wahlstrøm, R. R. (2024). Explaining Mortgage Defaults Using SHAP and LASSO. Computational Economics. https://doi.org/10.1007/s10614-024-10763-6

Parekh, N., Sen, A., Rajasekaran, P., Jayaseeli, J. D. D., & Robert, P. (2024). Network Intrusion Detection System Using Optuna. 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), 312–318.

Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. Neurocomputing, 239, 39–57. https://doi.org/10.1016/J.NEUCOM.2017.01.078

Roosenboom, P., van der Kolk, T., & de Jong, A. (2020). What determines success in initial coin offerings? Venture Capital, 22(2), 161–183. https://doi.org/10.1080/13691066.2020.1741127

Sarıkaya, A., Günel Kılıç, B., & Demirci, M. (2022). GRU-GBM: A combined intrusion detection model using LightGBM and gated recurrent unit. Expert Systems, 39(9), e13067. https://doi.org/10.1111/EXSY.13067

Schmid, L., Gerharz, A., Groll, A., & Pauly, M. (2023). Tree-based ensembles for multi-output regression: Comparing multivariate approaches with separate univariate ones. Computational Statistics & Data Analysis, 179, 107628. https://doi.org/10.1016/J.CSDA.2022.107628

Shekhar, S., Bansode, A., & Salim, A. (2021). A comparative study of hyper-parameter optimization tools. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 1–6.

Shrestha, P., Arslan-Ayaydin, Ö., Thewissen, J., & Torsin, W. (2021). Institutions, regulations and initial coin offerings: An international perspective. International Review of Economics & Finance, 72, 102–120. https://doi.org/10.1016/J.IREF.2020.10.014

Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3), 355–374. https://doi.org/10.2307/1882010

Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. Biomedical Signal Processing and Control, 73, 103456.

Stanley, M. (2019). The Application of Behavioural Heuristics to Initial Coin Offerings Valuation and Investment. The Journal of the British Blockchain Association, 2(1), 1–7. https://doi.org/10.31585/JBBA-2-1-(7)2019

Wang, J., Chen, R., Xu, W., Tang, Y., & Qin, Y. (2022). A document analysis deep learning regression model for initial coin offerings success prediction. Expert Systems with Applications, 210, 118367. https://doi.org/10.1016/J.ESWA.2022.118367

Wang, Y., & Wang, T. (2020). Application of Improved LightGBM Model in Blood Glucose Prediction. Applied Sciences 2020, Vol. 10, Page 3227, 10(9), 3227. https://doi.org/10.3390/APP10093227

Xu, W., Wang, T., Chen, R., & Zhao, J. L. (2021). Prediction of initial coin offering success based on team knowledge and expert evaluation. Decision Support Systems, 147, 113574. https://doi.org/10.1016/J.DSS.2021.113574

Yen, J. C., Wang, T., & Chen, Y. H. (2021). Different is better: how unique initial coin offering language in white papers enhances success. Accounting & Finance, 61(4), 5309–5340. https://doi.org/10.1111/ACFI.12760

Zhang, S., Aerts, W., Lu, L., & Pan, H. (2019). Readability of token whitepaper and ICO first-day return. Economics Letters, 180, 58–61. https://doi.org/10.1016/J.ECONLET.2019.04.010.