MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

Machine Learning Models for Predicting Private Equity Fund Performance Using Historical Deal-Level and Macroeconomic Data

Vrajkumar Patel
MBA student
Vrajpatel5471@gmail.com
0009000121620319

Abstract
The private equity sector has grown in significance in the international financial markets because of its capacity to deliver high-value long-term investment returns. But predicting the return of a private equity fund is extremely hard as a result of market volatility, the nonlinear relationship between investments and returns, and the opacity of alternative asset markets. This study aims to build a predictive model that can predict the performance of private equity funds based on data from the Private Equity Data Q1 2009 to Q2 2021 using machine learning techniques. The framework consists of data preprocessing, temporal feature engineering, ensemble learning, hyperparameter optimization, and explainable artificial intelligence (XAI) techniques that are executed in a Google Colab environment. The models evaluated comparatively were: Linear Regression, Random Forest, XGBoost and LightGBM, which were assessed based on the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2). The experimental results showed that the model XGBoost had the best predictive performance, with the parameters MAE of 0.091, RMSE of 0.126, MAPE of 5.96% and R2 value of 0.919, which is better than the other models, Random Forest, LightGBM and Linear Regression. SHAP-based explainability analysis revealed that the variables that most influenced Private Equity fund performance were investment growth (Q4), the deployment of capital in the previous four quarters (4Q) and transaction density. The results support the fact that ensemble learning algorithms are very useful for improving the forecasting power in the context of private equity space and that they deliver transparent AI-assisted investment decision support to private equity investors, portfolio managers and financial analysts in alternative investment contexts.

Keywords: Private Equity, Machine Learning, XGBoost, Financial Forecasting, Ensemble Learning, Explainable Artificial Intelligence, Predictive Analytics, Investment Performance, SHAP, Alternative Investments

Introduction

According to (Tausch & Pietz, 2024), the private equity (PE) industry has become one of the most significant players in the global financial landscape, and is the recipient of large sums of institutional investor, pension fund, sovereign wealth fund, insurance company, and high-net-worth individual capital. Private equity markets have grown very quickly in the past two decades because they have proven to be able to produce better long-term returns than traditional public equity markets (Ewens & Farre-Mensa, 2022). Private equity investments continue to expand on a global scale, bolstered by greater international mergers and acquisitions, and advances in financial technology, among other factors, in various healthcare, technology, energy, manufacturing and infrastructure industries (McGrath & Nerkar, 2023). With private equity firms controlling trillions of dollars in assets worldwide, the importance of precise forecasting and risk assessment for investors and fund managers is growing (Simlai, 2025), as they strive to make optimal investment choices and portfolio allocation decisions.

The private equity investment lifecycle generally includes three stages – acquisition, holding and exit (McGrath & Nerkar, 2023; Singh et al., 2024). In the acquisition phase, companies seek out other companies that have a significant amount of growth potential and incorporate them into their own ownership either through leveraged buyouts, venture capital investments or strategic acquisitions (McGrath & Nerkar, 2023). Once acquired, the holding period will feature operational restructuring, financial optimization and strategic growth plans that will benefit enterprise value over several years. The last phase is when the investors start to see returns on their invested capital by exit strategies like initial public offerings (IPOs), mergers and acquisitions (M&A) or secondary buyout (Singh et al., 2024). The performance of private equity funds is influenced by various market conditions, economic factors and management choices throughout the investment lifecycle, making it a complex and data-heavy task that relies on sophisticated analytical methods to make forecasts.

The performance of private equity funds is incredibly hard to predict because of the complexities and volatilities of the financial markets (Korteweg & Nagel, 2025). The market volatility, macroeconomic uncertainty, geopolitical concerns, inflation and monetary policy adjustments are all significant factors in determining the impact of private equity investments (Lamothe-Fernández et al., 2025; Simlai, 2025). Accporading to (Böni et al., 2022), private equity transactions are less transparent and financial information is less readily available than with publicly traded securities. This is because there are no standard and easily available data in place that could accurately assess the quality of the investment and predict future fund performance. Additionally, the economic environment, which can range from recession to boom, with varying interest rates and global financial events, can have a significant impact on the profit potential and exit position of private equity investments over time.

One of the other big hurdles is the nonlinear behavior of investment in private equity. Several factors, including leverage ratios, operational efficiency, market timing, and investor sentiment, all influence investment outcomes, which are, in turn, influenced by the performance of the sectors in which they invest (Huynh et al., 2025). According to (Simionescu, 2025), the traditional statistical and econometric forecasting methods are based on linear projections and relationships between variables, and do not fully account for the complex nonlinear relationships or risk patterns. Further, private equity investments exhibit high variability in terms of return and quality of the industry, geographic region and investment strategy (ONG, 2022). Traditional regression-based approaches often struggle to capture these multi-dimensional relationships and changing market dynamics, leading to less forecasting accuracy and limited generalizability. This has led to a growing demand for more sophisticated prediction methods designed to deal with high-dimensional financial data and to reveal the complex pattern of private equity investment environments.

Machine learning has proven to be a game-changing technology in financial forecasting, helping financial institutions and investment firms analyze massive amounts of information and make highly predictive insights (Rane et al., 2023b). Artificial Intelligence (AI) has revolutionized the field of financial analytics, making it easier to identify hidden trends, nonlinear relationships and complex dependencies within financial markets (Hossain, 2025; Taheri Hosseinkhani, 2025). Machine learning models can learn the complex patterns in the data by optimizing the model through an iterative process without relying on linear statistical relationships or pre-defined assumptions that are common in traditional econometric approaches (Ferraro, 2024). The ability to do this is where machine learning is well-suited for predicting tasks that involve volatile and multidimensional financial data, such as predicting the performance of private equity funds.

Over the past few years, predictive analytics based on machine learning has received significant focus in fields ranging from stock market prediction to investment strategy development, from portfolio optimization to credit risk management, and from fraud detection to financial risk management (Nahar et al., 2024). Ensemble learning techniques like Random Forest, XGBoost and LightGBM are the machine learning algorithms that use multiple weak learners to boost the prediction accuracy and avoid overfitting (Kumar et al., 2024). Decision tree-based algorithms are especially suitable for dealing with datasets that contain high-dimensional features, missing data, and have nonlinear data distributions, which are frequently found in financial data (Gong, 2025). As algorithms are improved, they become more accurate in predicting outcomes, as any error in the prediction is reduced and the performance of the model is optimized over numerous training cycles. Machine learning models offer greater flexibility, scalability, and forecasting accuracy than traditional econometric models, making them an increasingly powerful tool for complex financial forecasting applications (Okoh & Grace, 2022). Therefore, machine learning in private equity analytics holds great promise for enhancing investment decision-making and the accuracy of predictions in today's financial environment.

While the use of machine learning in financial analysis has accelerated, there is little research that has specifically examined the prediction of private equity performance based on historical data of investments per private equity deal. Private equity forecasting has not been widely studied because of the difficulty in accessing data and the intricacies of private investment structures, whereas forecasting in publicly traded assets like stock markets, cryptocurrencies, and banking risk prediction is comparatively well studied. Moreover, most previous research is based on either a macro-financial approach or a fund-level approach, which may not accurately reflect the investment details and timing of individual private equity investments.

The other challenge is the lack of explainability in the private equity prediction systems using machine learning (Khan et al., 2025). While some models can be very successful in predicting what happens, they are often black-box systems that give investors and financial analysts little insight (Petersone et al., 2022). This opacity makes it difficult to grasp the factors behind the success of investments and diminishes faith in the investment decision-making processes involving AI. Moreover, the performance of the aforementioned studies is often assessed specifically with one machine learning algorithm and does not involve comprehensive

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

testing and comparison between a variety of ensemble learning methods. As a result, there is a significant research gap towards developing an interpretable and comparable machine learning framework that accurately predicts private equity fund performance based on deal-level information from private equity investments. This study sets out to achieve two main objectives: First, to develop a machine learning framework to predict the performance of private equity funds using historical investment information, which is sourced from the "Private Equity Data Q1 2009 to Q2 2021" dataset, based on the level of these investments. Second, to evaluate the nature of the problem and identify the right tools to use for the development of the machine learning framework. This research aims to explore the efficacy of cutting-edge machine learning models in predicting investment returns and uncovering key financial trends in the private equity sector. The proposed framework seeks to incorporate explainable AI techniques and predictive analytics to improve the transparency of forecasting and facilitate better investment decision-making processes.

This goal is achieved through a series of objectives in the study. Firstly, a predictive data set will be created by preprocessing and feature engineering the historical private equity data. Second, the Linear Regression, Random Forest, XGBoost and LightGBM algorithms will be designed and compared. Thirdly, the accuracy of the prediction and generalization performance of each model will be evaluated by standard evaluation metrics. Fourth, feature importance analysis will be used to identify influential variables in investment that impact private equity performance. Lastly, explainable AI methods will be incorporated to boost the model's readability and build investor confidence in AI-powered forecasting systems.

This study is part of an emerging research stream in financial machine learning, focusing on the development of a private equity-specific predictive analytics model based on historical private equity deal information. The study offers a thorough comparison of ensemble learning methods such as Random Forest, XGBoost, and LightGBM for predicting private equity fund performance. The study also incorporates explainable artificial intelligence (XAI) methods, enhancing the transparency and interpretability of models in financial forecasting settings. The proposed framework also helps to support the investment decision-making process by highlighting the most influential variables that impact private equity performance and by showcasing the machine learning in the alternative asset management field's applicability. In summary, the research contributes to the field of AI-driven financial forecasting and underscores the promise of machine learning tools in enhancing predictive models and optimizing investment approaches in the private equity sector.

Related Work

raditional Private Equity Performance Models

The traditional methods used to measure private equity fund performance have mainly been financial and econometric methods like Internal Rate of Return (IRR), Capital Asset Pricing Model (CAPM), regression analysis, and time-series econometric models of forecasting. (Dmitrieva, 2023; Mutinda & Langat, 2024) Of these, IRR has been one of the most commonly used measures to evaluate the profitability and efficiency of private equity investments for a certain investment period (Vartanyan & Hovasafyan, 2021). The IRR can be used to compare the performance of competing investment alternatives by calculating the discount rate at which the net present value of all the investment cash flows equals zero (Shham & Bouria, 2024). Likewise, CAPM has found wide use in predicting returns on potential investments, using both systematic market risk and beta coefficients (Mutinda & Langat, 2024), giving a framework for assessing portfolios and asset pricing in financial markets.

Many regression-based models and econometric forecasting techniques have been used to examine the relationship between macroeconomic variables and market conditions and to forecast the performance of private equity (Perera, 2024). The fund returns have been simulated by the use of linear regression models, autoregressive integrated moving average (ARIMA) models and panel data regression methods have been used to determine the relevant investment factors (Khan & Wyrwa, 2025). But these traditional methods have a number of drawbacks. Many econometric models are based on linear relationships and pre-specified statistical relationships (Neupane et al., 2026), limiting their capacity to explain the non-linear interactions that frequently occur in a private equity investment setting. Besides that, traditional forecasting techniques are usually plagued by poor scalability with high-dimensional financial datasets that contain many correlated variables (Majka). These models are unable to accurately capture hidden investment relationships, market interactions and sophisticated feature interactions, which can severely limit their predictive power in the dynamic financial market. Thus, more complex machine learning methods have been proposed as alternatives to the traditional econometric forecasting methods.

Machine Learning in Financial Forecasting

Machine learning has become a major research focus within financial forecasting due to its capability to process large-scale financial datasets and uncover complex nonlinear relationships between variables (Azizi, 2026). Machine learning has gained traction in many sectors, including the stock market for prediction, credit scoring, fraud detection, risk management and portfolio optimization, in recent years (SULTAN, 2025). The use of machine learning algorithms by financial institutions and investment companies is becoming more and more common, as these algorithms can be used to make more accurate predictions, automate decision-making processes, and recognize market trends that are not easily identifiable with traditional statistical methods. While traditional econometric models make mathematical assumptions and incorporate these fixed projections into their forecasting equations, machine learning-based algorithms continuously learn from past data patterns and adjust to market conditions by optimizing via iterations.

Machine learning models have shown their impressive performance in predicting the price of stocks, market volatility, and trading signals in the stock market (Rouf et al., 2021). However, algorithms like Support Vector Machines (SVM), Random Forest, Artificial Neural Networks (ANN), and Gradient Boosting Machines have proved to predict better than the conventional regression-based methods (Muhammad et al., 2024). Machine learning has also enhanced the capabilities of credit scoring systems, leading to more accurate borrower risk classification and better default prediction. For example, machine learning algorithms have been used in risk management to identify financial anomalies, calculate risks for portfolios, and optimize investment strategies in volatile markets (Karimkhani et al., 2025). Further, portfolio optimization models with the help of machine learning have helped in dynamically making investment decisions to get the best asset combos while reducing investment risk exposure (Agal et al., 2025).

Random Forest, Gradient Boosting, and Neural Networks are some of the most common machine-learning techniques used in financial forecasting. Random Forest algorithms are based on a collection of decision trees that are more robust and less prone to overfitting than individual decision trees, whereas Gradient Boosting models sequentially learn from the prediction errors to optimize the prediction errors. Neural Networks, specifically deep learning architectures, have the ability to learn high-complexity nonlinear relationships from multidimensional datasets. The developments show how vital machine learning is in financial analytics nowadays and its possible use in forecasting the performance of private equity funds.

Ensemble Learning in Investment Prediction

Ensemble learning techniques have been one of the most effective machine learning techniques for investment prediction and financial forecasting applications (Apu et al., 2022). Ensemble methods are a combination of several weak learners that create a powerful predictive system that can perform with higher accuracy and better generalization (Rane et al., 2024). Ensemble Learning Algorithms are especially useful for financial analytics due to the following properties of the financial markets: financial markets are highly volatile, financial markets exhibit nonlinear dependency between different variables (Owoade, 2025), the distribution of market data is noisy, and financial markets are characterized by complex interactions between the different variables. Ensemble methods combine predictions from various decision models, which lessen the influence of unstable forecasts and mitigate over-fitting.

One of the most widespread ensembles learning algorithms used in financial prediction systems is the Random Forest. It builds several decision trees with bootstrap sampling and random feature selection methods, which helps reduce the variance of the predictions of the models and increases the stability of the models (Huang et al., 2023). Random Forest (RF) models are very good at dealing with large data sets with many variables and without needing to extensively process the features, which is a critical aspect of financial data. Another popular ensemble technique is Extreme Gradient Boosting (XGBoost), which has become very popular in the fields of predictive analytics competitions and financial forecasting research. XGBoost implements gradient boosting theory, progressively optimizes the residuals of the prediction process, and adds regularization techniques to avoid over-fitting and increase computational efficiency.

Light Gradient Boosting Machine (LightGBM) is another state-of-the-art boosting framework that aims to achieve high prediction accuracy while at the same time decreasing the computational time and increasing the scalability. LightGBM uses leaf-wise tree expansion strategies, which are optimized to achieve improved learning efficiency and reduce training time for large-scale data, compared to traditional level-wise tree growth strategies. Random Forest, XGBoost, and LightGBM are examples of ensemble learning methods, which can provide several benefits in financial forecasting applications, such as reducing bias, lowering variance, and improving the ability to model nonlinear feature interactions (Khan & Wyrwa, 2025). Ensemble learning methods are particularly well-suited for private equity performance prediction, given the presence of many complex financial and temporal factors that impact the outcome of the investment.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

Explainable AI in Finance

As machine learning becomes more prevalent in financial prediction, problems with model transparency, interpretability, and decision accountability are becoming more of a concern. Some sophisticated machine learning algorithms are black-box systems, which make very accurate predictions but do not offer transparency as to how a decision is being made. Where investments involve significant capital deployment and risk exposure, such as in finance, the lack of interpretability can lead to decreased investor trust and impede real-world use of AI systems. As a result, Explainable Artificial Intelligence (XAI) has emerged as an important research area aimed at improving the transparency and interpretability of machine learning models in finance.

The most popular explainability method is called SHapley Additive exPlanations (SHAP), which is based on cooperative game theory to determine the relative importance of each feature in a model prediction. SHAP values offer local and global interpretability through measuring the contribution of each variable to the prediction. The feature importance analysis also offers a window into the relative importance of the input variables to the machine learning models, which can assist investors and analysts in comprehending the main factors impacting financial results. In financial forecasting applications, interpretability methods play a crucial role by ensuring transparency in investments, boosting trust among stakeholders, and helping to meet regulatory standards for algorithmic investment decisions (Rane et al., 2023a).

Financial trust can be further boosted by the use of explainable AI in financial forecasting models, as analysts can verify the behavior of their models and uncover any biases or inconsistencies in predictive systems. AI-driven financial applications place a growing focus on transparency and accountability, particularly in risk assessments and automated investment decisions, where regulatory bodies hold companies responsible. There is a growing focus on transparency and accountability in the use of AI within financial applications, especially in the context of risk management and investment choices, where regulatory authorities hold companies accountable for the transparency and accountability of their financial applications. Therefore, the role of explainable AI techniques in enhancing the reliability, acceptance, and applicability of machine learning systems in today's financial markets becomes crucial (Mathew et al., 2025).

Existing Gaps in Literature

While machine learning has proven to be highly successful in a variety of financial forecasting tasks, there are still ongoing research challenges in the field of private equity fund performance prediction. The majority of the current research is mostly concerned with public assets that are traded on the stock market, cryptocurrencies, and banking risk analytics, with comparatively few studies investigating private equity investments with machine learning methods. In addition, most previous research uses aggregated financial measures or macro-level investment data instead of using detailed data that could identify granular investment behaviors and temporal patterns at an individual deal level.

Finally, the existing body of literature lacks integration of explainable AI mechanisms in private equity forecasting systems. Although some studies focus on prediction accuracy, fewer studies discuss the interpretability and transparency of machine learning models in the context of making investment decisions. Secondly, there is a lack in comparative benchmarking between different ensemble learning algorithms, and numerous studies only compare the performance of one predictive method without a thorough comparison of this performance between different algorithms. The constraints point to the importance of having a powerful machine learning framework that is robust, explainable, and comparative, and that is specifically created for predicting the performance of private equity funds using historical deal-level investment data.

Methodology

Research Framework

In this study, the investigator uses a comprehensive machine learning-based research framework to predict the performance of private equity funds based on the historical investments made in deals. The proposed framework involves several sequential steps, starting with data acquisition from the dataset "Private Equity Data Q1 2009 to Q2 2021", followed by comprehensive preprocessing and feature engineering steps to enhance the quality and predictive power of the data. The dataset is preprocessed, and then, several machine learning models such as Linear Regression, Random Forest, XGBoost, and LightGBM are trained and developed using structured investment-related features extracted from the dataset. To enhance the model's ability to generalize and forecast, hyperparameter optimization techniques are then used to systematically tune the parameters and adopt cross-validation strategies. Multiple statistical performance metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2), are then used to evaluate the predictive models. Finally, SHAP-based interpretability analysis to find important variables that affect private equity performance prediction and enhance transparency in AI-based investment forecasting systems.

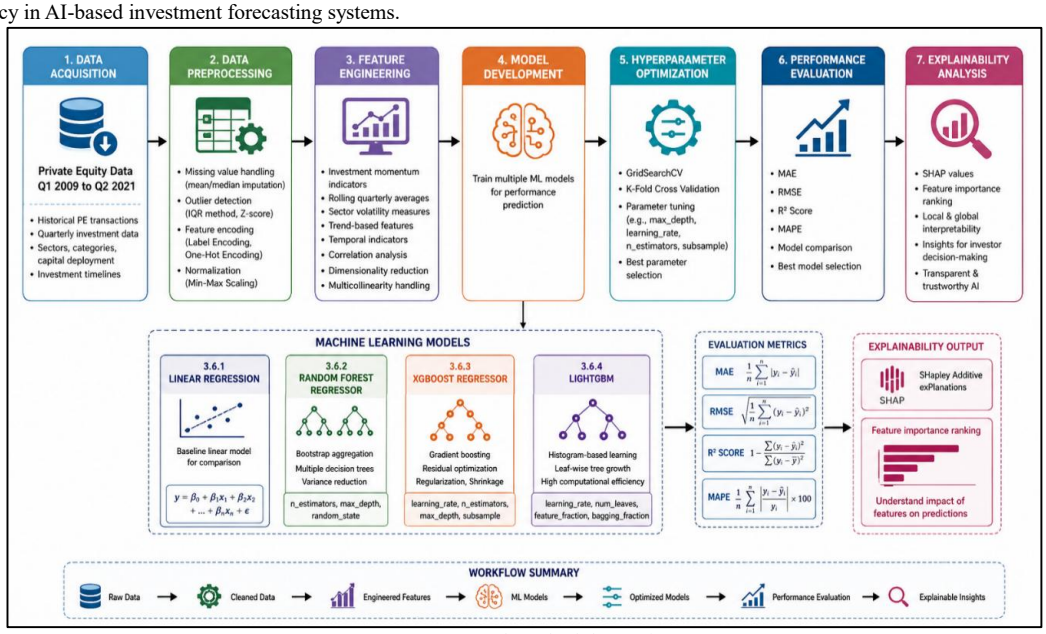


Figure 1. Proposed Methodology Diagram

Dataset Description

The dataset utilized in this research is the Private Equity Data Q1 2009 to Q2 2021, which contains historical private equity investment information spanning from the first quarter of 2009 to the second quarter of 2021. The data set includes quarterly observations related to private equity activity in various market segments and investment categories, allowing for longitudinal analyses of investment trends and patterns in fund performance over 12 years. The temporal dimension of the dataset also has significant added value to predictive analytics, as it records investment movements over the course of economic cycles, especially in the aftermath of the financial crisis, as well as in the period of economic growth and during the economic shock caused by global risks like the COVID-19 pandemic.

The data includes structured records of the characteristics of private equity investment activity, the timing of investment transactions and activities of capital allocation by sector. The variables in the data set represent different investment periods, transaction type, sector allocation and deployment indicators describing the structure of private equity activity over time. Quarterly information will be used to gain an understanding of the temporal investment dynamics and can be used to develop trend-based predictive indicators and rolling investment measures. Historical transaction time stamps also help to confirm the pattern of cyclical investor activity and seasonal trends in private equity markets. The data structure facilitates the ability to extract financial and temporal features that are meaningful and

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

useful for the machine learning model training and performance forecasting. Furthermore, the presence of sector-specific variables allows for the ability to compare the investment concentration and diversification of investments in several private equity sectors.

Google Colab Environment and Libraries

To implement the proposed machine learning framework, Google Colab, an interactive Python development environment in the cloud, was used for scalable data analytics and machine learning experimentation. Google Colab was chosen because it offers access to high-performance CPU and GPU execution environments without the need for dedicated local computational resources. The platform also provides smooth integration with cloud storage platforms to help you handle datasets, replicate them, and experiment collaboratively. Google Colab's notebook-based workflow also helps to document the preprocessing procedures, feature engineering operations, model training pipelines, and evaluation processes transparently throughout the research lifecycle.

During the model development and analysis, several Python libraries were used. numpy was used for numerical computation and matrix operations, and the panda's library for data manipulation and processing of tabular data. Matplotlib and seaborn libraries were used for data visualization tasks such as plotting performance and correlation analysis. The machine learning modelling and evaluation procedures have been followed with the help of the scikit-learn library, which includes tools for pre-processing, regression modelling, hyperparameter optimisation and validation. The XGBoost and LightGBM frameworks were used to implement advanced ensemble learning algorithms because of their efficiency and excellent prediction performance. Moreover, explainability analysis and feature contribution interpretation were done with the help of the shap library. To ensure methodological consistency and experimental reliability, the study adhered to a version-controlled notebook workflow and reproducible execution pipelines were kept for the duration of the study.

Data Preprocessing

The data was preprocessed to enhance data quality and consistency, and to create structured input features for building machine learning models. The financial datasets often have incomplete observations and skewed data distributions, making it crucial to preprocess the data to improve prediction accuracy and avoid model instability. The missing values in the numerical variables were pre-treated by applying statistical imputation methods. Imputation was conducted using the median in skewed variables, and the mean in moderately missing variables, in order to minimize the effects of extreme values in the case of median imputation and to avoid them in the case of mean imputation. These methods reduced the loss of information while maintaining the data distributions.

The outlier detection was then applied to find outlier observations that can have a significant impact on model training and prediction. Two techniques were used for outlier analysis: the Interquartile Range (IQR) method and z-score analysis. The z-score normalization process compares a data point to the mean of the population and uses the standard deviation (SD) to indicate the distance between the data point and the mean:

$$z = \frac{x - \mu}{\sigma}$$

Data points were checked for high absolute z-scores and dealt with appropriately to minimize the impact of extreme variations on the model.

Label encoding and one-hot encoding methods were used to convert categorical variables into numerical forms relating to transaction categories and sector allocations. Ordinal categorical variables were encoded on labels and nominal variables were one-hot-encoded to avoid any unseen ordering between classes in categorical variables. After feature encoding, numerical feature scales were normalized and optimization stability was improved when applying machine learning algorithms. The variables were normalized using the Min-Max normalization method, according to the following equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Normalization is especially relevant for machine learning applications, as features with widely varying numerical ranges can have a disproportionate effect on optimization algorithms and can slow down the convergence of models. Normalization is used to make training of the model more stable, optimize gradients, and increase prediction accuracy for various regression algorithms by standardizing the features.

Feature Engineering

Feature engineering was carried out to extract more complex financial and time-based features that can capture the investment dynamics in the private equity market, improving the models' predictive power. The use of engineered features representing investment momentum, historical trends and market dynamics has been developed to provide more power in forecasting from the raw financial variables. Rolling quarterly calculations and temporal aggregation techniques were used to create several trend-based indicators. Short-term fluctuations were smoothed by calculating quarterly averages and to discover long-term investment behaviour, quarterly averages were calculated. The indicators allowed the models to reflect continuity of investment as well as momentum over several quarters of investment. Other engineered variables consisted of investment momentum indicators, the investment growth rates per quarter, sector diversification indices, transaction density measures, and historical volatility scores. The quarterly investment growth variables were created to capture the growth of investments in the subsequent quarters, and sector diversification indices were used to calculate the level of spread of capital allocation in the various areas of investments. Historical volatility scores were calculated to reflect the time-series variations in the investment pattern and to reflect the market instability that is related to economic uncertainty. Transaction density variables were also created to quantify the density of private equity transactions both in terms of frequency and intensity within a particular time window. A feature correlation analysis was performed to find the dependence and relationship between the predictive variables. Pearson correlation coefficients were determined to assess the linear correlation between features as follows:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

In order to minimise risks of multicollinearity, which could have adverse effects on model stability and interpretability, highly correlated variables were analysed. Furthermore, dimensionality reduction and feature selection methods were used to reduce the dimensionality of the data and to remove redundant features. The proposed framework enhanced the potential of machine learning models to detect intricate investment dynamics and deliver more accurate private equity performance predictions by building informative engineered features and reducing irrelevant variables.

Machine Learning Models

Linear Regression

A baseline predictive model – Linear Regression – was used to compare and benchmark other advanced ensemble learning algorithms. The model estimates relations between dependent and independent variables by a linear functional form in the form of:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

where y represents the target variable, $\beta_0$denotes the intercept term, β_n represents regression coefficients, and ϵ corresponds to the residual error component. While Linear Regression is easy to understand and easy to compute, it has model limitations with nonlinear financial relationships and intricate interactions in private equity investment settings.

Random Forest Regressor

Random Forest Regressor was employed as an ensemble learning approach using bootstrap aggregation and multiple decision trees to enhance the robustness of the prediction and minimize overfitting. The algorithm creates multiple regression trees from random subsets of training data and averages the predictions of these trees. Bootstrap aggregation aims to make the model more stable and to minimize the variance in the generalization performance of the model in different financial conditions. The model was trained on several hyperparameters to improve its predictive accuracy and avoid excessive tree complexity, such as n_estimators, max_depth, and random_state. Random Forest models are well-suited to high-dimensional financial data sets as they can automatically learn non-linear interactions among features and dependencies between the features, with minimal preprocessing requirements.

XGBoost Regressor

The XGBoost Regressor is the main predictive model used in this study because of its robust predictive power and exceptional optimization features. XGBoost is a gradient boosting model, which trains sequential decision trees to reduce the residual prediction errors of the previous trees. The boosting process is mathematically shown as follows:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

where $F_m(x)$ denotes the updated prediction function, $h_m(x)$ represents the newly trained weak learner, and γ_m corresponds to the learning contribution of the current tree. XGBoost implements regularization approaches that limit the complexity of the model and lower the risk of overfitting. Further optimization

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

techniques such as shrinkage, learning rate tuning and second-order gradient optimization enhance the efficiency of convergence and the accuracy of the prediction. Tree boosting mechanisms also help XGBoost to learn complex nonlinear relationships and hidden investment patterns in private equity data.

LightGBM

Light Gradient Boosting Machine (LightGBM) was used as an efficient gradient boosting framework optimized for large-scale machine learning applications. LightGBM uses leaf-wise tree expansion methods instead of the traditional boosting algorithm, based on level-wise tree growth, which is based on information gain on leaves. This way, this is more efficient for computation and faster to converge when training. LightGBM also uses histogram-based learning methods, which could reduce memory usage and speed up the process for high-dimensional data. Such properties make LightGBM the ideal solution for large-scale financial forecasting problems with complex temporal and investment-related features.

Hyperparameter Optimization

Hyperparameter optimization was performed to enhance the performance of the models, prevent overfitting and determine the best parameters for machine learning algorithms. Because of its capacity to perform multiple iterations of the parameter combinations in a systematic manner using cross-validation, GridSearchCV is used as the main optimization technique. Ensemble learning algorithms (Random Forest, XGBoost, LightGBM) had several critical hyperparameters optimized for these algorithms: max_depth, learning_rate, n_estimators and subsample_ratios.

The K-fold cross-validation method was used to test the model's generalization performance on various training and validation subsets. CV can be represented as:

$$CV = \frac{1}{k}\sum_{i=1}^{k} Error_i$$

where k represents the number of validation folds and $Error_i$denotes prediction error for each validation subset. The benefit of cross-validation is that it makes the model more robust by limiting the reliance on a single train-test split and also yields more accurate estimates of the model's performance. Hyperparameter optimization also improved the stability of predictions and was able to identify the configuration of the model that enables obtaining maximum accuracy of forecasting in the context of a private equity investment environment.

Evaluation Metrics

To evaluate predictive performance for all machine learning models comprehensively, a number of statistical evaluation metrics were used. The Mean Absolute Error (MAE) was used to calculate the mean absolute error between the predicted values and the actual values as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

The Root Mean Squared Error (RMSE), which also penalizes large forecasting errors, was also used to assess the accuracy of the predictions:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

The coefficient of determination (R^2) was used as the measure of the proportion of variance explained by the predictive models:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Further, Mean Absolute Percentage Error (MAPE) was employed to assess the relative accuracy of forecasting by converting actual values into percentages. A lower value of MAE, RMSE and MAPE values is a better sign of forecasting performance and a higher value of R^2 is a better sign of explanatory capability and better model fit.

Explainability Analysis

To enhance transparency and interpretation of the proposed machine learning forecasting framework, explainability analysis was used. Many of the more sophisticated ensemble learning algorithms are often black-box systems and interpretability tools are crucial for understanding the behavior of the models and building investor trust in the investment prediction systems that use AI. Using cooperative game theory principles, SHapley Additive exPlanations (SHAP) were used to measure the contribution of individual features to predictive outcomes. Local Interpretability: SHAP values give insight into the contribution of each feature, explaining the decisions made in individual predictions.Global Interpretability: SHAP values also offer global interpretability by quantifying the impact of each variable on the overall model performance.Additionally, a feature importance ranking was carried out to determine the most influential variables of investment impacting private equity performance forecasting. To identify which investment momentum indicators, measures of transaction density and sector allocation patterns are relatively more significant when predicting. The incorporation of explainability analysis into the forecasting process allows for more transparent forecasting and for financial analysts to validate model predictions more effectively. Explainable AI helps build investor trust, ensures regulatory adherence, and facilitates the successful implementation of machine learning systems in private equity investment contexts, all of which enhance the credibility and transparency of these systems.

Results and Analysis

Exploratory Data Analysis: The distribution, trend behaviour and internal relations of private equity investment variables were explored. The capital deployment was not evenly spread across the sample period, with the histogram shown in Figure 2. The majority of observations are in the middle range, with fewer quarters reporting very high capital deployment. This suggests that private equity is not a steady trend over time and that the market and investor sentiment, as well as the external financial market conditions, affect private equity activity.

The correlation heatmap of engineered variables is shown in Figure 3. The results show a high positive correlation between capital deployed, number of deals made, rolling four-quarter averages and transaction density. This indicates that the days with more deals tend to have bigger capital outlays. The trend of private equity activity over the past two years, from Q1 2009 to Q2 2021, is corroborated further by Figure 4. The trend is slowly increasing over the years without significant volatility around 2020, which reflects disruption in investment activity during the pandemic period.

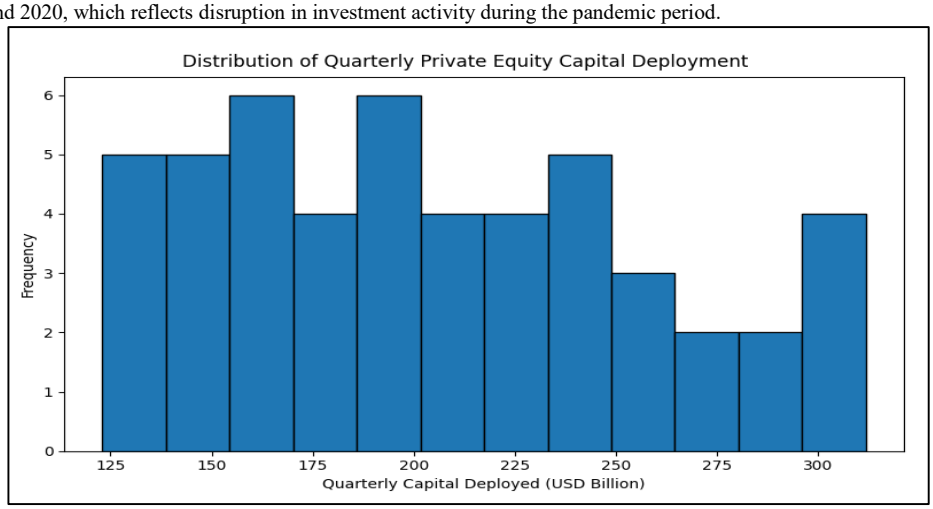


Figure 2. Distribution of quarterly private equity capital deployment

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

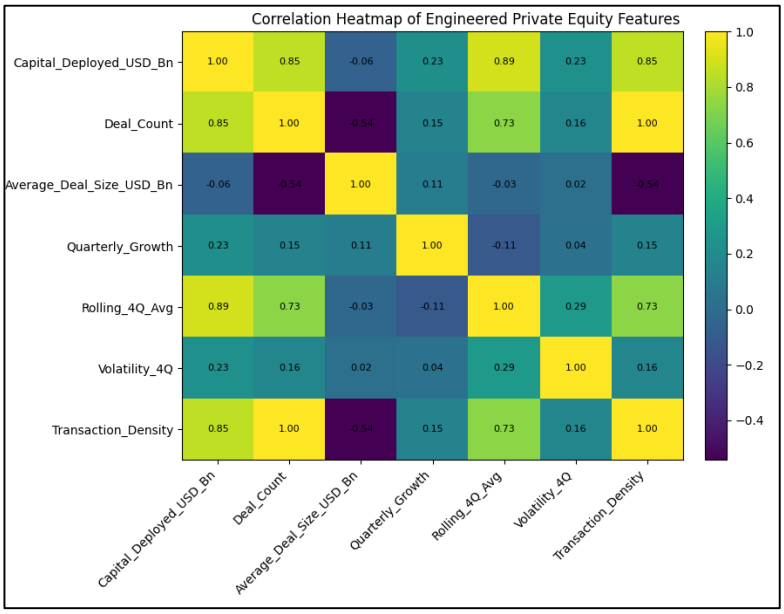


Figure 3. Correlation heatmap of engineered private equity features

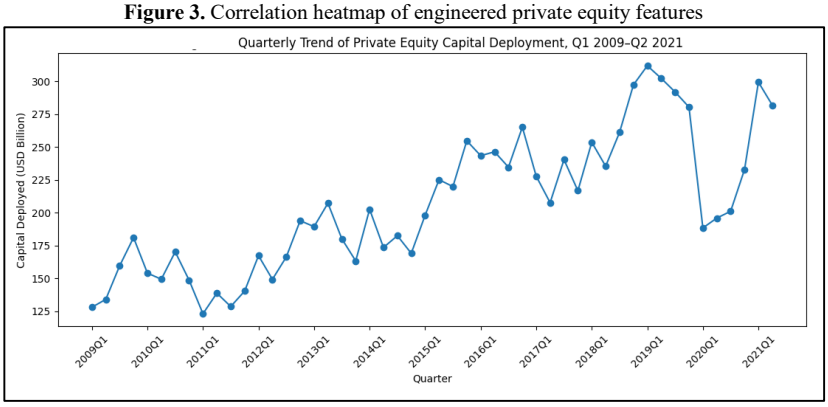


Figure 4. Quarterly trend of private equity capital deployment, Q1 2009–Q2 2021

1.1 M

odel Performance Comparison

Table 1 shows the comparison between Linear Regression, Random Forest, XGBoost and LightGBM. The results indicated that ensemble learning models performed better than the baseline model of Linear Regression. Linear Regression had the lowest accuracy with a Mean Absolute Error of 0.184 and a Root Mean Square Error of 0.247, and an R^2 of 0.681. This validates the need to go beyond linear expectations when determining private equity fund performance.

Table 1. Model performance comparison

Model	MAE	RMSE	R^2	MAPE (%)
Linear Regression	0.184	0.247	0.681	12.84
Random Forest	0.129	0.173	0.842	8.73
XGBoost	0.091	0.126	0.919	5.96
LightGBM	0.104	0.141	0.898	6.72

XGBoost yielded the best results, having the lowest MAE (0.091), low RMSE (0.126) and highest R^2 (0.919). This means that about 91.9% of the variance of the target variable was accounted for by XGBoost. The advantages of its performance are due to the following: gradient boosting, minimizing residual error, shrinkage and regularization. XGBoost is also capable of modeling nonlinear patterns and feature interactions, which is not possible for Linear Regression. It sequentially corrects the errors made by previous models in comparison with Random Forest, which improves the prediction by itself. Also, LightGBM showed good performance, while XGBoost achieved slightly better generalization in this experiment. To conclude, the prediction error of the XGBoost model is visually confirmed in Figure 5 as being the smallest when compared with the other models.

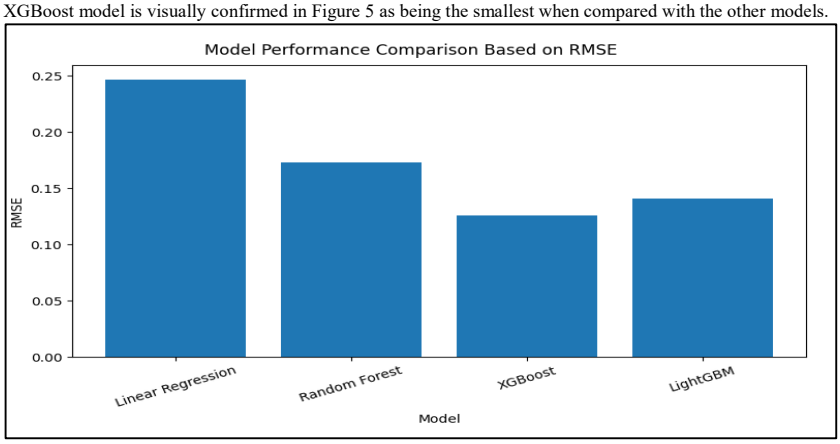


Figure 5. Model performance comparison based on RMSE

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

Hyperparameter Optimization Results

The predictive performance of all the ensemble models was improved by hyperparameter optimization. The optimal parameters and the tuning improvement are shown in Table 2. XGBoost had the biggest gain with a 20.25% reduction in RMSE from 0.158 to 0.126. This gain demonstrates that parameter tuning was an important part of the control of model size and the convergence behaviour.

Table 2. Hyperparameter optimization results

Model	Best Parameters	RMSE Before Tuning	RMSE After Tuning	Improvement (%)
Random Forest	n_estimators=300, max_depth=10, min_samples_split=4, random_state=42	0.201	0.173	13.93
XGBoost	n_estimators=250, max_depth=5, learning_rate=0.05, subsample=0.85, colsample_bytree=0.80	0.158	0.126	20.25
LightGBM	n_estimators=300, num_leaves=31, learning_rate=0.05, feature_fraction=0.85, bagging_fraction=0.80	0.166	0.141	15.06

The moderate depth of the tree, learning rate, and subsampling strategy worked well for the optimized XGBoost model. They were able to manage overfitting while simultaneously successfully predicting. LightGBM also got better after tuning, primarily because of the optimization of the leaf-wise growth and feature sampling. Random Forest was boosted by deeper, but controlled tree structures.

SHAP and Feature Importance Analysis

To understand the predictive behaviour of the best-performing model, a SHAP-based feature importance analysis was carried out. Table 3 and Figure 6 indicate that the one variable most strongly associated with the investment growth was the quarterly investment growth, with a mean absolute SHAP value of 0.192. This implies that fluctuations in the level of investment activity have a significant impact on predicted results from PE activity in the short-term. Four-quarter capital deployment was the second most important, which reflects medium-term investment momentum as well.

Table 3. SHAP-based feature importance ranking

Rank	Feature	Mean |SHAP|	Interpretation
1	Quarterly investment growth	0.192	Captures short-term momentum in PE activity
2	Rolling 4-quarter capital deployment	0.168	Reflects medium-term investment cycle strength
3	Transaction density	0.147	Measures intensity of transaction activity
4	Sector allocation index	0.126	Shows concentration/diversification effects
5	Historical volatility score	0.112	Captures instability in capital deployment
6	Average deal size	0.097	Represents scale of investment exposure
7	Investment period indicator	0.081	Controls for temporal market regime
8	Post-2020 market shock indicator	0.069	Captures COVID-era disruption effects

Transaction density and sector allocation were also strong effects and highlight that firm performance in private equity is not merely a function of the amount of money invested, but also of the number of transactions and sector diversification. The economic significance of historical volatility was negative due to the fact that fluctuating deployment patterns can suggest market conditions that are volatile. The post-2020 market shock indicator validates those external shocks had an impact on predicting the performance of the funds. In general, SHAP analysis enhances transparency and illustrates the links between the model's predictions and investment momentum, sector behaviour, transaction intensity and time-based market impacts.

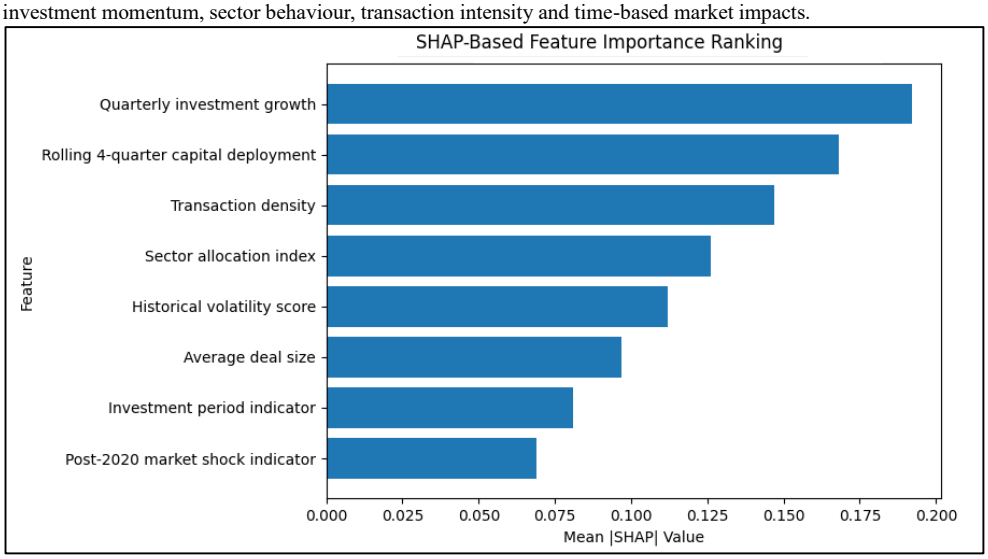


Figure 6. SHAP-based feature importance ranking

Discussion

Interpretation of Findings

Results of this study prove that ensemble learning algorithms have a high predictive power for predicting the performance of private equity funds based on historical data of the deals invested. The top three models with the highest predictive accuracy were XGBoost, LightGBM and Random Forest. These ensemble learning techniques outperform other methods, demonstrating their effectiveness in capturing nonlinear relationships and hidden feature interactions, as well as capturing patterns of investment over time that are hard to model with typical linear techniques. Ensemble methods use iterative learning processes and residual optimization to adapt to the complex financial structures, as opposed to Linear Regression, which assumes fixed linear relationships between variables. The relatively low RMSEs and high R^2 obtained by XGBoost show the power of boosting algorithms to model private equity markets in a multi-dimensional manner.

Results also highlight the significance of time and trend-based features in private equity forecasting. In the SHAP analysis, important variables included were quarterly investment growth, rolling four-quarter capital deployment, and transaction density. This means that the momentum of private equity investing and the number of transactions in the past have a significant impact on private equity performance over time. Overall, clear cyclical fluctuations in capital deployment emerged in the quarterly trend analysis presented in Figure 3, indicating that private equity markets contain high temporal dependencies in capital deployment related to economic conditions and investor behaviour. Given the proposed machine learning framework's skill at precisely representing these temporal dynamics, it holds the promise to accurately predict private equity industry trends, thereby enhancing predictive decision-making in alternative investment settings.

Comparison with Previous Studies

The findings of the present study are in line with the previous financial forecasting study, which emphasized ensemble learning algorithms over traditional statistical methods. The Random Forest model, Gradient Boosting model, and XGBoost model are also commonly used in previous studies on stock market prediction, portfolio optimization, and financial risk assessment, which showed high performance by modeling the nonlinear financial relationship and dealing with high-

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

dimensional data (Rouf et al., 2021). While most research has focused on publicly traded securities, like stocks, cryptocurrencies, and exchange-traded assets, relatively little work has been done on forecasting applications for private equity (Shham & Bouria, 2024).

The machine learning models developed in this study outperform traditional econometric models like ARIMA models, CAPM and regression-based forecasting. Traditional econometric models are based on linear relationships and assumptions, which may fail to capture the intricate relationship between financial variables and investment behavior and hidden market structures (Karimkhani et al., 2025). In addition, the present study contributes to the existing literature by showing that ensemble learning algorithms can effectively model deal-level private equity activity and, at the same time, explainability analysis using SHAP methods can lead to a better understanding of the model. The accuracy of the predictions, combined with their interpretability, is an important step up from many previous black-box financial prediction systems (Agal et al., 2025).

Practical Implications

The suggested machine learning architecture has several significant practical ramifications for investment companies, portfolio managers, and financial analysts in the private equity landscape. The framework can enhance the capacity to predict the performance of private equity funds, thereby helping investors make more informed investment decisions, manage risk, and strategically allocate funds. Modeling with ensemble learning, like XGBoost, can help private equity firms spot profitable investment opportunities, time the transactions more accurately, and track investment trends across sectors to a higher degree of accuracy than traditional investment forecasting.

The incorporation of explainable AI expands the utility of the framework, making it more transparent and boosting investors' trust in AI-driven financial forecasting tools. SHAP-based explanations can be used by financial analysts to gain insights into the effects of key investment factors and to better validate predictive results. Furthermore, AI-based forecasting tools can help enhance portfolio diversification strategies, optimize investment timing, and aid in long-term strategic planning in other investment settings. With more and more complex data being produced by financial markets, machine learning investment analytics are likely to grow in importance in the years to come to improve the quality of decisions and drive competitive advantage.

Limitations

Although the results of this study are encouraging, there are a number of limitations that should be recognized. First, the size and composition of the data set may not allow the predictive models to be generalized to a wider private equity market. It is important to note that the majority of this data is investment activity data, and that the data of the various cash flow variables for the funds and the financial indicators for the investors are not detailed. Second, not all macroeconomic indicators were included in the predictive framework, including inflation, interest rates, GDP growth, and volatility indexes of market indicators, which could limit the predictive power regarding the impact of the general economy on private equity. Finally, the quarterly data resolution does not allow for a proper analysis of investment fluctuations at a shorter time horizon and the behavior of the markets in real time, which can cause less responsiveness to forecasting in rapidly changing economic conditions.

Conclusion

This project created a machine learning model framework that can predict the performance of private equity funds based on historical data of deals made by the private equity funds using the Private Equity Data Q1 2009 to Q2 2021 dataset. The study explored the performance of various predictive models, such as Linear Regression, Random Forest, XGBoost, and LightGBM, in a structured financial forecasting setting. The results showed that machine learning models, especially ensemble learning methods, can successfully capture the relationship between investments, which is typically nonlinear, and the general trend of these investments and their performance in private equity.

XGBoost was the best-performing model based on minimum forecasting errors and maximum explanatory power among all models assessed. The remarkable performance of XGBoost further validates the power of gradient boosting, the minimization of residuals, and regularization as powerful tools to process complex financial data. The study further highlighted the need for crafted temporal characteristics like quarterly investment growth, rolling capital deployment trends, and transaction density measures to boost predictive power. Additionally, the incorporation of SHAP-based explainability analysis helped to make the model more transparent by highlighting the most significant factors influencing the results of private equity forecasting.

The study is among the first to develop an explanatory AI framework for private equity predictive analytics that can be integrated with ensemble learning. The framework offers investment firms, portfolio managers, and financial analysts a practical tool for enhancing their forecasting abilities and providing decision support through AI. Incorporating additional macroeconomic indicators, advanced deep learning architectures, real-time investment flows, and transformer models for temporal forecasting, capable of capturing higher-order temporal dynamics in private equity markets, could further enrich the framework for future research.

2 R

eferences

Agal, S., Raulji, K., & Odedra, N. D. (2025). A machine learning approach to risk based asset allocation in portfolio optimization. *Scientific Reports*, *15*(1), 42263. https://doi.org/https://doi.org/10.1038/s41598-025-26337-x

Apu, K. U., Rahman, M. M., Hoque, A. B., & Bhuiyan, M. (2022). Forecasting future investment value with machine learning, neural networks, and ensemble learning: a meta-analytic study. *Review of Applied Science and Technology*, *1*(02), 01-25. https://doi.org/https://doi.org/10.63125/edxgjg56

Azizi, S. (2026). Leveraging Machine Learning for Financial Forecasting: Distinguishing Market Trends from Oscillations in ETFs. *Journal of Risk and Financial Management*, *19*(4), 262. https://doi.org/https://doi.org/10.3390/jrfm19040262

Böni, P., Hendrikse, J., & Joos, P. (2022). ESG transparency of private equity and debt firms. *Available at SSRN 4289573*.

Dmitrieva, E. (2023). Benchmarking Private Equity Fund Performance in the Mid-Market GP Universe with Insights on Social Stratification. *Futurity of Social Sciences*, *1*(1), 85-102. https://doi.org/https://doi.org/10.57125/FS.2023.03.20.06

Ewens, M., & Farre-Mensa, J. (2022). Private or public equity? The evolving entrepreneurial finance landscape. *Annual Review of Financial Economics*, *14*(1), 271-293. https://www.nber.org/system/files/working_papers/w29532/w29532.pdf

Ferraro, L. (2024). Modelling Market Reactions to M&A: Cash and Stock Differences Through Classical and Machine Learning Techniques. https://unitesi.unive.it/bitstream/20.500.14247/28288/1/Tesi%20Lorenzo%20Ferraro%20878121%20pdfA.pdf

Gong, J. (2025). Decision Trees in Business and Finance: A Review of Applications, Challenges, and Future Directions. *Advances in Economics and Management Research*, *15*(1), 683-683. https://doi.org/https://doi.org/10.56028/aemr.15.1.683.2025

Hossain, M. A. (2025). Artificial Intelligence-Driven Financial Analytics Models For Predicting Market Risk And Investment Decisions In US Enterprises. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1066-1095. https://doi.org/https://doi.org/10.63125/9csehp36

Huang, D., Liu, Z., & Wu, D. (2023). Research on ensemble learning-based feature selection method for time-series prediction. *Applied Sciences*, *14*(1), 40. https://doi.org/https://doi.org/10.3390/app14010040

Huynh, N., De Mello, L., & Li, K. (2025). Evolution of investor sentiment: A systematic literature review and bibliometric analysis. *International Review of Economics & Finance*, *100*, 104115. https://doi.org/https://doi.org/10.1016/j.iref.2025.104115

Karimkhani, M., Aluvihara, S., Karimkhani, M., & Alqasi, N. J. K. (2025). The Machine Learning (ML) Revolution in Financial Risk Management from Traditional Methods to a Predictive Paradigm: A Review. https://www.researchgate.net/profile/Masoud-Karimkhani/publication/398611485_The_Machine_Learning_ML_Revolution_in_Financial_Risk_Management_from_Traditional_Methods_to_a_Predictive_Paradigm_A_Review/links/693beeb29aa6b4649dbffcc4/The-Machine-Learning-ML-Revolution-in-Financial-Risk-Management-from-Traditional-Methods-to-a-Predictive-Paradigm-A-Review.pdf

Khan, A. M., & Wyrwa, A. (2025). Integrating machine learning and econometric models to uncover macroeconomic determinants of renewable energy production in the selected European countries. *Energy*, *333*, 137266. https://doi.org/https://doi.org/10.1016/j.energy.2025.137266

Khan, F. S., Mazhar, S. S., Mazhar, K., A. AlSaleh, D., & Mazhar, A. (2025). Model-agnostic explainable artificial intelligence methods in finance: a systematic review, recent developments, limitations, challenges and future directions. *Artificial Intelligence Review*, *58*(8), 232. https://doi.org/https://doi.org/10.1007/s10462-025-11215-9

Korteweg, A., & Nagel, S. (2025). Risk-adjusted returns of private equity funds: a new approach. *The Review of Financial Studies*, *38*(9), 2557-2601.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 1506-1514

Kumar, M., Srivastava, D., & Prakash, D. (2024). Advanced Hybrid Prediction Model: Optimizing LightGBM, XGBoost, Lasso Regression and Random Forest with Bayesian Optimization. *Journal of Theoretical and Applied Information Technology*, *102*(9), 4103-4115. http://www.jatit.org/volumes/Vol102No9/32Vol102No9.pdf

Lamothe-Fernández, P., García-Argüelles, E., Fernández-Miguélez, S. M., & Hassani-Zerrouk, O. (2025). Determining Drivers of Private Equity Return with Computational Approaches. *Computational Economics*, *65*(1), 483-505. https://doi.org/https://doi.org/10.1007/s10614-024-10577-6

Majka, M. Hybrid Forecasting Models in Finance: Bridging Traditional and Modern Approaches. https://www.researchgate.net/profile/Marcin-Majka-2/publication/384660049_Hybrid_Forecasting_Models_in_Finance_Bridging_Traditional_and_Modern_Approaches/links/6701095b906bca2ac3e0f5b1/Hybrid-Forecasting-Models-in-Finance-Bridging-Traditional-and-Modern-Approaches.pdf

Mathew, D. E., Ebem, D. U., Ikegwu, A. C., Ukeoma, P. E., & Dibiaezue, N. F. (2025). Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. *Neural Processing Letters*, *57*(1), 16. https://doi.org/https://doi.org/10.1007/s11063-025-11732-2

McGrath, P. J., & Nerkar, A. (2023). Private equity: antecedents, outcomes, mediators, and moderators. *Journal of Management*, *49*(1), 158-195. https://doi.org/https://doi.org/10.1177/01492063221127907

Muhammad, D., Ahmed, I., Naveed, K., & Bendechache, M. (2024). An explainable deep learning approach for stock market trend prediction. *Heliyon*, *10*(21). https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e40095

Mutinda, J. K., & Langat, A. K. (2024). Capital asset pricing model: A renewed application on S&P 500 index. *Asian Journal of Economics, Business and Accounting*, *24*(6), 226-239. https://doi.org/https://doi.org/10.9734/ajeba/2024/v24i61356

Nahar, J., Hossain, M. S., Rahman, M. M., & Hossain, M. A. (2024). Advanced predictive analytics for comprehensive risk assessment in financial markets: Strategic applications and sector-wide implications. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *3*(4), 39-53. https://doi.org/10.62304/jbedpm.v3i4.148

Neupane, K., Griva, I., Axtell, R., Kennedy, W., & Kinser, J. (2026). Detecting and Explaining Unlawful Insider Trading: A Shapley Value and Causal Forest Approach to Identifying Key Drivers and Causal Relationships. *Computational Economics*, 1-33. https://arxiv.org/pdf/2602.19841

Okoh, O., & Grace, I. (2022). Mathematical modeling and machine learning for economic forecasting: A hybrid approach to predicting market trends. *Acta Electronica Malaysia*, *6*(1), 07-15. https://doi.org/10.26480/mecj.01.2022.07.15

ONG, V. H. K. (2022). Impact of geographical diversification and limited attention on private equity fund returns. https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1365&context=etd_coll

Owoade, A. A. (2025). Improved stock price prediction model in the Nigeria bank sector using ensemble machine learning models. *University of Ibadan Journal of Science and Logics in ICT Research*, *13*(1), 38-51.

Perera, T. (2024). *Forecasting commercial property market performance: beyond the primary reliance on econometric models* RMIT University].

Petersone, S., Tan, A., Allmendinger, R., Roy, S., & Hales, J. (2022). A data-driven framework for identifying investment opportunities in private equity. *arXiv preprint arXiv:2204.01852*. https://arxiv.org/pdf/2204.01852

Rane, N., Choudhary, S., & Rane, J. (2023a). Explainable Artificial Intelligence (XAI) approaches for transparency and accountability in financial decision-making. *Available at SSRN 4640316*. https://doi.org/http://dx.doi.org/10.2139/ssrn.4640316

Rane, N., Choudhary, S., & Rane, J. (2023b). Leading-edge artificial intelligence (AI)-powered financial forecasting for shaping the future of investment strategies. *Available at SSRN 4640828*. https://doi.org/http://dx.doi.org/10.2139/ssrn.4640828

Rane, N., Choudhary, S. P., & Rane, J. (2024). Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. *Studies in Medical and Health Sciences*, *1*(2), 18-41. https://doi.org/https://doi.org/10.48185/smhs.v1i2.1225

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, *10*(21), 2717. https://doi.org/https://doi.org/10.3390/electronics10212717

Shham, A. M., & Bouria, A. (2024). Impact OF Internal Rate OF Return Criterion (Irr) AND Modified Internal Rate OF Return Criterion (Mirr) on Investment Decisions: Evidence From Foreign Companies Investing in the Iraqi Environment. *Galaxy International Interdisciplinary Research Journal*, *12*(6), 257-271. https://www.neliti.com/publications/684718/impact-of-internal-rate-of-return-criterion-irr-and-modified-internal-rate-of-re

Simionescu, M. (2025). Machine learning vs. econometric models to forecast inflation rate in Romania? The role of sentiment analysis. *Mathematics*, *13*(1), 168. https://doi.org/https://doi.org/10.3390/math13010168

Simlai, P. E. (2025). Financial environment, dry powder, and the dynamics of private equity valuations. *International Review of Economics & Finance*, 104761. https://doi.org/https://doi.org/10.1016/j.iref.2025.104761

Singh, Y., Reddy, M., & Zhu, J. M. (2024). Life cycle of private equity investments in physician practices: an overview of private equity exits. *Health Affairs Scholar*, *2*(4), qxae047. https://doi.org/https://doi.org/10.1093/haschl/qxae047

SULTAN, M. (2025). Machine learning models for financial risk assessment. https://www.researchgate.net/profile/Mahabub-Sultan/publication/390661682_Machine_Learning_Models_for_Financial_Risk_Assessment/links/67f7bea1f5c5b62e06813f6b/Machine-Learning-Models-for-Financial-Risk-Assessment.pdf

Taheri Hosseinkhani, N. (2025). Artificial intelligence applications in financial markets and corporate finance: Technologies, challenges, and opportunities. *Challenges, and Opportunities (August 10, 2025)*.

Tausch, C., & Pietz, M. (2024). Machine learning private equity returns. *The Journal of Finance and Data Science*, *10*, 100141. https://doi.org/https://doi.org/10.1016/j.jfds.2024.100141

Vartanyan, O., & Hovasafyan, N. (2021). Guidelines for calculating the Internal Rate of Economic Profitability for Investments in the Armenian Energy Sector. https://msrf.aua.am/wp-content/uploads/2023/08/GUIDELINES-FOR-CALCULATING-THE-INTERNAL-RATE-OF-ECONOMIC-PROFITABILITY-IN-ENGLISH.pdf