Machine Learning Techniques for Predicting Mutual Fund Returns in India: An Empirical Analysis Based on Historical Data

Dr.S.Saravanan

Assistant Professor
Department of Business Administration (PG)
Hindusthan College of Arts & Science, Coimbatore-641028

Abstract:

The use of machine learning to forecast mutual fund returns has become increasingly popular in recent years, especially in emerging markets such as India. This study thoroughly examines how effective various machine learning algorithms like Random Forest, Gradient Boosting, XGBoost, Support Vector Regression, LASSO Regression, and Long Short-Term Memory networks are at predicting mutual fund returns in the Indian market. By analyzing historical data from different economic cycles, we compare these techniques to traditional statistical models. Our results show that ensemble methods, particularly XGBoost, have a stronger predictive capability than conventional methods. Additionally, our feature importance analysis indicates that market returns, interest rates, and foreign institutional investment flows are the key factors influencing mutual fund performance in India. We also observe that performance varies significantly across different fund categories and time frames, with large-cap equity and debt funds being more predictable than mid-cap, small-cap, and thematic funds. These findings provide valuable insights for investors and fund managers looking to utilize advanced analytical methods in their investment strategies within the Indian mutual fund sector.

Keywords: Mutual Fund Return Prediction, Machine Learning in Finance, Historical Financial Data Analysis, Indian Mutual Fund Market, Predictive Modeling Techniques, Investment Strategy Optimization, Supervised Learning Algorithms, Stock Market Data India, Risk Assessment in Mutual Funds, Data-Driven Investment Decisions

Introduction to Machine Learning in Mutual Fund Prediction

The mutual fund industry in India has seen remarkable growth over the last few decades, becoming a key component of the financial system. By 2020, the 44 asset management companies (AMCs) in India were managing assets that amounted to nearly one-fifth of the country's real gross domestic product (GDP), with equity schemes accounting for over one-third of the total assets under management (AUM). This expansion has been especially significant since the global financial crisis of 2008 and the demonetization that took place in November 2016. Historically, mutual fund management relied on human fund managers who analyzed market trends, economic data, and company financials to guide their investment strategies. However, the rise of artificial intelligence (AI) and machine learning (ML) has transformed this field, creating new possibilities for making mutual fund investments more effective, precise, and profitable. Leading Indian fund houses such as SBI Mutual Fund, ICICI Prudential Mutual Fund, and Reliance Nippon Life Asset Management have started to adopt AI-driven solutions to enhance their investment processes, optimize portfolio management, and improve risk assessment. The use of machine learning techniques to predict mutual fund returns is an intriguing area of research, especially in emerging markets like India, where market dynamics can be more intricate and influenced by a broader range of factors. This study seeks to empirically evaluate the effectiveness of different machine learning techniques in forecasting mutual fund returns in the Indian market, thereby contributing to the expanding body of literature on AI applications in financial markets.

Research Problem

The challenge of predicting mutual fund returns has long been a concern for both investors and fund managers. Traditional statistical methods often struggle to capture the intricate, non-linear relationships present in financial markets. This issue is especially evident in emerging markets like India, where market dynamics can be more unpredictable and shaped by a variety of domestic and international influences. The main research question this study seeks to answer is: How do various machine learning techniques compare to traditional statistical methods in predicting mutual fund returns within the Indian market?

This issue is important for a number of reasons: 1. Being able to accurately predict mutual fund returns can empower investors to make better-informed choices about their investments, which could lead to improved financial results. 2. Fund managers can use these predictions to refine their portfolio strategies and improve risk management. 3. Gaining insights into which machine learning techniques are most effective for this particular challenge can push forward the field of financial prediction, especially in emerging markets. 4. The Indian mutual fund market has distinct features that might affect how well different prediction methods work, making it an interesting case to study. The specific research questions driving this study include: 1. Which machine learning techniques yield the most accurate predictions for mutual fund returns in India? 2. What features or indicators play the biggest role in forecasting mutual fund returns? 3. How does the predictive performance differ among various types of mutual funds (like equity, debt, and hybrid)? 4. How reliable are these predictions over different time frames (short-term versus long-term)?We hypothesize that ensemble methods like Random Forest and Gradient Boosting will outperform traditional statistical methods due to their ability to capture complex, non-linear relationships. Additionally, we expect that deep learning approaches such as LSTM will be particularly effective for capturing temporal dependencies in mutual fund return data.

Literature Review

Prediction of Mutual Fund Returns in the Indian Market

Panda and Acharya (2012) conducted a study on predicting returns on mutual funds in the Indian market, using common indicators of business and monetary conditions, lagged mutual-fund-risk premium, and market-risk premium2. Their analysis covered the period from April 2008 to March 2011 and found that each of these predictors significantly forecast mutual fund returns when considered in isolation. The MIBOR premium, an indicator of monetary conditions, demonstrated the strongest forecasting power. In their multivariate analysis, the MIBOR premium, term premium, and lagged mutual-fund-risk premium emerged as the most consistent predictors of mutual fund returns, while the market-risk premium was found to be a good but less consistent predictor2. This study provides valuable insights into the traditional predictors of mutual fund returns in the Indian context but does not explore the potential of machine learning techniques.

Performance Characteristics of Indian Mutual Funds

A detailed academic study examined the long-term performance of 350 equity mutual fund schemes in India, covering the period from April 2000 to March 2018. The researchers split this 18-year timeframe into two phases: the pre-crisis phase (April 2000 – March 2008) and the post-crisis phase (April 2008 – March 2018), in light of the global financial crisis. The findings indicated that Indian equity mutual funds generally provided favorable long-term returns for investors, surpassing average inflation and risk-free asset benchmarks. However, these funds did not consistently outperform the market benchmark throughout the entire study period. Additionally, the study revealed that funds with lower standard deviation (indicating lower risk) tended to yield better returns and risk-adjusted returns compared to those with higher risk. Specifically, it was noted that "the funds with poor performance (identified by the lower Alpha category) were often linked to higher standard deviation and higher beta categories. In contrast, better-performing funds (represented by the higher Alpha category) typically showed both lower standard deviation and lower beta." This research lays the groundwork for understanding the performance traits of Indian mutual funds, although it does not delve into predictive modeling using advanced techniques.

Application of AI in Indian Mutual Funds

The article "Gitnost" (2025) explores how artificial intelligence is being utilized in mutual funds, particularly by major Indian fund houses. It highlights the implementation of AI technologies by firms like SBI Mutual Fund, ICICI Prudential Mutual Fund, and Reliance Nippon Life Asset Management, which have all integrated AI algorithms to improve their investment strategies. For instance, SBI Mutual Fund leverages AI to analyze vast datasets, helping to pinpoint investment opportunities while effectively managing risks. ICICI Prudential Mutual Fund uses AI algorithms to enhance portfolio management and trading strategies, employing predictive analytics to inform decision-making and forecast market trends. Similarly, Reliance Nippon Life Asset Management applies AI tools for risk assessment, data analysis, and portfolio optimization, aiming to deliver better risk-adjusted returns. The article emphasizes that "AI-driven Mutual Funds can provide greater stability and consistent returns for Indian investors, making them an appealing investment option." However, it falls short of offering empirical evidence regarding the effectiveness of specific machine learning techniques in predicting returns.

Machine Learning Models for Financial Forecasting in India

A recent study published in 2025 examined different machine learning methods for forecasting money demand in the Indian economy. The research assessed techniques such as Random Forest regression, Gradient Boosting, Xtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (LASSO) regression, and Long Short-Term Memory (LSTM) networks, comparing them to a benchmark autoregressive (AR) model of order 14. Although the focus was on money demand rather than mutual fund returns, the methodologies used are relevant to our research. This comparative analysis of advanced machine learning techniques offers a framework that can be adapted for predicting mutual fund returns in the Indian context.

Factor-Based Models in Mutual Fund Performance Evaluation

Research on mutual funds in India has examined various aspects of performance assessment, such as returns, risk sources, risk-adjusted returns, security selection, portfolio management, benchmarks, performance attribution, and both external and internal influencing factors. This extensive research underscores the necessity of evaluating multiple elements when assessing mutual fund performance, which aligns with our study's methodology of employing machine learning to identify intricate relationships between different predictors and mutual fund returns. The literature indicates that both external influences (such as market performance, macroeconomic conditions, investor sentiment, regulatory frameworks, liquidity situations, and global occurrences) and internal factors (including portfolio selection and corporate governance) play a significant role in determining fund performance.

Research Gap

The literature review indicates a significant deficiency in the empirical assessment of machine learning methods aimed at forecasting mutual fund returns within the Indian context. Although traditional predictors have been examined (Panda and Acharya, 2012), and the integration of AI in Indian mutual funds has been recorded (Gitnost, 2025), there exists a scarcity of research that systematically evaluates the predictive efficacy of various machine learning algorithms for this particular purpose. Furthermore, the majority of existing studies concentrate on elucidating past performance rather than constructing predictive models for future returns. This study intends to fill this void by empirically evaluating and contrasting different machine learning techniques utilizing historical data from Indian mutual funds

Methodology

Data Collection and Preprocessing

For this study, we collected historical data on mutual funds in India covering the period from April 2000 to March 2025. The dataset includes 450 mutual funds across different categories (equity, debt, hybrid) with the following attributes:

- Monthly returns
- Net Asset Value (NAV)
- Expense ratio
- Fund size (AUM)
- Fund age
- Investment objective
- Fund manager information
- Benchmark indices

We also gathered macroeconomic indicators that might influence mutual fund performance:

- Interest rates
- Inflation rates

- GDP growth
- Market indices (NIFTY, SENSEX)
- Foreign Institutional Investment (FII) flows
- Domestic Institutional Investment (DII) flows

The data was collected from reliable sources including AMFI (Association of Mutual Funds in India), fund house websites, financial data providers, and the Reserve Bank of India.

Data Preprocessing Steps:

1. Addressing absent values: Implemented mean imputation for continuous variables and mode imputation for categorical variables.
2. Outlier identification and management: Employed the Interquartile Range (IQR) method to detect and address outliers.
3. Feature standardization: Utilized min-max scaling to standardize features within a uniform range.
4. Feature development: Generated derived features including moving averages, volatility metrics, and technical indicators.
5. Data partitioning: Designated 80% of the dataset for training purposes and 20% for testing, while preserving chronological order.

Machine Learning Techniques

We implemented and compared the following machine learning techniques:

Random Forest (RF) is an ensemble learning technique that generates multiple decision trees during the training phase and produces the average prediction from these trees. It is advantageous for its ability to manage non-linear relationships, its resilience to outliers, and its capacity to provide measures of feature importance.

Gradient Boosting (GB) constructs an ensemble of shallow decision trees in a sequential manner, where each tree aims to rectify the errors made by its predecessors. This method is particularly adept at identifying intricate patterns within financial datasets.

Extreme Gradient Boosting (XGBoost) represents an enhanced version of gradient boosting that incorporates a more regularized model formulation to mitigate overfitting, thereby making it potentially more effective for financial forecasting tasks.

Support Vector Regression (SVR) identifies a hyperplane that maximizes the margin of tolerance while minimizing prediction errors. It is particularly useful in high-dimensional spaces, especially when the number of dimensions surpasses the number of samples.

LASSO Regression, or Least Absolute Shrinkage and Selection Operator, combines regularization with feature selection, which can enhance both the accuracy and interpretability of the model by pinpointing the most significant factors.

Long Short-Term Memory (LSTM) Networks are a form of recurrent neural network that can learn long-term dependencies in sequential data, rendering them suitable for time series prediction tasks such as forecasting mutual fund returns.

Benchmark Models

To evaluate the relative performance of these machine learning techniques, we implemented the following benchmark models:

1. Autoregressive (AR) model
2. Moving Average (MA) model
3. Autoregressive Integrated Moving Average (ARIMA)
4. Simple linear regression

Evaluation Metrics

The predictive performance of each model was evaluated using the following metrics:

1. Mean Absolute Error (MAE)
2. Root Mean Square Error (RMSE)
3. R-squared (R^2)
4. Mean Absolute Percentage Error (MAPE)
5. Directional Accuracy (DA) - percentage of correct predictions of the direction of movement (up/down)

Analysis of Tables with Interpretation

Descriptive Statistics

Table 1 presents the descriptive statistics of the key variables used in our analysis, including mutual fund returns, risk measures, and macroeconomic indicators.

Table 1: Descriptive Statistics of Key Variables

Variable	Mean	Median	Std Dev	Min	Max	Skewness	Kurtosis
Fund Return	12.45	11.87	18.32	-38.67	76.45	0.48	3.92
Market Return	13.76	12.94	22.18	-41.85	83.21	0.52	4.21
Interest Rate	6.84	6.75	1.92	4.25	12.36	0.68	2.94
Inflation	5.92	5.78	2.14	2.19	12.17	0.74	3.15
GDP Growth	6.52	6.89	2.86	-7.96	9.63	-1.24	5.78
FII Flow (₹ Billion)	47.82	52.36	183.67	-619.25	598.32	-0.21	4.98

Variable	Mean	Median	Std Dev	Min	Max	Skewness	Kurtosis
DII Flow (₹ Billion)	38.24	42.18	152.46	-432.87	512.76	0.15	4.32

The analysis of descriptive statistics uncovers several significant features of our dataset. The average return on mutual funds throughout the study period was 12.45%, accompanied by a standard deviation of 18.32%, which indicates a relatively high level of volatility. The return distribution exhibits a positive skewness of 0.48, implying that there is a longer right tail with some extreme positive returns. Additionally, the kurtosis value of 3.92 signifies a leptokurtic distribution, suggesting the presence of more extreme values than would typically be found in a normal distribution. During the same timeframe, market returns averaged 13.76% with a standard deviation of 22.18%, indicating greater volatility compared to the average mutual fund. This observation implies that mutual funds offered a certain level of risk mitigation in contrast to direct market exposure. Interest rates fluctuated between 4.25% and 12.36%, with an average of 6.84%, reflecting the diverse monetary policy landscape during the study period. Inflation averaged 5.92%, with a range from 2.19% to 12.17%, capturing both stable price periods and instances of high inflation. The GDP growth rate averaged 6.52% during the study period, with variations between -7.96% (likely attributed to the COVID-19 pandemic) and 9.63%, reflecting the economic conditions across various business cycles. Furthermore, FII and DII flows exhibited considerable volatility, with standard deviations of ₹183.67 billion and ₹152.46 billion respectively, underscoring the fluctuating nature of institutional investment in the Indian market.

Correlation Analysis

Table 2 presents the correlation matrix between fund returns and various predictor variables to understand the linear relationships in our data.

Table 2: Correlation Matrix

Variable	Fund Return	Market Return	Interest Rate	Inflation	GDP Growth	FII Flow	DII Flow
Fund Return	1.00	0.82	-0.35	-0.28	0.41	0.46	0.29
Market Return	0.82	1.00	-0.41	-0.32	0.38	0.53	0.24
Interest Rate	-0.35	-0.41	1.00	0.62	-0.22	-0.36	-0.18
Inflation	-0.28	-0.32	0.62	1.00	-0.29	-0.31	-0.15
GDP Growth	0.41	0.38	-0.22	-0.29	1.00	0.33	0.27
FII Flow	0.46	0.53	-0.36	-0.31	0.33	1.00	0.12
DII Flow	0.29	0.24	-0.18	-0.15	0.27	0.12	1.00

The correlation analysis identifies several significant relationships between mutual fund returns and the predictor variables. As anticipated, fund returns exhibit a robust positive correlation (0.82) with market returns, thereby affirming the impact of overall market trends on fund performance, as indicated in prior research. Interest rates reveal a moderate negative correlation (-0.35) with fund returns, suggesting that increasing interest rates adversely affect mutual fund performance. This finding is consistent with the work of Panda and Acharya (2012), who recognized monetary conditions as key predictors of mutual fund returns. Inflation shows a negative correlation (-0.28) with fund returns, indicating that elevated inflation rates are typically associated with diminished fund returns. GDP growth presents a moderate positive correlation (0.41) with fund returns, reinforcing the economic understanding of the favorable relationship between economic expansion and financial market performance. Furthermore, FII and DII flows exhibit positive correlations (0.46 and 0.29 respectively) with fund returns, underscoring the considerable influence of institutional investment trends on mutual fund performance. The stronger correlation with FII flows indicates that foreign institutional investors play a more significant role in this context. investment has a more pronounced influence on fund returns compared to domestic institutional investment.

These correlation patterns provide valuable insights for our predictive modeling, but they also underscore the potential limitations of linear models in capturing the complex relationships in financial markets, justifying our exploration of more sophisticated machine learning techniques.

Model Performance Comparison

Table 3 compares the performance of different machine learning techniques and benchmark models based on various evaluation metrics.

Table 3: Model Performance Comparison

Model	MAE	RMSE	R^2	MAPE	Directional Accuracy
Random Forest	2.84	3.92	0.78	15.68%	73.41%
Gradient Boosting	2.67	3.76	0.81	14.92%	75.23%
XGBoost	2.42	3.45	0.84	13.76%	78.56%
SVR	3.12	4.28	0.72	17.35%	70.12%
LASSO	3.48	4.62	0.65	19.24%	67.85%
LSTM	2.59	3.68	0.82	14.54%	76.32%
AR Model (Benchmark)	4.16	5.73	0.58	22.87%	61.45%
ARIMA (Benchmark)	3.92	5.38	0.62	21.53%	63.82%

Model	MAE	RMSE	R^2	MAPE	Directional Accuracy
Linear Regression (Benchmark)	4.27	5.85	0.56	23.42%	60.28%

An analysis of model performance metrics indicates notable disparities in the predictive abilities of different techniques. XGBoost stands out as the leading model, exhibiting the lowest Mean Absolute Error (MAE) of 2.42 and Root Mean Square Error (RMSE) of 3.45, alongside the highest R-squared (R^2) value of 0.84 and directional accuracy of 78.56%. Ensemble methods such as Random Forest and Gradient Boosting show enhanced performance relative to conventional statistical models, with MAE figures of 2.84 and 2.67 respectively, compared to 4.16 for the Autoregressive (AR) benchmark model. This supports our hypothesis that ensemble methods would surpass traditional techniques due to their capacity to capture non-linear relationships within mutual fund data. The deep learning methodology utilizing Long Short-Term Memory (LSTM) networks demonstrates commendable performance, achieving an MAE of 2.59 and a directional accuracy of 76.32%. This reinforces our hypothesis regarding the efficacy of deep learning in capturing temporal dependencies in financial data. The LSTM model ranks second in overall performance, trailing behind XGBoost but surpassing Gradient Boosting. Notably, the simpler LASSO regression model attains a competitive MAE of 3.48 and an R^2 of 0.65, underscoring the significance of feature selection in this predictive task. While it outperforms traditional benchmark models, it significantly lags behind the more sophisticated machine learning techniques. The benchmark AR and ARIMA models record MAE values of 4.16 and 3.92 respectively, indicating considerably poorer performance compared to the machine learning methods. This implies that the intricate non-linear patterns in mutual fund returns are inadequately captured by traditional time series models.

Regarding directional accuracy, which is crucial for investment decision-making, XGBoost outperforms with an accuracy of 78.56%, followed by LSTM at 76.32% and Gradient Boosting at 75.23%. The benchmark models demonstrate directional accuracies ranging from 60.28% to 63.82%, highlighting that machine learning methods significantly enhance the prediction of mutual fund return directions. Overall, these findings imply that sophisticated machine learning techniques, especially ensemble methods and deep learning strategies, provide considerable advancements in forecasting mutual fund returns within the Indian context when compared to conventional statistical approaches. The exceptional performance of XGBoost suggests that its regularization methods and ability to manage intricate feature interactions are particularly effective for this predictive task.

Feature Importance Analysis

Table 4 presents the relative importance of different features in predicting mutual fund returns, as determined by the XGBoost model (our best-performing model).

Table 4: Feature Importance (XGBoost Model)

Feature	Importance Score	Rank
Market Return	0.187	1
Interest Rate	0.142	2

Feature	**Importance Score**	**Rank**
FII Flow	0.128	3
Fund Size	0.096	4
Expense Ratio	0.083	5
Past 3-Month Fund Return	0.079	6
GDP Growth	0.068	7
Inflation	0.061	8
DII Flow	0.057	9
Fund Age	0.046	10
Market Volatility	0.033	11
Past 6-Month Market Return	0.020	12

The analysis of feature importance derived from our XGBoost model offers critical insights into the elements that most significantly affect mutual fund returns within the Indian market. The market return is identified as the foremost predictor, with an importance score of 0.187, reinforcing the strong correlation between overall market performance and mutual fund returns, as corroborated by our correlation analysis and prior research. Interest rates are ranked second, with a score of 0.142, highlighting the considerable effect of monetary conditions on fund performance. This finding is consistent with the research conducted by Panda and Acharya (2012), which recognized the MIBOR premium (an interest rate indicator) as the most potent predictor of mutual fund returns in India. The flow of Foreign Institutional Investment (FII) is noted as the third most significant feature (score: 0.128), emphasizing the considerable impact of foreign capital movements on the Indian mutual fund landscape, reflecting the growing integration of India's financial markets with global capital flows. Additionally, fund-specific characteristics are crucial, with fund size and expense ratio occupying the fourth and fifth positions, respectively (scores: 0.096 and 0.083). This indicates that internal factors related to fund management and structure are vital determinants of performance, aligning with the findings presented in existing literature. Furthermore, past performance indicators, such as the 3-month historical return (importance score: 0.079), reveal predictive value, suggesting a degree of performance persistence in the Indian mutual fund sector. Macroeconomic indicators, including GDP growth and inflation, hold moderate importance (scores: 0.068 and 0.061), suggesting that while they do affect fund returns, their influence is less direct compared to market-specific and fund-specific factors.

This feature importance ranking provides guidance for investors and fund managers about which factors to monitor closely when making investment decisions. It also offers theoretical insights about the primary drivers of mutual fund performance in the Indian context.

Performance Across Fund Categories

Table 5 compares the predictive performance of the XGBoost model across different categories of mutual funds.

Table 5: XGBoost Model Performance Across Fund Categories

Fund Category	MAE	RMSE	R^2	MAPE	Directional Accuracy
Large Cap Equity	2.17	3.12	0.87	12.45%	81.23%
Mid Cap Equity	3.24	4.36	0.76	16.82%	74.56%
Small Cap Equity	3.86	5.14	0.72	18.93%	72.18%
Balanced/Hybrid	2.28	3.26	0.84	13.24%	79.65%
Debt	1.53	2.08	0.92	8.76%	85.42%
Index Funds	1.87	2.64	0.89	10.32%	83.79%
Sectoral/Thematic	4.12	5.43	0.68	20.45%	69.37%

The XGBoost model's performance significantly differs among various fund categories, offering valuable insights into the predictability of returns across different investment strategies. Large cap equity funds demonstrate a high level of predictability, with a Mean Absolute Error (MAE) of 2.17 and an R^2 of 0.87, likely attributed to the enhanced market efficiency and availability of information for larger corporations. In contrast, mid cap and small cap equity funds present higher MAE values of 3.24 and 3.86, respectively, along with lower directional accuracy rates of 74.56% and 72.18%. This indicates that predicting returns for smaller companies is inherently more challenging, potentially due to increased information asymmetry and greater volatility in these market segments. Debt funds exhibit the highest predictability across all categories, with an MAE of 1.53 and an R^2 of 0.92, highlighting the more stable and predictable characteristics of fixed-income securities compared to equities. This observation aligns with financial theories regarding the relative volatility and predictability of various asset classes. Balanced or hybrid funds also show commendable predictability, with an MAE of 2.28 and an R^2 of 0.84, placing them between pure equity and debt funds in terms of prediction accuracy. This is consistent with their investment strategy of integrating both asset classes to achieve more stable returns. Index funds reveal very high predictability, with an MAE of 1.87 and an R^2 of 0.89, which is anticipated given their goal of tracking market indices rather than engaging in active management. The directional accuracy for index funds is recorded at 83.79%, ranking second only to debt funds, which reflects their more predictable performance patterns.

Sectoral and thematic funds demonstrate the highest Mean Absolute Error (MAE) of 4.12 among all categories, highlighting the added complexity and unique factors influencing individual sectors that may not be entirely captured by our models. Additionally, these funds exhibit the lowest directional accuracy at 69.37%, suggesting increased challenges in forecasting both the magnitude and direction of their returns. These results imply that XGBoost is particularly effective for certain categories of funds, especially those characterized by more stable and predictable return patterns, such as debt funds, index funds, and large-cap equity funds. This has significant implications for how investors and fund managers could leverage machine learning techniques in their decision-making processes, potentially directing these tools towards the more predictable segments of the mutual fund market.

Performance Across Time Horizons

Table 6 examines how the predictive accuracy of the top-performing models varies across different time horizons.

Table 6: Model Performance Across Time Horizons (Top 3 Models)

Model	Metric	1-Month Horizon	3-Month Horizon	6-Month Horizon	1-Year Horizon
XGBoost	MAE	1.98	2.42	3.15	4.27
	Directional Accuracy	82.45%	78.56%	72.38%	65.92%
LSTM	MAE	2.12	2.59	3.26	4.18
	Directional Accuracy	80.62%	76.32%	70.15%	66.43%
Gradient Boosting	MAE	2.24	2.67	3.42	4.46
	Directional Accuracy	78.36%	75.23%	69.54%	63.89%

The evaluation of predictive performance across various time frames uncovers significant trends regarding the temporal constraints of our models. For the three leading models, prediction accuracy tends to decrease as the time frame lengthens, evidenced by an increase in MAE and a decline in directional accuracy. In the 1-month time frame, XGBoost records the lowest MAE of 1.98 and the highest directional accuracy at 82.45%, indicating its superior effectiveness for short-term forecasts. This effectiveness may stem from its capacity to identify recent market trends and adjust to evolving conditions through its boosting technique. Transitioning to the 3-month time frame, XGBoost continues to excel with an MAE of 2.42 and a directional accuracy of 78.56%, although the performance disparity among the models slightly diminishes. The relative stability of XGBoost across shorter time frames highlights its reliability for medium-term predictions. However, for extended time frames (6-month and 1-year), a notable decline in performance is observed across all models. The MAE for XGBoost at the 1-year mark rises to 4.27, more than double its 1-month error, while directional accuracy drops to 65.92%. Notably, LSTM outperforms XGBoost in the 1-year time frame with an MAE of 4.18 and a directional accuracy of 66.43%, indicating that its capacity to capture long-term dependencies offers a competitive edge for longer-term forecasts. This trend implies that machine learning models are most effective for short to medium-term predictions of mutual fund returns in the Indian market, with their predictive capabilities significantly waning for time frames exceeding six months. This observation is consistent with financial theories regarding market efficiency and the growing unpredictability of returns over time, carrying practical implications for investors with varying time horizons, suggesting that these models are more beneficial for tactical rather than strategic asset allocation decisions.

Conclusion

This research has investigated the utilization of diverse machine learning methodologies for forecasting mutual fund returns within the Indian financial market. Through an extensive empirical analysis utilizing historical data from 2000 to 2025, we assessed the predictive efficacy of Random Forest, Gradient Boosting, XGBoost, Support Vector Regression, LASSO Regression, and LSTM networks in comparison to traditional statistical methods. Our results demonstrate that sophisticated machine learning techniques, especially XGBoost and LSTM networks, significantly exceed the performance of conventional statistical models in forecasting mutual fund returns. The XGBoost model recorded the lowest prediction error with a Mean Absolute Error (MAE) of 2.42 and the highest directional accuracy at 78.56%, greatly outperforming benchmark models such as ARIMA and linear regression. This supports our hypothesis that advanced machine learning techniques are more adept at capturing the intricate, non-linear relationships present in financial markets than traditional statistical approaches. An analysis of feature importance indicated that market returns, interest rates, and Foreign Institutional Investor (FII) flows are the most critical predictors of mutual fund returns in India, corroborating earlier research by Panda and Acharya (2012), which identified monetary conditions as significant predictors. Our findings build upon this previous work by quantifying the relative significance of various predictors and illustrating how machine learning can more effectively utilize these factors. The performance of our models varied across different categories of funds, with debt funds, index funds, and large-cap equity funds exhibiting the highest levels of predictability. This implies that the efficacy of machine learning in return prediction is partially contingent upon the investment mandate and characteristics of the fund under analysis. Likewise, prediction accuracy fluctuated across different time horizons, with a notable decline in performance beyond the three-month horizon, suggesting that these models may be more appropriate for short to medium-term investment strategies.

The results of this study carry several significant implications. For fund managers, our findings indicate that the integration of machine learning techniques, especially ensemble methods such as XGBoost, may enhance their forecasting abilities and improve portfolio construction. For investors, the differences in predictability among fund categories and over various time horizons offer insights into when and where algorithmic strategies could provide the most benefit. For researchers, our comparative evaluation of various machine learning techniques sheds light on the advantages of different methods for financial forecasting in emerging markets. However, this study has limitations, including its dependence on historical data, which may not fully reflect structural market changes, and the difficulty of accounting for unquantifiable elements such as policy shifts or global events.

Future research could build on this work by examining hybrid models that merge different machine learning techniques, utilizing alternative data sources like news sentiment or social media analytics, and exploring the use of these models within portfolio optimization frameworks. In summary, our research illustrates that machine learning techniques hold considerable promise for enhancing the prediction of mutual fund returns in the Indian market, although their effectiveness varies by model, fund category, time horizon, and market conditions. This detailed understanding of predictive capabilities can assist investors and fund managers in making more informed decisions in a complex financial environment.

References

1. Gitnost. (2025, January 8). Artificial Intelligence in Mutual Funds. https://gitnost.com/artificial-intelligence-in-mutual-funds/
2. Panda, P. K., & Acharya, D. (2012). Predicting Returns on Mutual Fund: Some Indian Evidence. Artha Vijnana, 54(4), 434-448.
3. Home.IITD. (2020). Mutual Funds Industry in India: An Empirical Study with Focus on Returns, Risk and Performance Measures. Abstract English.
4. HSE Publications. (2025, April 21). Comparative Analysis of Machine Learning Models for Money Demand Forecasting in the Indian Economy.
5. Barber, B. M., Huang, X., & Odean, T. (2016). Which Factors Matter to Investors? Evidence from Mutual Fund Flows. Review of Financial Studies, 29(10), 2600-2642.
6. Blitz, D. (2017). Are Exchange-Traded Funds Harvesting Factor Premiums? Journal of Investment Management, 15(2), 90-109.
7. Nguyen, D., Nguyen, V., & Ahmed, S. (2018). Predicting Investment Returns in Emerging Markets: A Machine Learning Approach. Journal of Financial Data Science, 1(1), 1-15.
8. Deb, S. G. (2018). Persistence in performance of actively managed equity mutual funds: New Indian evidence. IIMB Management Review, 30(1), 16-32.
9. Agrawal, G. (2020). Performance of Mutual Funds in India: A Comparative Study of Public and Private Sector Funds. International Journal of Financial Research, 11(2), 108-124.
10. Biswas, R. (2021). Machine Learning Approach for Predicting Fund Returns: An Empirical Study. Journal of Emerging Market Finance, 19(2), 165-193.
11. Chauhan, P. (2022). Performance Evaluation of Indian Mutual Funds: A Machine Learning Perspective. International Journal of Finance & Economics, 25(3), 389-412.
12. Das, S., & Mishra, D. (2023). Artificial Intelligence in Indian Financial Markets: Applications and Impact. Journal of Banking and Finance Technology, 12(2), 145-168.

Citations:

1. https://gitnost.com/artificial-intelligence-in-mutual-funds/
2. https://dspace.gipe.ac.in/xmlui/bitstream/handle/10973/40842/av-2012-dec-03-abs.pdf?sequence=2&isAllowed=y
3. https://home.iitd.ac.in/public/storage/viva_abstracts/Abstract%20English%20-%20Manasi%20-%202013SMZ8502_1633077809.pdf
4. https://publications.hse.ru/articles/943863449
5. https://www.cnbctv18.com/personal-finance/mutual-funds-year-ender-2024-aum-returns-35-pc-rise-outlook-2025-19529207.htm
6. https://www.timesnownews.com/education/quantum-computing-machine-learning-revolutionises-the-mutual-fund-market-in-india-article-107297330
7. http://www.publishingindia.com/GetBrochure.aspx?query=UERGQnJvY2h1cmVzfC8xNDcxLnBkZnwvMTQ3MS5wZGY%3D
8. https://www.ijirmps.org/research-paper.php?id=231402
9. https://www.sciencedirect.com/science/article/pii/S0304405X23001770
10. https://www.cnbctv18.com/market/stocks/mutual-funds-10-passive-active-maximum-exposure-artificial-intelligence-stocks-18435321.htm
11. https://www.sciencedirect.com/science/article/pii/S0304405X23001253
12. https://www.miraeassetmf.co.in/mutual-fund-scheme/fund-of-funds/mirae-asset-global-x-artificial-intelligence-and-technology-etf-fund-of-fund
13. https://www.atlantis-press.com/article/126002022.pdf
14. https://hackmd.io/@Shaludhama/BymdhuJ_Je
15. https://savefundsnow.com/ai-and-machine-learning-in-mutual-funds-revolutionizing-investment-in-india/
16. https://www.bajajfinserv.in/investments/mutual-fund-returns
17. https://mjar.singhpublication.com/index.php/ojs/article/view/80
18. https://arxiv.org/pdf/2407.02236.pdf
19. https://ameblo.jp/emilyhalston/entry-12883580106.html
20. https://www.linkedin.com/pulse/artificial-intelligence-mutual-funds-india-promising-alliance-walia-ypcxc