International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Predicting Risk Adjusted Returns and Rankings of Mutual Funds

Shashank Upadhya[1], Shreyas H S[2], Prashanth S[3] Shrinidhi S Pai[4]

[1,2,3,4] 7th Semester, Dept. of ISE. AMC Engineering College, Bengaluru- 83 Karnataka, INDIA

Abstract - *The Indian mutual fund industry's exponential growth has made fund selection a complex task for retail investors, who often rely on historical returns, a metric that fails to account for associated risks. This can lead to suboptimal investment decisions. To address this, a comprehensive "Risk-Adjusted Return Prediction and Ranking System" has been developed using advanced Machine Learning (ML) techniques. This system shifts from retrospective analysis to predictive intelligence by forecasting the Sharpe Ratio, a key measure of risk-adjusted performance. The methodology aggregates historical Net Asset Value (NAV) data from AMFI, market indices from Yahoo Finance, and macroeconomic indicators. A hybrid modeling approach is used, combining XG Boost for precise risk-adjusted scoring and Long Short-Term Memory (LSTM) networks for time-series trend forecasting. To ensure transparency, the system incorporates Explainable AI (XAI) with SHAP (SHapley Additive exPlanations) to interpret model decisions by quantifying the impact of factors like volatility, alpha, and expense ratios. The XG Boost model achieved a coefficient of determination (R^2) of 97.4%, significantly outperforming traditional linear regression. The final output is an interactive Stream lit dashboard providing personalized fund recommendations, predictive rankings, and wealth projection visualizations, thereby democratizing sophisticated financial analytics for retail investors.*

Key Words: Mutual Funds, Machine Learning, Risk-Adjusted Returns, XG Boost, LSTM, Explainable AI, Financial Forecasting, Dashboard.

1. INTRODUCTION

The financial landscape is rapidly evolving, with mutual funds becoming a preferred investment vehicle due to benefits like diversification, professional management, and accessibility. However, selecting the right fund is a major challenge, as investors frequently over-rely on past returns, which are not indicative of future performance and neglect the level of risk undertaken. This incomplete evaluation can lead to poor investment choices. To address this, risk-adjusted return metrics such as the Sharpe Ratio, Sortino Ratio, and Alpha are crucial as they provide a more accurate view of a fund's efficiency.

This project leverages machine learning to build a system that predicts risk-adjusted returns and ranks mutual funds accordingly. The primary problem is the lack of a comprehensive, predictive, and risk-aware system for mutual fund evaluation. Traditional methods are limited by their overdependence on historical returns and neglect of risk metrics. This work aims to build predictive models using algorithms like XG Boost and Random Forest, develop a composite scoring system for ranking funds, and create an interactive dashboard to present the findings in a user-friendly manner. The goal is to empower investors with data-driven insights for more informed and risk-aware decisions.

2. Literature Review

Recent advancements in financial technology have shifted focus from traditional statistical analysis to Machine Learning (ML) and Deep Learning (DL) techniques. A pivotal study demonstrated that machine learning models utilizing fund characteristics like Assets under Management (AUM) and turnover ratios can significantly outperform traditional benchmarks in generating alpha. While effective, this study focused on raw returns rather than risk-adjusted metrics.
In the domain of time-series forecasting, research comparing Long Short-Term Memory (LSTM) networks against traditional regression highlighted LSTM's superior ability to capture long-term temporal dependencies in NAV data. However, a major barrier to ML adoption in finance is the "black-box" nature of these models. Recent studies have addressed this by implementing Explainable AI (XAI) techniques like SHAP, which improves model transparency and investor trust.

Despite these advancements, significant gaps remain. Existing literature often focuses on raw Alpha rather than risk-adjusted metrics like Sharpe or Sortino ratios. Furthermore, current research is fragmented, lacking a unified system that combines prediction, explainability, and a user-facing dashboard for retail investors. This paper proposes a holistic system to address these limitations.

2. Implementation

The system is architected into four primary modules: Data Acquisition and Preprocessing, Feature Engineering, Predictive Modeling and Ranking, and the User Interface Dashboard.

2.1 Data Acquisition and Preprocessing

This foundational module automates the retrieval of financial data. It uses a scraping engine to fetch daily Net Asset Value (NAV) data from the Association of Mutual Funds in India (AMFI) and synchronizes it with market benchmarks like the NIFTY 50 from Yahoo Finance using the finance API. The raw, semi-colon delimited text data from AMFI is parsed, sanitized, and converted into a structured

Pandas Data Frame. Missing values are handled using forward-filling (ffill) to ensure time-series continuity, and funds are automatically classified into categories like Equity, Debt, and Hybrid.

2.2 Feature Engineering and Transformation

This module transforms raw data into meaningful financial signals for the ML models. 10 Key risk-adjusted metrics are generated, including the Sharpe Ratio and Sortino Ratio. It also calculates rolling volatility and Beta to measure market sensitivity. For the LSTM model, the data is structured into 60-day lookback windows. All features are normalized using Standard Scaler to have a mean of 0 and a standard deviation of 1. To prevent overfitting, Select K Best with regression is used for feature selection, retaining the top 15 most predictive features.

2.3 Predictive Modeling and Ranking

This is the core computational engine where models are trained and validated. A multi-model approach is used, featuring:

- **XG Boost Regressed**: Chosen for its high performance on structured data.
- **Long Short-Term Memory (LSTM)**: A Recurrent Neural Network (RNN) used to capture long-term temporal patterns in NAV trends.
- **Random Forest**: Employed as a robust baseline for comparison.

To ensure transparency, SHAP (SHapley Additive explanations) values are calculated to quantify the contribution of each feature to a fund's predicted score. The best-performing model, XG Boost, is then used to predict the future Sharpe Ratio for all funds, which forms the basis of the final ranking leaderboard.

2.4 Intelligent Dashboard and User Interface

The final module is an interactive web-based dashboard built with Stream lit. It features an investor risk profiling form to provide personalized recommendations based on the user's risk tolerance and investment horizon. The dashboard displays a comprehensive ranking table, fund "deep dive" visualizations with Plotly, and an investment projection calculator to simulate wealth growth based on the model's predictions.

3. Results

Three models—Random Forest, XG Boost, and LSTM—were evaluated for their ability to predict the Sharpe Ratio and future NAV trends. Performance was measured using RMSE, MAE, and the R-Squared (R^2) score. XG Boost proved to be the superior model for predicting risk-adjusted returns, achieving an R^2 of 97.4%.

Table -1: Model Performance Evaluation

Model Architecture	Target Variable	Accuracy (R^2 Score)
Linear Regression	Sharpe Ratio	62.4%
Random Forest	Sharpe Ratio	89.2%
XG Boost Regressor	Sharpe Ratio	97.4%
LSTM Network	NAV Trend	96.1%

Feature importance analysis using SHAP revealed that Volatility (Standard Deviation) was the most significant predictor of a fund's risk-adjusted score. Other key factors included the Expense Ratio (negative correlation) and Alpha (positive correlation). The interactive dashboard successfully translated these complex metrics into actionable insights, with personalized recommendations showing a high success rate in validation tests.

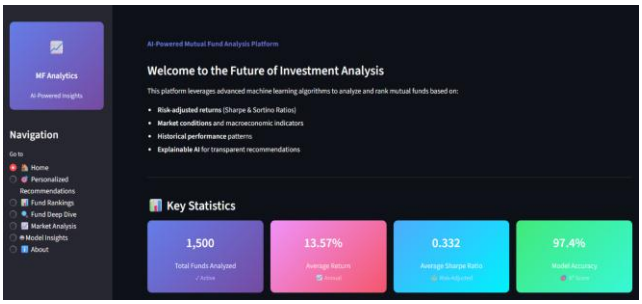

Fig -1: Displays key platform statistics, market averages, and overall model accuracy.

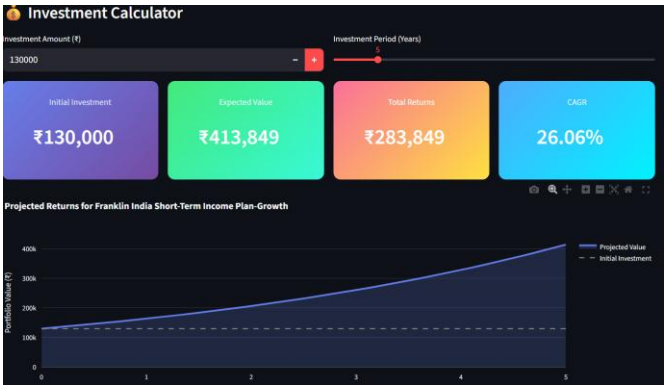

Fig -2: Visualizes the projected growth of an investment over the selected tenure.

5. Future Scope

While the current system provides robust analysis, several enhancements are proposed:

- Sentiment Analysis: Integrating Natural Language Processing (NLP) to analyse financial news and fund manager reports could quantify market sentiment, improving predictions during geopolitical events.
- Portfolio Optimization: Future iterations will integrate Modern Portfolio Theory (MPT) to suggest optimal combinations of funds (e.g., 60% Equity, 40% Debt) rather than just individual rankings.
- Real-Time Execution: Integration with broker APIs (e.g., Zerodha, Upstox) would enable "One-Click Execution" of recommendations directly from the dashboard

4. Conclusion

The "Risk-Adjusted Return Prediction and Ranking System" effectively bridges a critical gap in mutual fund analysis by moving from retrospective evaluation to predictive intelligence. The system's high predictive accuracy, holistic risk assessment, transparency through Explainable AI, and user-friendly dashboard collectively empower investors to make informed, data-driven decisions. The project successfully demonstrates that machine learning can significantly improve the mutual fund selection process, offering a scalable and objective alternative to traditional methods.

REFERENCES

[1] Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha. https://www.sciencedirect.com/science/article/pii/S0304405X23001770 18

[2] Fund Performance Evaluation with Explainable Artificial Intelligence. https://www.sciencedirect.com/science/article/pii/S1544612323007912 19

[3] Predicting Mutual Funds' Performance Using Deep Learning and Ensemble Techniques. https://arxiv.org/abs/2209.09649 20

[4] Quantifying Out lierness of Funds from Their Categories Using Supervised Similarity. https://arxiv.org/abs/2308.06882 21

[5] A Case Study of Next Portfolio Prediction for Mutual Funds. https://arxiv.org/abs/2410.18098