JOURNALS | Journal of Information and Communications Technology: Algorithms, Systems and Applications
Volume 2 Issue 3 (September 2026)
DOI: https://doi.org/10.64189/ict.26312
J. Inf. Commun. Technol. Algorithms, Syst. Appl., 2026, 2(3), 26312

Research Article

OPEN ACCESS

AI-Driven Unified Investment Guidance Platform for Diverse Asset Classes

Mohammad Ali Ansari,* Huzaifa Zahid Husein Shah, Talha Ansari, Aditya Vijay Kamble and Ahlam Ansari

Department of Computer Engineering, M. H. Saboo Siddik College of Engineering, Byculla, Mumbai, Maharashtra, 400008, India

*Corresponding Author
Mohammad Ali Ansari
ali.221204.co@mhssce.ac.in

Received: 10 May 2026
Revised: 17 August 2026
Accepted: 24 August 2026
Published Online: 25 August 2026

Citation
M. A. Ansari, H. Z. Husein Shah, T. Ansari, A. V. Kamble, A. Ansari, AI-driven unified investment guidance platform for diverse asset classes, *Journal of Information and Communications Technology: Algorithms, Systems and Applications,* 2026, 2(3), 26312, https://doi.org/10.64189/ict.26312.

Abstract
Many investors in India face difficulties in choosing the suitable investment avenue among diverse options. This happens due to fragmented financial platforms, diverse asset behaviors, and limited access to intelligent decision support tools. Existing research mainly focuses on single-asset forecasting leaving a gap for a unified, multi-asset investment guidance system. This research proposes an AI supported unified investment guidance platform that integrates all major asset classes such as stock, metals, real estate, mutual funds, cryptocurrency into a single user-friendly framework. This platform utilises historical and real-time data collected through APIs, web scraping, and public datasets, combined with data analysis, machine learning and deep learning models for accurate presentation and forecasting. Random Forest is employed for structured datasets such as gold and real estate, while Long Short-Term Memory (LSTM) networks are used for time-series prediction of stocks, cryptocurrency and mutual fund NAVs. To enhance contextual understanding, a sentiment analysis module processes financial news and market headlines, and an AI-based advisory layer converts analytical outputs into natural language investment insights. Experimental research exhibits strong predictive performance, achieving R^2 values of 0.92 for gold, 0.88 for real estate, 0.78 for stocks, 0.80 for cryptocurrency, and 0.99 for mutual funds. The results confirm that combining real time and historical Data along with integrating quantitative forecasting with sentiment-aware and AI-driven advisory mechanisms significantly improves interpretability and decision support. The proposed platform contributes toward democratizing financial guidance in India by enabling personalized, data-driven, and holistic investment decision-making.

Keywords: Finance; Multi assets; Wealth management; Investment; Investment guidance.

https://doi.org/10.64189/ict.26312

1. Introduction

1.1 Background and importance of investment decision-making

Over the past decade, the growth of global financial markets has dramatically increased the number of investment opportunities available to individual investors, and while this trend is most evident in developed countries, it also represents increasing complexity for individual investors in developing economies such as India. Today, individual investors are faced with a multiplicity of asset classes, including stocks, mutual funds, gold, real estate, and cryptocurrencies, each with its own unique characteristics of both risk and return. Sound investment decision-making requires investors to continuously interpret this diverse and fast-changing information across asset classes, a task that directly influences long-term wealth creation, retirement planning, and financial security for households.[1] As financial markets in India continue to deepen and diversify, the ability to make timely, well-informed investment decisions has become an increasingly important determinant of individual financial well-being.

1.2 Current challenges in investment

Traditional approaches to investing rely heavily on historical data, manual review, or platform-specific analysis that does not take into account all of the key asset classes at once, therefore rendering these traditional approaches ineffective at reflecting changing market conditions and the diverse preferences of individual investors. Individual investors in India face additional practical obstacles: financial platforms remain fragmented across asset classes, with dedicated applications for stocks, mutual funds, real estate, and cryptocurrency operating largely in isolation from one another.[2-8] Each asset class also exhibits distinct behavioral and statistical characteristics, so models developed for one asset class, such as real estate[2-4,9] or gold,[5-7] rarely transfer well to another. Compounding this fragmentation, a large proportion of Indian retail investors have limited financial literacy and limited access to professional advisory services, leaving them without the intelligent decision-support tools needed to compare opportunities across the full range of available asset classes.

1.3 Role of artificial intelligence in financial services

The advent of artificial intelligence (AI), machine learning (ML), and other related technologies provides a promising opportunity to improve the quality of investment analysis through data-driven forecasting, automated decision support, and personalized recommendations. Prior work has shown that ML and deep learning models can outperform traditional statistical approaches for price and return prediction across individual asset classes, including regression and ensemble methods for real estate and gold,[2-4,9] deep learning and hybrid architectures for time series forecasting,[5-7,9] sentiment-aware models that incorporate news and social media signals,[8,10] and reinforcement-learning-based systems for portfolio optimization and personal finance management.[11,12] While there are numerous models being utilized in financial applications that show great promise, very few models are currently used in practice. This limited use is attributed to the fact that there are no integrated platforms that allow users to access multiple asset classes in real time. The current state of fragmentation in the Indian financial technology landscape, combined with low levels of financial literacy, highlights the need for a comprehensive, user-centric investment guidance system.

1.4 Research gap

A closer look at the literature reveals that most AI-driven financial forecasting research is confined to a single asset class: real estate,[2-4,9] gold,[5-7] or cryptocurrency,[10,11] with relatively few studies addressing multiple asset classes within one unified framework.[1] Even when multiple-asset or portfolio-level approaches exist,[1,12] they tend to sacrifice interpretability for predictive performance, rely on narrow sentiment sources such as a single social media platform,[10] or remain largely conceptual without deployment or validation against real, live markets.[13] A detailed discussion of these gaps, organized by modeling approach, is presented in Section 2.6. In summary, no existing system combines multi-asset price forecasting, news-based sentiment analysis, and natural-language advisory generation within a single, deployed, user-facing platform tailored to the Indian market.

1.5 Proposed methodology

To address this gap, this research proposes an AI-supported unified investment guidance platform that integrates five major asset classes-stocks, mutual funds, gold and metals, real estate, and cryptocurrency-into a single, user-friendly framework. The platform combines historical and real-time data collected through API, web scraping, and public datasets with a multimodel machine learning and deep learning pipeline: Random Forest regression is applied to structured, slower-moving datasets such as gold and real estate, whereas LSTM networks are used for the more volatile, sequentially dependent time series of stocks, cryptocurrency, and mutual fund NAVs. A FinBERT-based sentiment analysis module processes financial news headlines to capture market mood, and a large language model (LLM)-based advisory layer converts the combined analytical and sentiment outputs into personalized, natural-language investment guidance. The full methodology is described in Section 3.

1.6 Key contributions

The key contributions of this research are as follows:

- Integration of five major asset classes-stocks, mutual funds, gold/metals, real estate, and cryptocurrency-within a single analytical and advisory platform, rather than the single- or dual-asset focus of prior work. [2-10,13]
- A domain-tailored multi-model forecasting architecture that pairs a random forest model with structured, low-frequency asset data (gold, real estate) and LSTM networks with high-frequency, sequential asset data (stocks, cryptocurrency, mutual funds).
- A FinBERT-based sentiment analysis module that draws on diverse financial news sources rather than a single social media platform, addressing a limitation identified in prior sentiment-based forecasting research.[10]
- An LLM-based advisory layer built on the NVIDIA Nemotron model that converts quantitative forecasts and sentiment signals into interpretable, natural-language investment guidance, addressing the performance–interpretability trade-off noted in prior deep learning studies.[1]
- A working, deployed platform validated against baseline investment strategies (Naïve Forecasting and Simple Moving Average), rather than a purely theoretical model, addressing the deployment gap identified in the literature.[11]

1.7 Paper organization

The remainder of this paper is organized as follows. The rest of Section 2 presents a detailed review of the literature across five methodological categories, identifies the research gap in greater depth, and differentiates the proposed platform from existing investment applications. Section 3 describes the methodology, including the data collection, model design, sentiment analysis module, AI advisory layer, data visualization, and system architecture. Section 4 presents the evaluation metrics and experimental results for each asset class and concludes with a discussion of their implications in Section 5. Section 6 concludes the paper and outlines directions for future work.

2. Literature review

The application of artificial intelligence (AI) and machine learning (ML) in financial forecasting has developed into an area that is among the most active and popular in computational finance. A brief review of some of the major contributions for gold, real estate, stocks and cryptocurrencies is presented below.

2.1 Traditional and regression-based models

Traditional and regression-based approaches, including linear regression, multiple regression, and tree-based regressors such as XGBoost and Random Forest -form the baseline methodology for structured, tabular financial datasets such as year-by-year real estate prices and daily gold prices. These models typically use price history together with location-based or macroeconomic features as predictors and are typically evaluated using the coefficient of determination (R^2) alongside error-based metrics such as the RMSE and MAE. A summary of the traditional and regression-based machine learning methods applied to the prediction of gold and real estate prices can be seen in Table 1.

2.2 Ensemble and hybrid techniques

Ensemble and hybrid techniques combine multiple base learners-for example, stacking regression models with boosting algorithms or pairing deep neural networks with traditional machine learning models-to reduce forecasting error and capture relationships that a single algorithm may miss. These approaches are typically applied to real estate, multi-asset portfolios, and broader financial market datasets and are evaluated using the same error-based metrics (RMSE, MAE) and R^2 used for single-model approaches, alongside portfolio-level metrics such as risk-adjusted return where applicable. Table 2 shows some examples of different ensemble methods and hybrid approaches that can be used for financial forecasting.

Table 1: Traditional and regression-based models

Authors (Ref.)	Asset	Key Findings	Limitations
Maloku *et al.*[2]	Real Estate	They found that random forest algorithms could be bet- ter at predicting real estate prices then linear regression algorithms when they used location-based features	Their sample size was limited and their approach is static.
Srikanth and Ramchander [3]	Real Estate	Srikanth and Ramchander's use of xgboost to predict real estate prices had an R^2 of 0.934 which is quite high	They did not incorporate any dynamic macroeconomic indicators into their model.
Satyam and Suresh[4]	Real Estate	Satyam and Suresh were able to demonstrate very good real-time prediction capabilities in terms of predicting real estate prices (R^2 = 0.9012)	Their dataset was static and did not contain as many con- textual indicators as it could have.

Table 2: Ensemble and hybrid techniques

Authors (Ref.)	Asset	Key Findings	Limitations
Deepu *et al.*[9]	Real Estate	By combining multiple models together, Deepu *et al.* were able to reduce errors in their forecasting	This approach does increase the complexity of the overall model.
Mahajan[1]	Multi-Asset	Mahajan's results indicated that deep neural network models generated returns that were significantly greater than those of tra- ditional machine learning models when applied to multii-asset portfolios	This approach reduces the interpretability of these models.
Jay *et al.*[14]	Cryptocurrency	Jay *et al.* applied a stochastic neural network architecture to model the inherent randomness of cryptocurrency price movements, improving prediction accuracy over deterministic neural network baselines.	Manual hyperparameter tuning of the model can be inefficient.

2.3 Deep learning and time series forecasting

Deep learning approaches, particularly LSTM networks and their bidirectional variant (BiLSTM), use gated recurrent architectures to capture long-range temporal dependencies in sequential financial data. These models are trained on historical price sequences-typically using a fixed look-back window-for assets such as gold, silver, and cryptocurrency and are evaluated using the RMSE, MAE, and R^2 computed on a held-out test period. Table 3 provides a detailed view of how deep learning methods are being used to forecast time series data using LSTMs and BiLSTMs.

2.4 Sentiment analysis and external data integration

Sentiment analysis approaches extract investor mood from unstructured text-typically social media posts or news headlines-using either lexicon-based scoring or pretrained language models and combine the resulting sentiment score with historical price data as an additional predictive feature. These hybrid sentiment-price models are typically applied to cryptocurrency and real estate stock forecasting and are evaluated by comparing prediction accuracy (e.g., Root Mean Square Error (RMSE) and directional accuracy) against an equivalent model trained on price data alone. Table 4 presents a summary of the literature that examines the use of sentiment analysis for predicting stock prices.

2.5 Reinforcement learning and practical applications

Reinforcement learning (RL) approaches frame investment decision-making as a sequential decision

Table 3: Deep learning and time series forecasting

Authors (Ref.)	Asset	Key Findings	Limitations
Zangana & Obeyd[5]	Multi-asset	Zangana & Obeyd found that Bi-LSTM models were more effective than LSTM models at forecasting multivariate data with an R^2 of 0.95	They did not explore other external indicators in addition to historical data.
Amini & Kalantari[6]	Gold & Silver	Amini & Kalantari found that they were able to capture the extreme movements in the long-term data of gold and silver prices using their deep learning methodology	The methodology they developed is computation-ally expensive.
Nagata *et al.*[7]	Gold	Nagata *et al.* found that an approach that combined both the Gold and Daily data provided the best forecast of future stock prices	Did not fully utilize either the volume or macro factors.
Tanwar *et al.*[11]	Crypto (Altcoins)	Tanwar *et al.* used a technique known as dependency to improve their ability to predict the price of Altcoin cryptocurrencies based on the price of Bitcoin.	It had model complexity and scalability issues

Table 4: Sentiment analysis and external data integration

Authors (Ref.)	Asset	Key findings	Limitations
Arslan[10]	Crypto	His study demonstrated that incorporating sentiment from social media posts resulted in a significant improvement in his model's ability to predict cryptocurrency prices relative to a model that only utilized historical price data.	The source of the sentiment information in this study was limited to Twitter and only utilized data from a relatively short period of time.
Fan & Chen[8]	Real Estate	The authors reported that their model resulted in higher prediction accuracy than a model that did not utilize sentiment.	The model may not be applicable to other markets because it was trained using data collected in the Chinese stock market.

problem in which an agent learns an asset-allocation or budgeting policy by maximizing a reward signal-typically portfolio return-through repeated interaction with historical or simulated market data. These approaches are typically evaluated by benchmarking backtested portfolio returns against a market index or a rule-based baseline. Table 5 lists some examples of how reinforcement learning can be applied to develop AI systems that support decision-making in the area of personal finance.

The literature indicates a movement toward the use of ensemble and deep learning techniques for modeling and prediction. In terms of India's investment landscape, there appears to be a common theme among the literature regarding the need for a unified platform for providing investment advice by combining the various types of assets into a single interface, using external economic indicators and balancing predictive accuracy with interpretability.

2.6 Identified research gap

Several research areas have been identified as gaps on the basis of the collective findings of the above literature:

- Limited datasets - Most studies use relatively small, static datasets that do not accurately reflect the complexities of real-world situations.
- Omission of key external factors-Macroeconomic trends, liquidity and trading volume-is typically omitted from models when determining how they may impact investment decisions.
- Oversimplification of models - While some models are multivariate and consider relationships between different asset classes, many remain univariate and do not account for the interplay between related assets.
- Limitations of sentiment analysis - Sentiment analysis models are limited because of their reliance on narrow sources (e.g., Twitter), which limits their ability to provide robust sentiment-based models.
- Inefficiencies in optimization - Hyperparameters are manually tuned, and simplifying assumptions are made to optimize models, limiting their overall accuracy.
- Trade-offs between performance and interpretability - Highly complex models provide very little transparency and therefore diminish investors' confidence in using the results in real-world applications.
- Gaps in deployment - many models developed exist only in theory; few have been deployed into live markets to validate their effectiveness.
- Focus on short-term forecasts - Most models focus on predicting investment outcomes over the short-term horizon, making them less relevant to investors who make investment decisions based on longer-term strategies.
- Validation of profitability - Few studies evaluate the practical profitability of investment portfolios and/or investor outcomes on the basis of the investment recommendations generated by the model.

2.7 Differentiation from existing platforms

Popular investment analysis platforms do not include all major asset classes; most of them focus on only 1–3 asset

Table 5: Reinforcement learning and practical AI applications

Authors (Ref.)	Domain	Key contributions	Limitations
Leung *et al.*[12]	Portfolio Opt.	Leung *et al.* developed a model that utilized reinforcement learning to deter- mine optimal asset allocation portfolios. Their model produced returns that exceeded those of the S&P 500 index during backtesting.	They also noted that liquidity constraints and long-term market risks are not well represented by the model.
Deepthi *et al.*[13]	Personal Finance	Deepthi *et al.* developed a system that utilized reinforcement learning to assist individuals with budgeting and managing their portfolios.	The system remains largely conceptual and lacks empirical evidence that demonstrates its effectiveness in real-world markets.

classes, such as Groww (Stocks, Mutual Funds, Metals),MagicBricks (Real Estate), and Binance (Crypto). Additionally, they do not combine technologies such as market sentiment analysis, an AI advisory section, future price prediction, or cross-asset comparison within a single platform.

The proposed system directly addresses the abovementioned gaps. The proposed solution integrates five major asset classes-stocks, mutual funds, metals, real estate, and cryptocurrency-under one analytical roof. It employs different ML-based future prediction models (LSTM and random forest), incorporates a sentiment analysis module drawn from diverse financial news sources beyond social media alone, and includes an additional LLM-based advisory layer for personalized investment guidance.

2.8 Scope of the study

This research investigates the development and implementation of an artificial intelligence (AI)-driven unified investment guidance platform that can support the needs of the Indian market. This proposed system will incorporate multiple asset classes, including but not limited to stocks, mutual funds, real estate, cryptocurrencies and metals, within a single analytical structure. The system uses supervised machine learning models to develop classification models, deep learning-based time series models to predict future price movements, and both financial news sentiment analysis and a large language model–based advisory layer to generate personalized, interpretable and data-driven investment recommendations. The platform focuses on generating medium- to long-term forecasts and providing individual and comparative assessments of asset classes rather than facilitating high-frequency trading.

3. Methodology

3.1 Data collection

This dataset is an aggregation of five major asset types, namely, Stocks, Mutual Funds, Real Estate, Cryptocurrency and Metals. Additionally, they are based on multiple time frames and provide a foundation for analyzing investments.

Stock price data were obtained from Alpha Vantage's API and included real-time stock prices gathered through web scraping on Google Finance using Selenium and Beautiful Soup. Cryptocurrency data from the past 10 years were retrieved from the CoinGecko API. NAV data for mutual funds were collected from the Association of Mutual Funds in India (AMFI), utilizing the mftool Python library.

Data concerning real estate (from 2015–2025) were collected from MagicBricks; specifically, the costs of real estate and yields were based upon localities. Historical past 10 years of gold price data were collected from Kaggle, while real-time gold and silver prices were collected from The Economic Times. Historical data concerning the price of silver were collected from SilverPrice.org. Table 6 summarizes the data type, source, and coverage period used for each asset class.

3.2 Data splitting

The train/test split strategy was tailored to each asset class's dataset frequency and structure rather than a single split applied uniformly across all the models. For assets with daily historical –data-gold and mutual fund – NAVs-training and test sets were split chronologically, with approximately 80% of each dataset used for training and the remaining 20% (all records after a fixed cutoff date) reserved for testing, to preserve temporal order and prevent data leakage. For stock price prediction, the model was trained on historical data from a representative subset of NSE-listed companies and evaluated out-of-sample on AAKASH. NS, a company not included in training, over a fixed forward test window (see Section 4.2). For cryptocurrency price prediction, the model was trained on historical data up to a fixed date and evaluated over a distinct, held-out forward test window (see Section 4.5). Given its yearly resolution dataset (11 observations from 2015–2025 for the evaluated locality), an 80/20 random train-test split was used (8 training and 3 test observations) for the real

Table 6: Data sources and types by asset class

Asset class	Data type	Source	Coverage/Frequency
Stocks	Real-time and historical OHLCV prices	Alpha Vantage API; Google Finance (via Selenium and Beautiful Soup web scraping)	Real-time, daily
Cryptocurrency	Historical daily price data	CoinGecko API	Past 10 years (2015–2025)
Mutual Funds	Net Asset Value (NAV)	Association of Mutual Funds in India (AMFI), via the mftool Python library	Daily NAV history
Real Estate	Locality-wise average price per sq. ft.	MagicBricks	2015–2025 (yearly)
Gold	Historical daily closing price	Kaggle	Past 10 years (2015–2025)
Gold and Silver	Real-time spot price	The Economic Times	Real-time
Silver	Historical price data	SilverPrice.org	Historical

estate model; because of the very small sample size at yearly resolution, this split is not chronological, and the resulting test-set performance should be interpreted with this limitation in mind (see Section 4.3).

3.3 Supervised learning algorithms

Several supervised learning algorithms have been implemented to evaluate and predict investment returns across multiple categories of assets. These include:

3.3.1 Structured data models

- Random Forest Regressor utilized for gold price prediction when nonlinear associations exist between variables in a structured dataset. It is also used to forecast real estate prices by utilizing historical year-by-year price trends for individual localities. For gold, the model used the past 10 years (2015–2025) of gold price data as input features, with 100 decision trees (n_estimators = 100). The dataset was split chronologically into records up to 2024 for training and beyond 2024 for testing to preserve temporal order and prevent data leakage.
- Long short-term memory (LSTM) networks are applied to stock, mutual funds, and cryptocurrency data to learn sequential patterns in time series data. Daily closing price data were fetched via the Alpha Vantage, mftool, and CoinGecko APIs. Each model consisted of a single LSTM layer with 50 units and a dense output layer trained using the Adam optimizer for 200 epochs. The look-back window was tuned to each asset's data characteristics: a 90-day window for stock price prediction and a 60-day window for both mutual fund NAV and cryptocurrency price prediction.

All the models were tuned for specific domains to improve the overall forecasting accuracy and flexibility of the models to support the unified investment advisory framework.

3.4 Sentiment analysis module

In addition to quantitative forecasting, a sentiment analysis module was built to understand how news and media coverage can influence investor perception across different asset classes.

For sentiment scoring, a lexicon-based approach is used where each token is matched against a list of predefined positive and negative financial keywords. In addition, the ProsusAI/FinBERT model, which is a BERT model specifically trained on financial text, yields more contextually accurate sentiment classifications than simple keyword matching does.

The net sentiment score is calculated as follows:

$$\text{Sentiment Score} = \frac{\text{Positive Words} - \text{Negative Words}}{\text{Total Words}} \quad (1)$$

Scores are classified as positive, neutral or negative. The positive, neutral and negative classifications in Fig. 1 represent the prevailing investor sentiment for a specific asset class or sector. The combined sentiment index is intended to provide additional context for short-term price movements, complementing the quantitative forecasts; a comparative evaluation of forecasting accuracy with and without sentiment features was not performed as part of this study.

3.5 AI advisory integration

AI Advisory uses a decision support tool that transforms large amounts of information into personal investment advice presented as natural language. It uses the NVIDIA Nemotron Nano 9B V2 model provided by the OpenRouter API to provide a bridge from raw analytical output to actionable user insights. The workflow is as follows:

Input capture: A user requests investment advice via a JSON-formatted string (i.e., "Should I invest in TATA Motors?").

Query formatting: The query is converted into a formal financial format with the following structure: Introduction; Market Overview; Pros and Cons; and Bottom Line.

...

	headline	positive_words	negative_words	sentiment
0	Vodafone Idea clarifies on 'AGR relief' report...	2	0	Positive
1	Vodafone Idea Share: वीआई- वोडाफोन आइडिया शेयर...	0	0	Neutral
2	Stocks to Watch Today: HUL, Vodafone Idea, Pin...	0	0	Neutral
3	Filatex Fashions Sees Exceptional Trading Volu...	1	0	Positive
4	Easy Trip Planners Stock Falls to 52-Week Low	0	6	Negative
5	Multibagger jewellery stock under ₹20 rallies ...	2	0	Positive
6	YES Bank shares may see $250 mn inflows on Nif...	1	0	Positive
7	HAL Share Price Today: Record Order Book and S...	4	0	Positive

Fig. 1: Sentiment analysis output with positive and negative word counts.

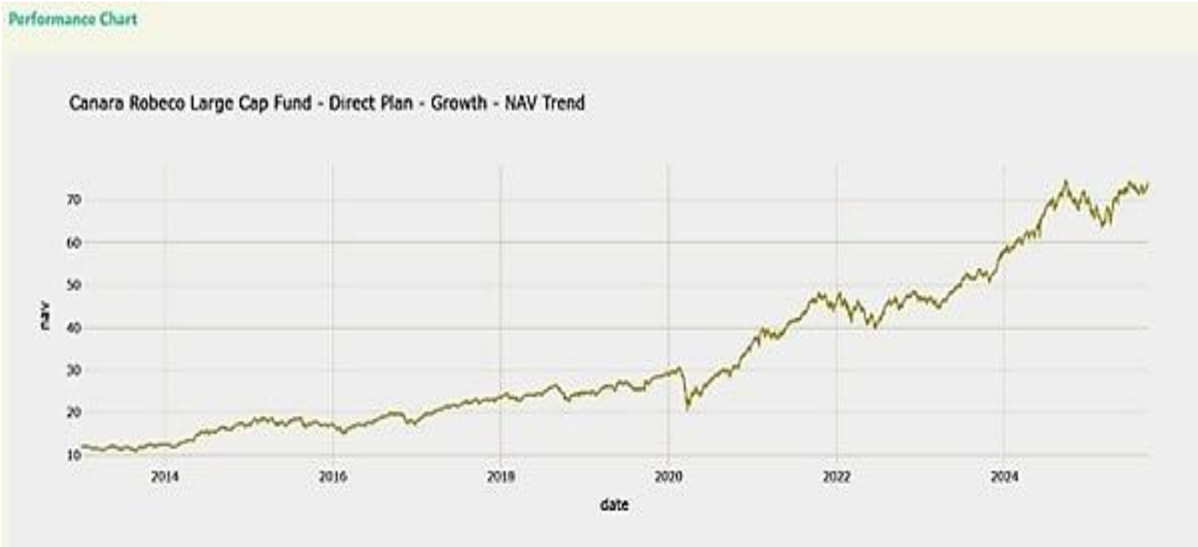

Fig. 2: NAV trend visualization (Canara Robeco Large Cap Fund).

Model request: The system sends a secure request to the model, requesting that it generate a markdown-formatted response on the basis of real-time financial information.

Response processing: The system cleanses the response and formats it as a JSON-formatted string and displays it on the dashboard.

This integration is intended to transform raw sentiment indicators and predictions into understandable advisory statements, with the goal of increasing the accessibility of the platform's outputs for users who are not able to interpret raw statistical metrics directly. This capability provides interactive, rationalized investment suggestions alongside traditional numerical forecasting tools; a formal evaluation of its effect on user decision-making or perceived credibility was not conducted in this study (see Section 5).

3.6 Data visualization

The proposed platform will present its analysis in two ways: tables and graphs. Graphs can also help create an intuitive understanding of how an asset has performed historically and is expected to perform in the future by displaying its past and future values on a line chart.

3.7 Graphical representation

The line charts are based on the date and net asset value (or price movement). The system shown in Fig. 2 uses an interactive plot to show the data against net asset value (nav) and has a dark theme to allow better visibility and readability.

3.8 System architecture

The system architecture, as shown in Fig. 3, is focused on providing an integrated approach to combining multiple data sources and utilizing an artificial intelligence-driven analytical model designed to provide real-time, individualized and data-driven investment recommendations for all asset classes.

The platform is built around a Flask-based backend that acts as the central controller connecting all the components. On the AI side, it manages three modules-

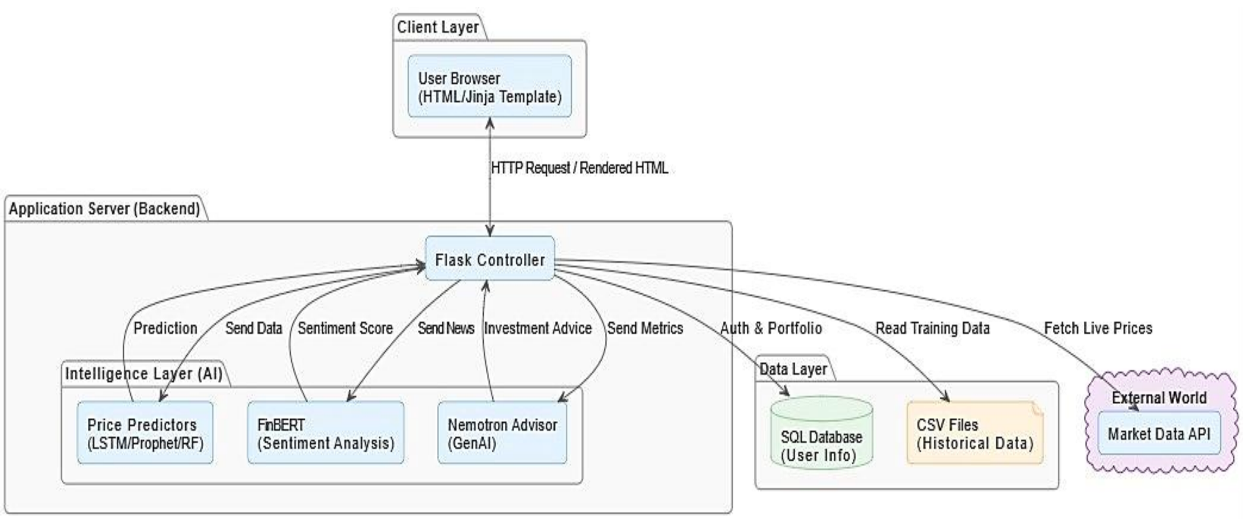

Fig. 3: System architecture of the proposed ai-driven unified investment guidance platform.

the price prediction models (LSTM and Random Forest), the FinBERT sentiment analysis module, and the Nemotron GenAI advisor that generates natural language investment advice. On the data side, it reads user and portfolio information from an SQL database, pulls historical price data from CSV files, and fetches live market prices from external APIs. All processed outputs are then rendered and delivered to the user through a browser-based HTML interface.

For reproducibility, each of the two prediction model families (LSTM and Random Forest) is trained independently on the corresponding asset-class dataset described in Section 3.1, using the training/test split and hyperparameters detailed in Section 3.3, and is loaded by the Flask backend to serve predictions on request rather than being retrained for every user query. When a user requests a forecast for a given asset, the Flask controller identifies the corresponding asset class, retrieves the relevant historical data from the CSV files or live prices from the external market APIs, applies the same pre-processing steps used during training-chronological splitting for random forest-and min–max normalization and windowing for LSTM-and returns the resulting prediction to the front end. The FinBERT sentiment module and the Nemotron-based GenAI advisor operate as separate analytical services within the same backend: FinBERT scores incoming financial news headlines using the process described in Section 3.4, while the Nemotron model is queried through the OpenRouter API using the structured prompt format (Introduction; Market Overview; Pros and Cons; Bottom Line) described in Section 3.5. User and portfolio information, used to personalize recommendations, is stored in and retrieved from an SQL database, while all analytical outputs-price predictions, sentiment scores, and advisory text-are combined and rendered through a Jinja-templated, browser-based HTML interface. This modular design, in which each AI component is trained and served independently, allows the forecasting, sentiment analysis, and advisory layers to be reproduced, retrained, or replaced without requiring changes to the rest of the platform.

4. Results

4.1 Evaluation metrics

To assess the predictive performance of each forecasting model, three standard regression evaluation metrics are used throughout this study: the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). For a test set of n observations, where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the actual values, these metrics are defined as follows.

The Root Mean Square Error (RMSE) measures the average magnitude of the prediction error in the same unit as the target variable, with larger errors penalized more heavily:

$$\mathrm{RMSE} = \sqrt{((1/n) \cdot \Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2)} \tag{2}$$

The Mean Absolute Error (MAE) is a measure of the average absolute difference between the predicted and actual values and is less sensitive to large outliers than the RMSE is:

$$\mathrm{MAE} = (1/n) \cdot \Sigma_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

The Coefficient of Determination (R^2) is the proportion of variance in the actual values explained by the model, where an R^2 closer to 1 indicates stronger predictive performance:

$$R^2 = 1 - [\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2 / \Sigma_{i=1}^{n} (y_i - \bar{y})^2] \tag{4}$$

In addition to these standard regression metrics, Section 4.7 also reports Directional Prediction-whether a model correctly predicted the direction (upward or downward) of price movement over the evaluated test window-to complement the error-based regression metrics above with a decision-relevant, direction-based comparison.

4.2 Performance of stock forecasts by LSTMs

A predictive model is trained using historical stock price information from a representative subset of companies listed on the National Stock Exchange (NSE) rather than a company-specific model so that a single generalized model can be built across the many companies available. Using the Open, High, Low, Close, and Volume information as input parameters, the model has a 90-day look back time frame to represent both short-term and medium-term price movements. To avoid the possibility of overfitting, an early stop in the training process is used. AAKASH. NS, the stock used for out-of-sample evaluation in this section, was not included in the training set; the reported test performance therefore reflects the model's ability to generalize to a company it had not previously seen, rather than performance on data used during training.

To assess the predictive performance of the model's predictions over a one-month testing period (September 22nd, 2025 through October 22nd, 2025), the model's predictions were assessed using common metrics for regression models. For this model, the results were as follows:

- Root Mean Square Error (RMSE): 0.1487
- Mean Absolute Error (MAE): 0.1181
- Coefficient of Determination (R^2): 0.78

In Fig. 4, we can see how well the model predicts the stock prices (orange dashed line) compared to the actual prices of the stock (green line). Although some small variations in stock prices occur during sudden changes in the market, it is clear that the model is able to track the direction of the stock price for the majority of the time.

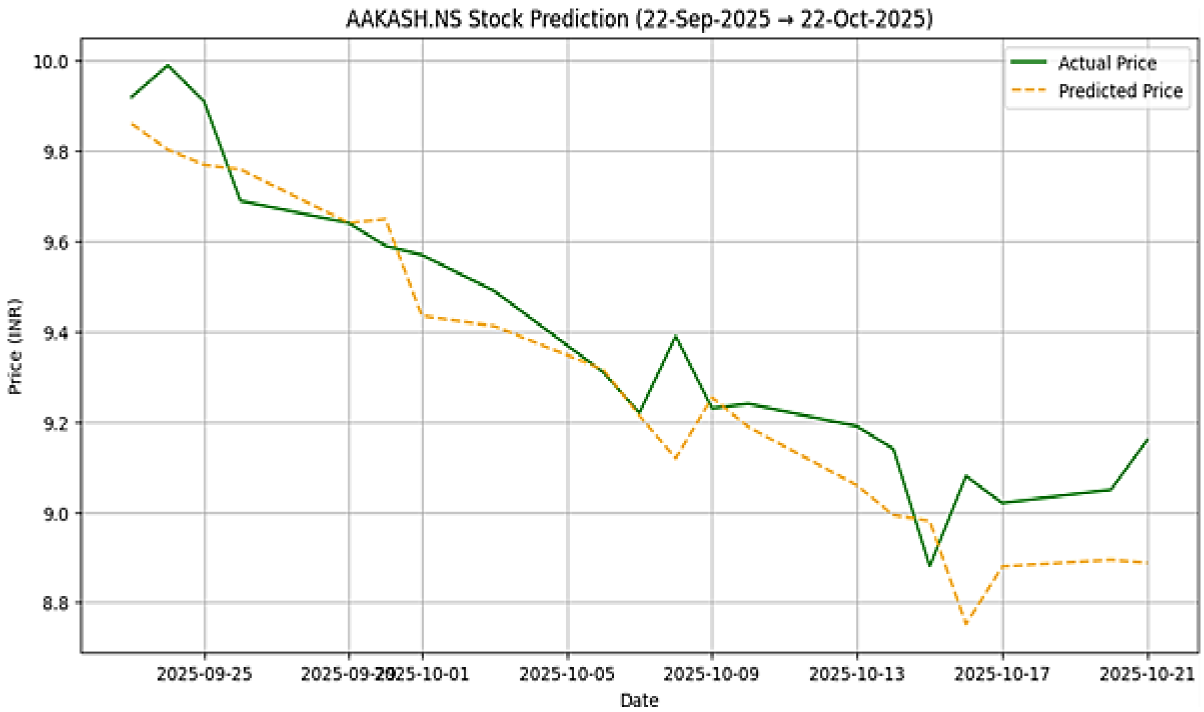

Fig. 4: AAKASH. NS stock future prediction trend.

This demonstrates the potential for the use of an LSTM architecture for the short-term prediction of stock prices. The results show that LSTM models are capable of finding patterns within a dataset (such as temporal dependencies or nonlinear relationships) that can be useful in predicting the near-term prices of stocks. As this evaluation is based on a single out-of-sample stock (AAKASH. NS) over a one-month test window, these findings should be regarded as a preliminary indication of the model's generalization ability rather than a general claim of predictive performance across the broader stock market.

4.3 Forecasting prices for real estate in Mumbai

Price forecasting for the Mumbai real estate market was completed using a random forest regressor trained on yearly average price data for Andheri West from 2015 to 2025. The model used both the year and a nonlinear feature ($year^2$) derived from the year to reflect long-term trends of increasing prices. The optimal combination of hyperparameters (number of trees, maximum tree depth, and minimum samples per split/leaf) was identified using GridSearchCV with 3-fold cross-validation, optimizing for R^2.

Given the yearly resolution of the underlying dataset, the 11 available observations for Andheri West (2015–2025) were split using an 80/20 random train-test partition (8 training and 3 test observations); the model was fit on the training partition, and all reported metrics reflected the performance on the 3 held-out test observations rather than the in-sample training fit. On this test set, the model achieved an MAE of 1330.74 INR, an RMSE of 1434.79 INR, and an R^2 of 0.88, indicating that 88% of the variance in property prices on the held-out observations is accounted for by the model. Given the very small test-set size, these results should be interpreted as indicative rather than statistically robust.

The comparison of the predicted price trend from the fitted random forest model to the actual price trend for Andheri West across the full 2015–2025 dataset period is shown in Fig. 5. The close alignment between the predicted and actual trends supports the model's ability to capture the long-term real estate price trend, which is consistent with the R^2 of 0.88 reported above. These results suggest that this method can be applied effectively for modeling real estate price trends for specific urban submarkets.

4.4 Gold price forecasting

Random forest regression was used to predict the gold closing price on each trading day (GLD) using historical data. To increase the accuracy of the forecasts, additional temporal information, such as the year, month, day and price of the previous day, was also included in the model. Consistent with the chronological split described in Section 3.3, the data were split into two parts: a training set up to 2024 and a testing set post-2024 to evaluate how well the model would generalize.

The model's predictive performance was evaluated on the post-2024 test set using the regression metrics defined in Section 4.1:

- Coefficient of Determination (R^2): 0.92
- Mean Absolute Error (MAE): 0.9011

An R^2 of 0.92 indicates that the model explains more than 92% of the variance in daily gold closing prices, reflecting strong agreement between the predicted and actual price

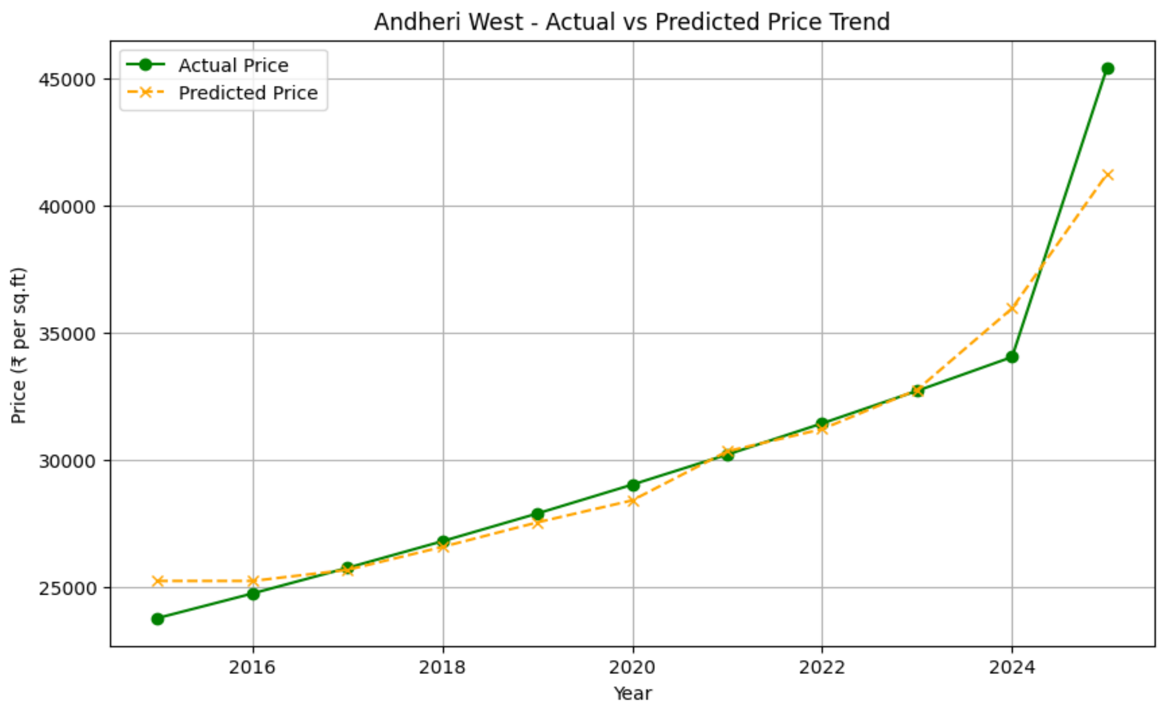

Fig. 5: Andheri West - Actual vs. Predicted price trend.

series shown in Fig. 6. Building on this fit, the model was also used to generate a 30-day forecast of future price trends (Fig. 7), allowing proactive decision-making.

4.5 Cryptocurrency price prediction

To evaluate the ability of the proposed LSTM model to forecast cryptocurrency prices, historical Bitcoin price data retrieved via the CoinGecko API (see Section 3.1) were used to train and test the model, following the same LSTM architecture and chronological train/test split described in Section 3.3, with a 60-day look-back window consistent with the window used for mutual fund NAV prediction. The model's predictions were evaluated over a test period from January 22, 2026, to March 22, 2026, and compared against a naïve forecast baseline to assess the predictive value beyond a simple persistence model. The model achieved a coefficient of determination (R^2) of 0.80 on the test set, indicating that the LSTM captures a substantial share of the variance in Bitcoin's price movements over the evaluation period, including the sharp downward correction observed in early February 2026 and the subsequent partial recovery.

Fig. 6: Actual vs. predicted gold price (test set).

Fig. 7: Gold price future prediction trend analysis.

In Fig. 8, the predicted price (red line) is compared against the actual price (blue line) and the Naïve forecast baseline (dashed gray line) for the test period. Compared with the naïve baseline, the LSTM model tracks the overall downward and recovery trend of the actual price series more closely during the sharpest price movements, particularly around the early February decline, although-which is consistent with the stock and mutual fund results in Sections 4.2 and 4.6-some lag is visible during the most volatile short-term fluctuations. This is consistent with the broader finding across asset classes that LSTM-based models are well suited to capturing directional and medium-term trends in volatile, high-frequency financial time series, even where they cannot fully anticipate sudden short-term shocks.

4.6 Mutual funds prediction

To assess how well the proposed LSTM model performs in predicting mutual fund net asset value (NAV), we use the Canara Robeco Large Cap Fund–Direct Plan (Growth)

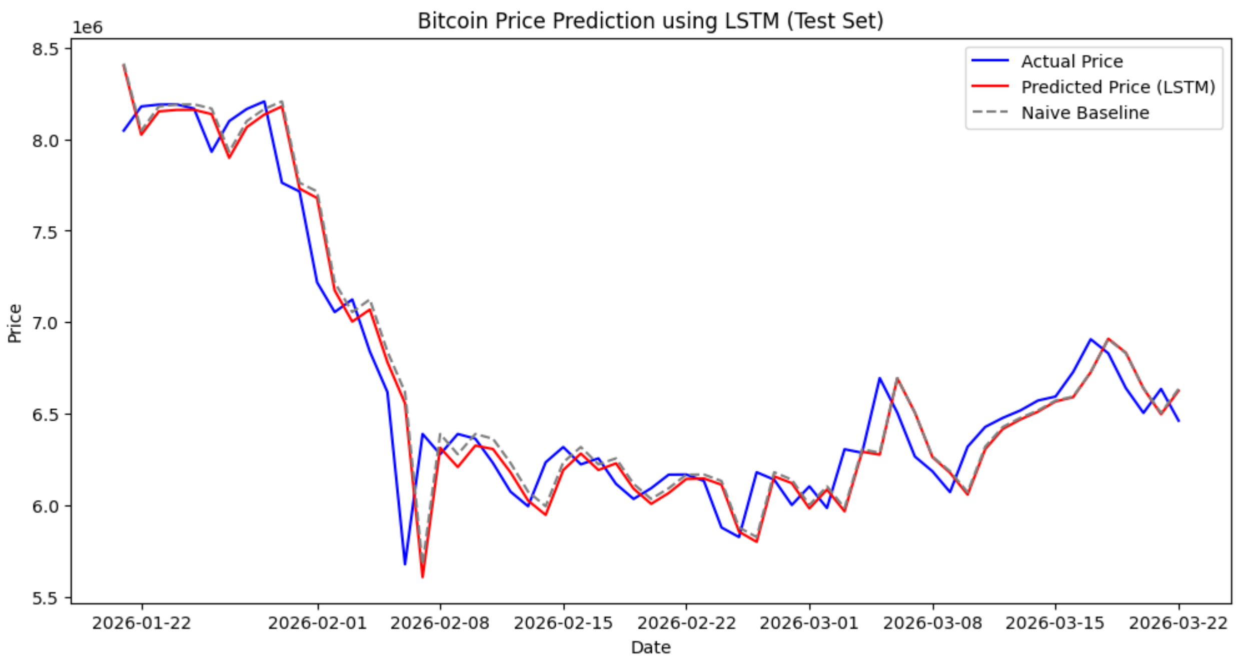

Fig. 8: Bitcoin price prediction using LSTM (test set).

dataset. We used the historical NAV data to normalize it to a range from zero to one using the min–max method. We also used a 60-day window in time to measure the relationship between the input features and output features in terms of time dependence. We trained our model using 80% of the data and tested it using the other 20%.

Our model performed satisfactorily in terms of prediction accuracy. Here are some of the key statistics about how well our model did:

- R^2 score: 0.99
- MAE: 0.7398
- RMSE: 0.9575

In Fig. 9, we compare the actual NAV (blue line) to the predicted NAV (red line) for the test period. The red line follows the blue line very closely in terms of the overall trend of the actual NAV; this finding demonstrates that our model is able to learn and identify relationships between the input variables (the time series data) and the output variable (the NAV).

Although minor prediction errors occur during periods of high volatility, the LSTM network effectively captures the long-term growth trend of the mutual fund. Therefore, while we cannot claim to have developed the best possible approach to mutual fund NAV prediction, these results demonstrate that deep learning models such as LSTMs can be useful for financial time series forecasting, including mutual fund NAV prediction; this provides insight into what may happen to the performance of a mutual fund moving forward.

4.7 Comparison with baseline investment strategies

To provide a preliminary assessment of the practical usefulness of the proposed LSTM model, its predictions for AAKASH are presented. NS (Aakash Exploration Services Limited) were compared against two baseline strategies over the test period (22-Sep-2025 to 22-Oct-2025): Naïve Forecasting and Simple Moving Average (SMA-20).

As shown in Table 7, the proposed LSTM model correctly predicted the downward direction of AAKASH. NS over the test period and produced a predicted end price of ₹8.89, compared with the actual end-of-period price of ₹9.16. This prediction was closer to the actual end price than those produced by the Naïve forecast and Simple Moving Average (SMA-20) baselines were. The Naïve forecast, which carried forward the most recently observed price, predicted an upward direction, whereas the SMA-20 baseline showed a delayed response to the downward price movement. These results provide a preliminary comparison of the directional prediction performance between the proposed LSTM model and the selected baseline strategies. However, because the comparison is limited to a single stock and a one-month

Table 7: Comparison of the proposed LSTM model against baseline investment strategies (AAKASH. NS)

Method	Directional prediction	Predicted end price (INR)	Actual end price (INR)
LSTM (Proposed)	Correct (Downward)	₹8.89	₹9.16
Naïve Forecast	Incorrect	₹9.93	₹9.16
SMA-20	Delayed	~₹9.50	₹9.16

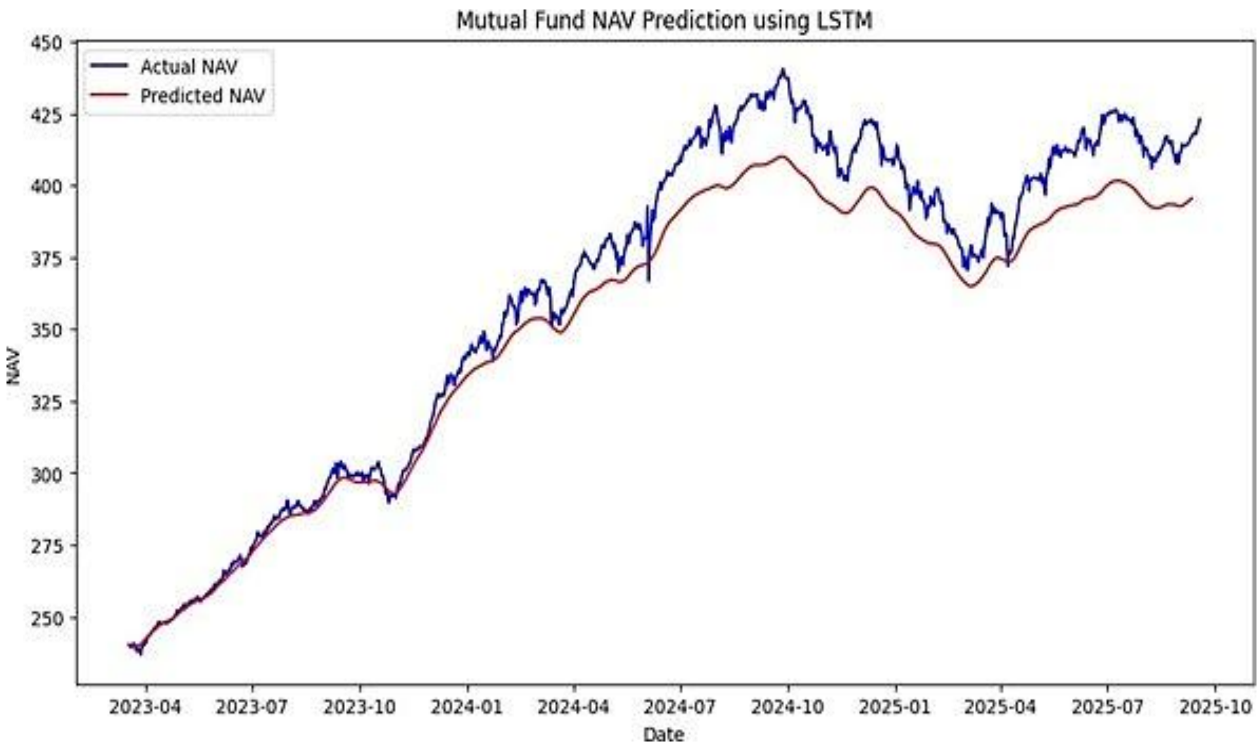

Fig. 9: Mutual fund NAV prediction using LSTM (Canara Robeco large cap fund).

test period, the findings should not be generalized to broader market performance. Further evaluation across multiple securities, asset classes, and longer test periods is needed to assess the generalizability and practical applicability of the proposed approach (see Section 5).

5. Discussion

The experimental results across the five asset classes reveal a consistent pattern: model performance is closely tied to the statistical structure of the underlying asset. The random forest and LSTM models applied to gold (R^2 = 0.92) and mutual fund NAVs (R^2 = 0.99) achieved the strongest fit, which is consistent with the relatively smooth, trend-dominated behavior of these series. The real estate model (R^2 = 0.88) performed similarly well over most of the 2015–2025 period but showed larger deviations toward the end of the forecast window (Section 4.3, Fig. 5), suggesting that longer-horizon forecasts of illiquid, low-frequency markets such as real estate are more sensitive to unmodeled structural shifts than shorter-horizon forecasts are. In contrast, the stock (R^2 = 0.78) and cryptocurrency (R^2 = 0.80) models-both applied to the most volatile and high-frequency asset classes in this study-achieved comparatively lower predictive performance, with R^2 values indicating that the models explain a smaller share of price variance than for the more structured asset classes. This finding is consistent with the broader literature on financial time series forecasting, where higher-frequency, sentiment-sensitive assets are inherently harder to predict than structured, slower-moving assets are.[5-7,9,10]

This pattern has direct implications for the proposed platform’s design: rather than applying a single model architecture uniformly across all asset classes, the results support the platform’s domain-tailored approach of pairing random forest with structured, low-frequency data and LSTM networks with volatile, high-frequency data (Section 3.3). Moreover, the comparison against baseline strategies in Section 4.7 shows that even when absolute error metrics are moderate, as with stocks, the LSTM model correctly predicted the price direction where the naïve and lagging moving-average baselines did not, indicating that the predictive value for real-world investment decisions cannot be judged from R^2 or RMSE alone.

The integration of the FinBERT-based sentiment module and the LLM-based advisory layer (Sections 3.4–3.5) addresses a limitation identified throughout the reviewed literature: that highly accurate models often sacrifice interpretability, reducing investor trust in real-world use.[1] By converting raw sentiment scores and quantitative forecasts into structured, natural-language advisory output, the platform aims to make model outputs more actionable for retail investors without requiring them to interpret raw statistical metrics directly. However, this interpretability layer has not been formally evaluated in this study-for example, through a user study assessing whether the generated advisory text improves investors' decision-making relative to raw forecasts alone-and represents an important direction for future validation.

Several limitations of the current study should also be acknowledged. First, the sentiment analysis module currently draws on financial news headlines rather than a broader set of external indicators, such as macroeconomic data, trading volume, or liquidity measures, that prior work has identified as omitted from many existing models;[2-8,13] incorporating such indicators could further improve forecasting accuracy, particularly for real estate and gold. Second, the baseline comparison in Section 4.7 is limited to a single stock (AAKASH. NS) over a one-month test window; broader validation across a larger set of securities, asset classes, and time horizons is needed before the platform's practical profitability can be more confidently established, addressing a validation gap noted in prior literature. [11,12] Third, while the platform has been implemented and deployed as a working system rather than remaining purely theoretical, addressing another gap identified in the literature review,[11] this study has not yet evaluated the platform under live, real-time market conditions over an extended period. Addressing these limitations-particularly through longer-horizon backtesting, broader asset and instrument coverage, and formal evaluation of the advisory layer's usefulness to end users-represents a promising direction for future work.

6. Conclusion

This study set out to address a specific gap in AI-driven financial technology: the absence of a unified platform capable of guiding individual investors, particularly in India, across multiple asset classes rather than a single market. To this end, in this research, an AI-driven Unified Investment Guidance Platform that integrates five major asset classes-stocks, mutual funds, gold, real estate, and cryptocurrency-is proposed and implemented within a single analytical and advisory framework, combining domain-tailored forecasting models (Random Forest for structured, low-frequency data; LSTM networks for volatile, high-frequency data) with a FinBERT-based sentiment analysis module and an LLM-based natural-language advisory layer. The experimental results demonstrate promising predictive performance across the five evaluated asset classes, although performance varies by asset type and evaluation setting: an R^2 of 0.92 for gold, 0.88 for real estate, 0.78 for stocks, 0.80 for cryptocurrency, and 0.99 for mutual fund NAV prediction. Beyond these standard regression metrics, a preliminary comparison against baseline investment strategies (Section 4.7) limited to a single stock over a one-month test window revealed that the proposed LSTM model correctly predicted the direction of price movement where the naïve and moving-average baselines did not. This is an encouraging early indicator of decision-relevant value, although broader validation across more securities, asset classes, and longer horizons is needed before general conclusions about the platform's practical investment performance can be drawn. The key contributions of this work are the integration of five asset classes into a single platform, a forecasting architecture tailored to the statistical characteristics of each asset class, a sentiment analysis module that draws on diverse financial news sources, and an LLM-based advisory layer that translates quantitative and sentiment-based outputs into interpretable, natural-language investment guidance. Taken together, these contributions address several of the gaps identified in the literature review (Section 2.6), including the lack of unified multiasset platforms, the trade-off between predictive performance and interpretability, and the shortage of deployed, validated systems. From a practical standpoint, the proposed platform contributes to democratizing financial guidance in India by providing individual investors-including those with limited financial literacy or access to professional advisory services-a single, interpretable point of reference for comparing investment opportunities across asset classes. This study has several limitations, discussed in more detail in Section 5, including the platform's reliance on financial news sentiment rather than a broader set of macroeconomic indicators, a baseline comparison limited to a single stock over a short test window, and the absence of a formal evaluation of the advisory layer's usefulness to end users. Future work should extend the baseline validation to a larger and more diverse set of securities and asset classes over longer time horizons, incorporate additional macroeconomic and liquidity indicators into the forecasting models, and conduct a user-centered evaluation of the AI advisory layer to assess its impact on investor decision-making in practice.

CRediT Author Contribution Statement

Mohammad Ali Ansari: Conceptualization, Investigation, Methodology, Software, Writing - Review & editing. **Huzaifa Zahid Husein Shah:** Formal analysis, Investigation, Validation, Writing - Original draft. **Talha Ansari:** Investigation, Visualization. **Aditya Vijay Kamble:** Data curation, Investigation. **Ahlam Ansari:** Supervision; Writing - Review & editing. All authors have read and agreed to the published version of the manuscript.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The datasets used in this study were derived from publicly accessible sources, including the Alpha Vantage API, Google Finance (accessed via web scraping), CoinGecko API, AMFI (via the mftool Python library), MagicBricks, Kaggle, The Economic Times, and SilverPrice.org, as described in Section 3.1.

Conflict of Interest

There is no conflict of interest.

Artificial Intelligence (AI) Use Disclosure

The authors declare that artificial intelligence (AI)-assisted tools were used only for language refinement, grammar improvement, and manuscript structuring purposes during the preparation of this work. All technical content, experimental implementation, results, and interpretations were independently developed and verified by the authors.

Supporting information

Not Applicable

References

[1] K. Mahajan, Transforming financial decision-making with artificial intelligence: a comprehensive study on ai-driven algorithms for investment, trading, and portfolio management, *Journal of Electrical Systems*, 2024, **20**, 5036–5044.

[2] F. Maloku, B. Maloku, A. A. D. Kumar, A. Agarwal, House price prediction using machine learning and artificial intelligence, *Journal of Artificial Intelligence & Cloud Computing*, 2024, **3**, 1–10, doi: 10.47363/JAICC/2024(3)357.

[3] N. Srikanth, M. Ramchander, Machine learning approach for house price prediction, *Juni Khyat*, 2023, **13**, 81-88, available at: https://www.junikhyat.org.in/no_1_Online_23.html

[4] K. Satyam, B. Suresh, Predicting real estate prices: A machine learning approach, *International Research Journal of Modernization in Engineering Technology and Science*, 2025, 7, 5776–5780, available at: https://www.irjmets.com/uploadedfiles/paper/issue_4_april_2025/73009/final/fin_irjmets1745312600.pdf

[5] H. M. Zangana, S. R. Obeyd, Deep learning-based gold price prediction: A novel approach using time series analysis, *Sistemasi: Jurnal Sistem Informasi*, 2024, **13**, 2581-2591.

[6] A. Amini, R. Kalantari, Gold price prediction by a

CNN-Bi-LSTM model along with automatic parameter tuning, *PLOS ONE*, 2024, **19**, e0298426, doi: 10.1371/journal.pone.0298426.
[7] A. B. Nagata, M. S. Hidajat, D. A. Wibowo, W. Widyatmoko, N. B. Yaacob, Predicting gold price movement using long short-term memory model, *Journal of Applied Intelligent System*, 2024, **9**, 19-28, doi: 10.62411/jais.v9i1.10305.
[8] X. Fan, J. Chen, Stock price forecasting in real estate industry based on investor sentiment, *Frontiers in Business, Economics and Management*, 2022, **6**, 54–59, doi: 10.54097/fbem.v6i3.3311.
[9] S. Deepu, A. Verma, A. Raj, M. Kumar, S. Sen, K. Bharat, performance evaluation of real estate price predictor using machine learning, Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024), 2024, doi: 10.2139/ssrn.4856312.
[10] S. Arslan, Bitcoin price prediction using sentiment analysis and empirical mode decomposition, *Computational Economics*, 2025, **65**, 2227–2248, doi: 10.1007/s10614-024-10588-3.
[11] S. Tanwar, N. P. Patel, S. N. Patel, J. R. Patel, G. Sharma, I. E. Davidson, Deep learning-based cryptocurrency price prediction scheme with interdependent relations, *IEEE Access*, 2021, **9**, 143284–143295, doi: 10.1109/ACCESS.2021.3117848.
[12] M.-F. Leung, A. Jawaid, S.-W. Ip, C.-H. Kwok, S. Yan, A portfolio recommendation system based on machine learning and big data analytics, *Data Science in Finance and Economics*, 2023, **3**, 152–165, doi: 10.3934/DSFE.2023009.
[13] Deepthi C. G., J. J. Shetty, I. B. S. Jain, Lakshmi C. R. Kiran Kumar D. R., AI-powered finance management platform, 2026 International Conference on Smart Futuristic Technology, Bengalore, India, 2026, 1-5, doi: 10.1109/ICSFT66733.2026.11506666.
[14] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, M. Alazab, Stochastic neural networks for cryptocurrency price prediction, *IEEE Access*, 2020, **8**, 82804–82818, doi: 10.1109/ACCESS.2020.2990659.