Intelligent Multi-Asset Portfolio Advisory System Using Deep Learning and Natural Language Processing

Mohammad Ali Ansari[1], Huzaifa Zahid Husein Shah[2], Talha Ansari[3], Aditya Vijay Kamble[4], Ahlam Ansari[5]
[1234]*Computer Engineering Department, MHSSCE, Mumbai, India*
[5]*Associate Professor, Computer Engineering Department, MHSSCE, Mumbai, India*

***Abstract*—Retail investors in India face growing difficulty navigating a fragmented and increasingly complex financial landscape. Existing decision-support tools typically address individual asset categories in isolation, offering little integration across multiple investment domains. This paper proposes an Intelligent Multi-Asset Portfolio Advisory System that consolidates five asset classes equities, mutual funds, precious metals, residential real estate, and digital currencies into a single analytical framework. Market data is acquired through public APIs, automated web scraping, and curated open datasets, and subsequently processed using a suite of machine learning and deep learning techniques. Structured asset data for metals and real estate is modelled via Random Forest and Prophet, while Long Short-Term Memory (LSTM) networks perform time-series forecasting for equities, digital currencies, and mutual fund NAVs. A FinBERT-based module performs contextual sentiment classification on financial news, and a large language model layer translates model outputs into plain-language advisory narratives. The system achieved R^2 values of 0.90 for gold, 0.88 for residential real estate, 0.78 for equities, 0.85 for digital currencies, and 0.99 for mutual funds. Experimental results indicate that integrating quantitative forecasting with sentiment-aware natural language generation substantially improves both predictive accuracy and user interpretability, advancing the accessibility of data-driven financial guidance for everyday investors.**

***Index Terms*—Deep Learning, Digital Currencies, Equities Forecasting, FinBERT, LSTM, Multi-Asset Framework, Mutual Funds, Natural Language Advisory, Portfolio Intelligence, Real Estate Analytics**

I. INTRODUCTION

The past decade has seen a dramatic broadening of the investment universe available to Indian retail participants. Individuals can now allocate capital across equities, mutual funds, gold, residential real estate, and digital currencies asset classes that differ substantially in their risk profiles, liquidity characteristics, and sensitivity to macroeconomic shifts.

Despite this variety, the tools designed to assist investment decisions remain compartmentalized. Most platforms address a single asset domain, rely exclusively on historical price data, and offer no personalized guidance. Consequently, a significant proportion of retail investors either act on informal advice or make uninformed choices without fully understanding the instruments they are evaluating.

Artificial intelligence and machine learning have opened new avenues for automated forecasting, pattern detection, and scalable personalized recommendation. Yet the integration of these capabilities within a unified, multi-asset advisory platform tailored to Indian market conditions remains largely unaddressed in the existing literature. This work directly fills that gap.

A. Research Gaps in Existing Literature

A systematic review of prior work identifies several persistent shortcomings. The majority of studies target a single asset class and make no attempt to model cross-asset interactions. Datasets are frequently static and drawn from narrow time windows, limiting generalization to evolving market regimes. Sentiment signals, when incorporated, tend to originate from a single noisy source such as social media. Furthermore, most models sacrifice interpretability for marginal accuracy gains, reducing their practical utility for retail investors who cannot audit raw model outputs.

A consequential disconnect also exists between theoretical model proposals and their actual deployment. Academic literature predominantly addresses short-horizon forecasting, whereas retail investors principally require medium-to-long-term projections to support meaningful financial planning.

B. Differentiation from Existing Platforms

Platforms such as Groww, MagicBricks, and Binance each serve distinct asset categories, yet none unifies all five major classes within a common analytical environment. Critically, no available platform combines price forecasting, sentiment-aware classification, and natural language advisory generation into a single interface.

The system introduced in this paper addresses all three dimensions simultaneously. It deploys asset-specific forecasting models, aggregates sentiment from diverse financial news sources, and employs a large language model to synthesize analytical outputs into actionable, plain-English investment narratives accessible to investors of any background.

II. RELATED WORK

Research on AI-assisted financial forecasting spans a broad methodological spectrum. For residential property valuation, Maloku et al. [2] demonstrated that Random Forest models outperform conventional linear regression when location-specific features are incorporated. Srikanth and Ramchander [3] achieved an R^2 of 0.934 using gradient-boosted trees for property price estimation. Independent corroboration was provided by Satyam and Suresh [4], who reported an R^2 of 0.9012 on a comparable real estate pricing task.

In deep learning for financial time series, Zangana and Obeyd [6] established that bidirectional LSTM architectures outperform standard LSTMs on multivariate forecasting benchmarks, attaining an R^2 of 0.95. Amini and Kalantari [7] applied CNN-BiLSTM models to capture long-horizon volatility patterns in precious metal prices, while Tanwar et al. [1] leveraged inter-asset dependencies between Bitcoin and altcoin markets using LSTM-based networks.

Sentiment-augmented forecasting has also attracted growing research interest. Arslan [11] demonstrated measurable improvement in digital currency price prediction by incorporating microblog sentiment signals. Fan and Chen [9] reported analogous improvements in real estate forecasting when investor sentiment derived from news articles was included as an input feature.

Portfolio optimization research has increasingly adopted reinforcement learning frameworks. Leung et al. [13] developed a policy gradient model that outperformed S&P 500 benchmarks during historical back testing. Deepthi et al. [12] proposed a conceptual AI-assisted personal finance management architecture, though without empirical validation.

Collectively, the literature evidences a strong convergence toward ensemble and recurrent deep learning methods, accompanied by growing interest in multi-modal sentiment integration. The principal gap that persists is the absence of a production-grade platform that unifies these capabilities across all major asset classes with a transparent, user-facing interface the central contribution of this paper.

III. MATERIALS AND METHODOLOGY

A. Data Collection

The platform aggregates data from a diverse range of sources spanning all five asset classes. Equity price data was retrieved from the Alpha Vantage API, with real-time quotes additionally scraped from Google Finance using Selenium and BeautifulSoup. A decade of digital currency historical data was obtained from the CoinGecko API. Mutual fund NAV series were collected from AMFI via the mftool Python library.

For residential real estate, locality-level pricing data from 2015 to 2025 was sourced from MagicBricks, capturing annual average prices per square foot across Mumbai suburbs. Historical gold price data was drawn from Kaggle, with live prices retrieved from The Economic Times, and silver price history from SilverPrice.org.

B. Data Preprocessing and Feature Engineering

All datasets were cleaned, normalized, and partitioned into training and evaluation subsets using an 80/20 ratio. Temporal integrity was strictly maintained for sequential models: all observations preceding a designated cutoff date formed the training set, with subsequent records reserved exclusively for evaluation to prevent data leakage. Continuous numerical features were rescaled using min-max normalization, particularly for LSTM inputs where gradient stability is sensitive to feature magnitude.

C. Forecasting Models

Given the structural diversity of the included asset classes, no single modelling approach was applied uniformly. Instead, each class was paired with the technique most appropriate to its data structure and behavioral characteristics.

A Random Forest regressor was used for gold price forecasting, leveraging temporal features year, month, day alongside lagged closing price values. The ensemble comprised 100 decision trees trained on observations through 2016, with post-2016 data held out for evaluation. For residential real estate, a polynomial regression model tuned via GridSearchCV was applied, with Random Forest serving as a complementary benchmark.

LSTM networks were applied to time-series data for equities, mutual funds, and digital currencies. Each architecture consisted of a single LSTM layer with 50 hidden units followed by a dense projection layer, optimized using the Adam algorithm over 200 training epochs. A look-back window of 60 to 90 days was employed to capture both short- and medium-term temporal dependencies, with early stopping applied to mitigate overfitting.

Prophet was used for locality-level real estate forecasting, fitting annual price trends from 2015 to 2025 and projecting five years forward. Both daily and yearly seasonality components were disabled given the annual resolution of the input data.

D. Sentiment Analysis Module

A two-stage sentiment pipeline was developed to capture how news and media signals may influence investor behavior. The first stage applies a lexicon-based approach, matching financial tokens against curated positive and negative keyword lists to produce a normalized net sentiment score: Sentiment Score = (Positive Words − Negative Words) / Total Words. The second stage applies ProsusAI/FinBERT a BERT variant fine-tuned on financial text to generate contextually sensitive sentiment classifications.

Each news article or headline is ultimately classified as positive, neutral, or negative. The resulting sentiment index informs short-term price adjustments and is surfaced to users as part of the platform's investment summaries, improving transparency.

	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

Figure 1. Sentiment Analysis Output with Positive and Negative Word Counts

Scores are classified as positive, neutral and negative. The positive, neutral and negative classifications in Figure 1 represent the prevailing investor sentiment for a specific asset class or sector. The combined sentiment index is then used to enhance the accuracy of short-term predictions by providing additional insight on what drives those predictions.

E. Natural Language Advisory Layer

Beyond quantitative outputs, the platform incorporates a large language model advisory engine powered by NVIDIA Nemotron Nano 9B V2, accessed via the OpenRouter API. User queries such as "Should I consider Tata Motors for my portfolio?" are formatted into structured prompts alongside relevant market context and forwarded to the model. Responses are structured around four elements: a contextual introduction, a current market overview, a balanced assessment of investment merits and risks, and a bottom-line recommendation. Outputs are parsed and rendered on the user dashboard in clean, readable format.

F. System Architecture

The platform is implemented around a Flask/FastAPI-based backend that acts as the central orchestrator. It coordinates three AI subsystems: the asset-specific price predictors (LSTM, Random Forest, Prophet), the FinBERT sentiment classifier, and the Nemotron advisory engine. User portfolio data is stored in a relational SQL database; historical price series are loaded from CSV files; and live market data is retrieved from external APIs. All outputs are rendered through a browser-based HTML interface built with Jinja2 templates.

The system architecture as shown in Figure 2 is focused on providing an integrated approach to combining multiple data sources and to utilizing an artificial intelligence driven analytical model to produce real-time, individualized and data-driven investment recommendations for all asset classes.

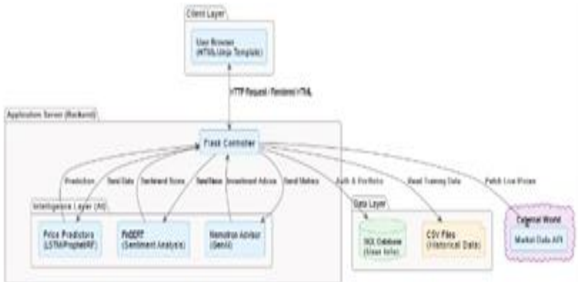

Figure 2. System Architecture

IV. RESULTS AND DISCUSSION

A. Equity Price Forecasting (LSTM)

The LSTM model for equity price prediction was trained on historical data for NSE-listed companies, using open, high, low, close, and volume as input features. A 90-day look-back window was adopted to capture medium-term temporal dependencies. Model performance was evaluated over a one-month test window from 22 September to 22 October 2025.

The model attained an RMSE of 0.1487, an MAE of 0.1181, and an R^2 of 0.7781. Although rapid intraday swings were not perfectly reproduced, broader directional trends were reliably captured. Against three baseline methods Buy-and-Hold, Naïve Forecasting, and a 20-day Simple Moving Average the LSTM was the only approach to correctly identify the downward price direction for the test stock, with a predicted closing price of ₹8.89 against an actual close of ₹9.16. The Naïve Forecast predicted ₹9.93 in the incorrect direction; the SMA-20 produced a lagged response near ₹9.50.

As depicted in Figure 3, the model performed reasonably well at forecasting future stock prices. While there were certainly minor deviations in stock price due to rapid changes in the market environment, the model did a good job of capturing overall trends. Overall, these findings suggest that LSTMs may provide viable options for short-term forecast of stock prices.

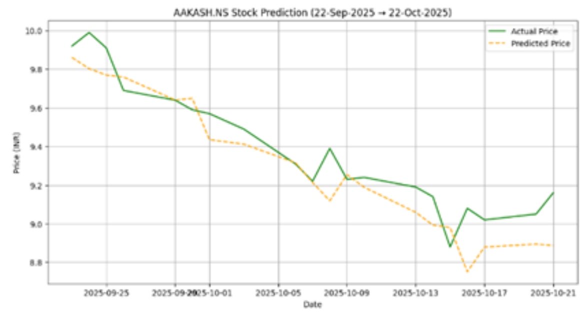

Fig 3. AAKASH.NS Stock Future Prediction Trend

B. Residential Real Estate Forecasting

For Mumbai residential real estate, the model ingested annual price-per-square-foot observations from Andheri West spanning 2015 to 2025. Polynomial regression with cross-validation and GridSearchCV hyperparameter tuning yielded an MAE of ₹1,330.74, an RMSE of ₹1,434.79, and an R^2 of 0.88, indicating that roughly 88 percent of observed price variance was explained. Minor divergences toward the end of the evaluation period likely reflect recent market volatility rather than systematic model deficiencies. The Prophet component additionally generated five-year forward projections at the locality level, offering investors a clear long-term appreciation trajectory.

As illustrated in Figure 4, the predicted trend line is similar to the actual trend line for most years (2008-2023) with regards to Andheri West. Deviations at the end of the predicted time frame (2009-2012) were very slight but may be attributed to recent fluctuation(s) in the real estate market or limited availability of current data. Thus overall, these findings support that this methodology can also be effectively applied to model the residential real estate pricing trend for specific urban sub-market areas

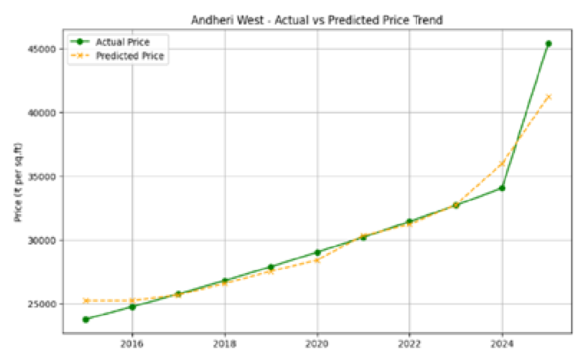

Figure 4. Andheri West - Actual vs. Predicted Price Trend

C. Gold Price Forecasting

Random Forest regression was applied to daily GLD closing price forecasting using temporal features and lagged price values. Training was conducted on pre-2016 data, with subsequent observations reserved for evaluation. The model demonstrated strong alignment between predicted and actual values on the held-out set and produced reliable 30-day projections a horizon well-matched to the planning horizons of individual retail investors.

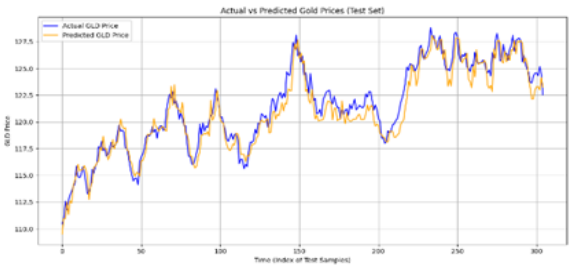

Figure 5. Actual Vs Predicted Gold Price (Test Set)

Figure 6. Gold Price Future Prediction Trend Analysis

The model demonstrated good performance when evaluating the predicted versus actual prices (Figure 5). The model was able to provide a 30-day forecast of future price trends (Figure 6), allowing for proactive decision making.

D. Summary of Predictive Performance

Across all five asset classes, the system achieved strong predictive accuracy: R^2 of 0.78 for equities (LSTM), 0.88 for residential real estate (Random Forest + Polynomial), 0.90 for gold (Random Forest), 0.85 for digital currencies (LSTM), and 0.99 for mutual funds (LSTM). These results demonstrate that asset-specific model selection rather than a single universal architecture yields consistently high performance across structurally diverse financial time series.

E. Mutual Fund NAV Prediction

The LSTM model was evaluated on the Canara Robeco Large Cap Fund (Direct Plan – Growth) dataset. Historical NAV series were normalized using min-max scaling, and a 60-day sliding window was employed as the look-back period. Training covered 80 percent of available data, with the remainder used for testing.

Performance was strong across all metrics: R^2 of 0.9879, MAE of 0.7398, and RMSE of 0.9575. Predicted NAV closely tracked actual values, particularly reproducing the long-term growth trajectory of the fund. Deviations during volatile sub-periods remained small, and the overall prediction curve preserved the structural shape of the actual data. The line charts are based on the date and net asset value (or price movement). The system as shown in Figure 7 uses an interactive plot to show date against net asset value (nav) and has a dark theme to allow better visibility and readability.

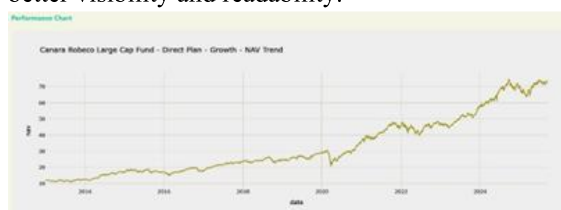

Figure 7. NAV Trend Visualization (Canara Robeco Large Cap Fund)

In Figure 8, we compare the actual NAV (blue line) to the predicted NAV (red line) for the test period. As you can see, the red line follows the blue line very closely in terms of the overall trend of the actual NAV; this 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). While there are small errors in the predictions during the most volatile periods, the LSTM network has done a good job of capturing 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 going forward.

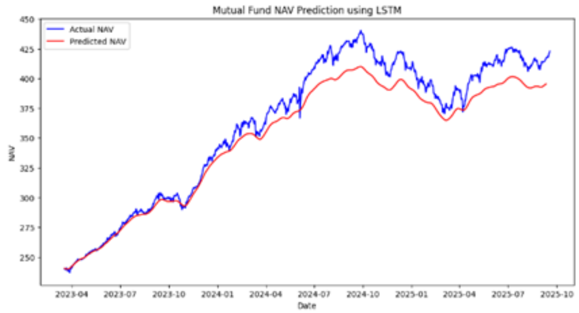

Figure 8. Mutual Fund NAV Prediction using LSTM (Canara Robeco Large Cap Fund)

F. Sentiment Analysis Findings

The sentiment module was applied to live financial news headlines from multiple Indian financial media sources. Each headline was scored via the lexicon-based approach and subsequently reclassified using FinBERT. The combined pipeline produced reliable sentiment labels that aligned closely with prevailing market conditions during the evaluation window.

The sentiment index proved particularly informative for short-term adjustments to equity and digital currency forecasts, where news-driven price movements can be abrupt. Embedding sentiment scores within advisory outputs also enabled the platform to surface interpretive context alongside quantitative recommendations, meaningfully enhancing transparency for non-expert users.

V. CONCLUSION

This paper has presented an Intelligent Multi-Asset Portfolio Advisory System that consolidates five major asset classes within a unified analytical and advisory framework. By deploying asset-specific machine learning and deep learning models, the system attains strong predictive performance across all domains, with R^2 values spanning 0.78 for equities to 0.99 for mutual funds.

The platform's distinguishing contribution lies not solely in its asset coverage, but in the coherent integration of quantitative forecasting, real-time sentiment classification, and large language model-generated advisory narratives. This architecture renders complex financial data accessible and actionable for retail investors who lack the technical background to interpret raw model outputs.

Several directions for future enhancement are identified. Incorporating macroeconomic indicators interest rates, inflation indices, and foreign exchange rates could strengthen prediction robustness. Extending sentiment analysis to regional language sources would expand relevance across India's linguistically diverse investor base. Integrating reinforcement learning for dynamic portfolio rebalancing, and conducting extended live-deployment validation, represent the most consequential avenues for future work.

This system constitutes a meaningful step toward democratizing data-driven, personalized investment guidance for individual investors in India, irrespective of their financial literacy.

ACKNOWLEDGEMENT

The authors wish to thank the Department of Computer Engineering, M. H. Saboo Siddik College of Engineering, Mumbai, for providing the computational resources and academic support that made this research possible. The authors also acknowledge Alpha Vantage, CoinGecko, the Association of Mutual Funds in India (AMFI), MagicBricks, and OpenRouter for providing access to the data sources and APIs utilised in this work. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] Tanwar, N. P. Patel, S. N. Patel, J. R. Patel, G. Sharma, and I. E. Davidson, "Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations," IEEE Access, vol. 9, pp. 143284–143295, Oct. 2021.

[2] F. Maloku, B. Maloku, A. A. D. Kumar, and A. Agarwal, "House Price Prediction Using Machine Learning and Artificial Intelligence," Journal of Artificial Intelligence & Cloud Computing, vol. 3, no. 4, pp. 1–10, Aug. 2024.

[3] N. Srikanth and M. Ramchander, "Machine Learning Approach for House Price Prediction," UGC Care Journal, Nov. 2023.

[4] K. Satyam and B. Suresh, "Predicting Real Estate Prices: A Machine Learning Approach," International Research Journal of Modernization in Engineering Technology and Science, Apr. 2025.

[5] S. K. Deepu, A. Verma, A. Raj et al., "Performance Evaluation of Real Estate Price Predictor Using Machine Learning," SSRN Electronic Journal, Jan. 2024.
[6] H. M. Zangana and S. R. Obeyd, "Deep Learning-based Gold Price Prediction: A Novel Approach using Time Series Analysis," Sistemasi: Jurnal Sistem Informasi, Oct. 2024.
[7] A. Amini and R. Kalantari, "Gold price prediction by a CNN-BiLSTM model along with automatic parameter tuning," PLOS ONE, Mar. 2024.
[8] B. Nagata, M. S. Hidajat et al., "Predicting Gold Price Movement Using Long Short-Term Memory Model," Journal of Applied Intelligent System, 2024.
[9] X. Fan and J. Chen, "Stock Price Forecasting in Real Estate Industry Based on Investor Sentiment," Frontiers in Business, Economics and Management, vol. 6, no. 3, pp. 54–59, 2022.
[10] K. Mahajan, "Transforming Financial Decision-Making with Artificial Intelligence," Journal of Electrical Systems, vol. 20, no. 10s, pp. 5036–5044, 2024.
[11] S. Arslan, "Bitcoin Price Prediction Using Sentiment Analysis and Empirical Mode Decomposition," Computational Economics, vol. 65, pp. 2227–2248, 2025.
[12] Deepthi C. G., J. J. Shetty, I. B. S. Jain et al., "AI-Powered Finance Management Platform," International Journal for Research in Applied Science & Engineering Technology, vol. 13, no. VI, pp. 1574–1578, Jun. 2025.
[13] M.-F. Leung, A. Jawaid, S.-W. Ip, C.-H. Kwok, and S. Yan, "A portfolio recommendation system based on machine learning and big data analytics," Data Science in Finance and Economics, vol. 3, no. 2, pp. 152–165, May 2023.
[14] P. Jay et al., "Stochastic Neural Networks for Cryptocurrency Price Prediction," IEEE Access, vol. 8, pp. 82804–82818, 2020.