NIVESH AI: A MULTI-AGENT, LLM-AUGMENTED INVESTMENT DECISION SUPPORT SYSTEM FOR INDIAN RETAIL STOCK MARKETS

[1]Divyansh Jindal, [2]Rishabh Sharma, [3]Naman Sharma, [4]Om Sharma
[1234]Department of Computer Science Engineering and Allied Branches
[1234]RD Engineering College, Ghaziabad, Uttar Pradesh, India
divyansh022ai0027@rdec.in, rishabh022ai0061@rdec.in, naman022ai0050@rdec.in, om022ai0056@rdec.in
Supervisor: Lav Kumar Dixit | lavkumardixit@rdec.in

Abstract: *Retail participation in Indian equity markets has surged post-pandemic, yet most individual investors lack access to institutional-grade analytical tools. This paper presents Nivesh AI, a full-stack, real-time, multi-agent investment decision support system built specifically for the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). The system orchestrates four specialised agents—a Technical Analysis Agent, a Fundamental Analysis Agent, a Price Prediction Agent, and a Live News Sentiment Agent—coordinated by a Master Orchestrator that fuses their verdicts into a single Nivesh AI Consensus signal. Price forecasting employs an ensemble of Linear Regression (weight 0.55) and ARIMA(5,1,2) (weight 0.45), further adjusted by a real-time sentiment score derived from both rule-based keyword analysis and Google Gemini 2.5 Flash LLM inference on live news headlines. The system attains a 30-day forecasting error as low as 2.84% MAPE on out-of-sample trials across a diverse set of NSE-listed equities. A prediction override tracker records every forecast outcome and marks it correct when directional accuracy is achieved within a 3% error envelope. The platform is deployed as a Streamlit web application backed by MongoDB Atlas and features WhatsApp OTP authentication, a Razorpay wallet, broker deep-links, and an AI Q&A chatbot. Nivesh AI demonstrates that multi-agent orchestration coupled with LLM-powered sentiment can meaningfully outperform single-model baselines for retail-oriented stock decision support.*

Index Terms — multi-agent system, stock market analysis, LLM sentiment, ARIMA, ensemble forecasting, Indian equity markets, retail investor, technical analysis, fundamental analysis, Gemini, LangChain

I. INTRODUCTION

The Indian equity market has witnessed an unprecedented democratisation over the past five years. According to the Securities and Exchange Board of India (SEBI), registered demat accounts crossed 160 million in 2024, with more than 70% belonging to first-generation retail investors operating without professional advisory support [1]. Despite this growth, retail participation remains hampered by (i) information asymmetry relative to institutional players, (ii) lack of affordable, integrated analytical platforms tailored to NSE/BSE instruments, and (iii) cognitive biases that arise when combining multiple, often contradictory, signals from disparate sources such as news, price charts, and financial statements.

Existing retail platforms—Zerodha Kite, Groww, and Upstox—provide raw data and basic charting but do not synthesise technical signals, fundamental metrics, predictive forecasts, and macroeconomic news into a unified verdict. Professional terminals such as Bloomberg or Refinitiv Eikon are prohibitively expensive. Academic literature on algorithmic trading and sentiment-driven forecasting is rich [2][3][4], yet translation of these advances into accessible, end-to-end systems for the Indian retail segment remains sparse.

This paper makes the following contributions:
- We design and implement a four-agent orchestration architecture that independently computes technical, fundamental, predictive, and sentiment verdicts, then fuses them through a weighted voting mechanism.
- We propose a sentiment-adjusted ensemble forecasting model that combines Linear Regression on engineered financial features with ARIMA(5,1,2), and modulates the 30-day price trajectory by a real-time LLM-derived sentiment score bounded to a ±1.5% drift envelope.
- We introduce a PredictionOverrideTracker that continuously records, evaluates, and reports per-ticker directional accuracy, enabling transparent model accountability.
- We demonstrate that the system is accessible to non-expert users through WhatsApp OTP login, vernacular-friendly UI, and rupee-denominated portfolio analytics.

II. RELATED WORK

A. Technical and Fundamental Analysis Systems

Automated systems applying rule-based technical analysis have been studied since the early 1990s. Lo et al. [5] demonstrated that technical indicators such as RSI, MACD, and moving averages carry statistically significant predictive content over medium horizons. More recent work by Chourmouziadis and Chatzoglou [6] combined multiple indicators into a fuzzy rule system, showing improved accuracy over single-indicator strategies. Fundamental scoring models—inspired by Piotroski's F-Score [7]—rank equities on profitability, leverage, and efficiency metrics; Nivesh AI adapts this paradigm to seven India-specific metrics including Debt-to-Equity norms prevalent in BSE-listed firms.

B. Machine Learning for Stock Price Forecasting

Supervised learning approaches for price prediction have explored linear regression [8], support vector machines [9], and deep learning architectures such as LSTM [10] and Transformers [11]. ARIMA and its variants remain strong baselines for univariate time-series tasks [12]. Hybrid ensemble methods consistently outperform single models: Zhang [13] showed that ARIMA–ANN combinations reduce forecasting error relative to either model alone. Our ensemble of Linear Regression and ARIMA(5,1,2) follows this tradition while adding a sentiment modulation layer.

C. News Sentiment and Social Media for Market Prediction

Bollen et al. [3] found that Twitter mood states predicted Dow Jones Industrial Average movements with 87.6% accuracy using a Granger causality framework. Ding et al. [2] demonstrated that structured event extraction from financial news improved short-term stock movement prediction beyond bag-of-words baselines. Schumaker and Chen [4] used SVMs on financial news to achieve statistically significant excess returns. More recently, large language models (LLMs) have been applied to financial text: BloombergGPT [14] and FinBERT [15] show superior sentiment classification on domain-specific corpora. Nivesh AI leverages Google Gemini 2.5 Flash as a zero-shot financial analyst, extracting structured JSON risk assessments from live multi-source news headlines.

D. Multi-Agent Systems in Finance

Multi-agent frameworks have been applied to portfolio optimisation [16] and algorithmic trading [17]. FinAgent [18] and similar LLM-orchestrated frameworks decompose financial reasoning into specialised sub-agents. Nivesh AI extends this paradigm by integrating heterogeneous analytical modalities—rule-based, ML, and LLM—under a single orchestrator for the retail Indian market context.

E. Gap Identified

Prior systems rarely integrate all four modalities (technical, fundamental, predictive, and sentiment) into a unified signal accessible to non-expert retail investors. Furthermore, LLM-based sentiment modulation of ensemble price predictions—with explicit drift bounding—has not been studied in the NSE/BSE context.

III. SYSTEM ARCHITECTURE

The Nivesh AI platform is structured as a layered, multi-agent system with five conceptual tiers:

Fig. 1: High-level multi-agent architecture of Nivesh AI. The Master Orchestrator collects independent verdicts from four specialised agents and fuses them into the Nivesh AI Consensus signal.

Tier 1 — Data Ingestion Layer. Raw market data (OHLCV, live price, market capitalisation, volume) is fetched via the yfinance Python library. NSE ticker symbols are automatically suffixed with .NS and BSE symbols with .BO. News data is aggregated from Yahoo Finance RSS feeds, Google News RSS, and structured buckets covering RBI announcements, Sensex/Nifty macro feeds, and global geopolitical news.

Tier 2 — Specialised Agent Layer. Four agents operate independently and in parallel: TechnicalAnalysisAgent, FundamentalAnalysisAgent, PricePredictionAgent, and LiveNewsAgent.

Tier 3 — Orchestration Layer. The MasterAgent collects verdicts (Buy / Hold / Sell) from each specialised agent, applies a majority voting mechanism, and resolves conflicts between technical and prediction signals. The output is the Nivesh AI Consensus with a plain-language conflict explanation when signals disagree.

Tier 4 — Persistence Layer. MongoDB Atlas hosts five collections: users, watchlist, price_alerts, portfolio, and wallet_transactions. Session persistence is implemented via streamlit_cookies_controller; a one-time disclaimer/terms gate is recorded per user.

Tier 5 — Presentation Layer. The frontend is a Streamlit web application featuring interactive Plotly candlestick charts, real-time price cards, a sector browser (20+ sectors), portfolio P&L dashboard, price alert management, investment calculator, AI Q&A chatbot (Gemini via LangChain), broker deep-links (Zerodha, Groww, Upstox, Angel One, 5Paisa), and Razorpay wallet integration for premium credits.

A. Authentication

Users authenticate via one of two OTP channels: WhatsApp OTP — delivered through the AiSensy API, enabling frictionless mobile-first login without requiring an email address; and Email OTP — dispatched via Hostinger SMTP as a fallback channel. Credentials and OTP hashes are stored in MongoDB with bcrypt hashing; sessions are maintained via secure HTTP-only cookies with a configurable TTL.

IV. METHODOLOGY

A. Technical Analysis Agent

The TechnicalAnalysisAgent retrieves 180 days of daily OHLCV data and computes eight indicators using the finta library: RSI (14-period), MACD (12/26/9 EMA), SMA-20, SMA-50, EMA-20, ADX (14-period), Momentum (10-period), CCI, and OBV. Rule-based verdicts are generated for two horizons:

Short-Term (1–2 weeks):

$$\begin{aligned} &Buy \Leftarrow RSI < 30 \ \ AND \ \ MACD > 0 \ \ AND \ \ Momentum > 0 \\ &Sell \Leftarrow RSI > 70 \ \ AND \ \ MACD < 0 \ \ AND \ \ Momentum < 0 \end{aligned}$$

Long-Term (3–12 months):

$$\begin{aligned} &Buy \Leftarrow Price > SMA50 \ \ AND \ \ ADX > 25 \ \ AND \ \ OBV\Delta > 0 \\ &Sell \Leftarrow Price < SMA50 \ \ AND \ \ ADX > 25 \ \ AND \ \ OBV\Delta < 0 \end{aligned}$$

Any state not matching a Buy or Sell pattern is labelled Hold. A volatility badge is additionally computed as annualised standard deviation of daily returns: Low (<20%), Medium (20%–40%), High (>40%).

B. Fundamental Analysis Agent

The FundamentalAnalysisAgent scores seven metrics against sector-calibrated thresholds (Table I). Each metric contributes +1 (favourable), −1 (unfavourable), or 0 (neutral) to a weighted aggregate score S_F. The verdict is determined by: $Verdict_F$ = Long-term Buy if $S_F \geq 4$; Long-term Sell if $S_F \leq -2$; Hold otherwise.

TABLE I. Fundamental Scoring Metrics and Thresholds

Metric	Buy Signal	Sell Signal
P/E Ratio	< 25	> 50
Profit Margin	> 10%	< 0%
ROE	> 15%	< 5%
Current Ratio	> 1.5	< 1.0
Debt-to-Equity	< 1.0	> 2.0
FCF Margin	> 5%	< 0%
Earnings Growth	> 10%	< -5%

C. Live News Sentiment Agent

1) News Aggregation

The LiveNewsAgent fetches headlines from two buckets: India-specific (stock-specific RSS, RBI policy updates, Sensex/Nifty macro feeds, FII/DII flows, rupee exchange rate, CPI inflation, GDP announcements) and Global macro (US Federal Reserve decisions, geopolitical conflicts, Brent/WTI crude prices, China trade data, IMF/World Bank outlooks). Deduplication is performed via URL fingerprinting and title similarity hashing before analysis.

2) Rule-Based Sentiment Scoring

A curated lexicon of domain-specific positive (e.g., record profit, FII inflow, rate cut) and negative (e.g., earnings miss, FII outflow, rate hike, crude surge) keywords is applied to each headline. The score is:

$$score\_rule = (N\_pos - N\_neg) / (N\_pos + N\_neg + \varepsilon) \ \in [-1.0, +1.0]$$

3) LLM-Powered Analysis

All deduplicated headlines are passed to Google Gemini 2.5 Flash via a structured prompt requesting a JSON response containing: overall_sentiment (bullish/bearish/neutral), key_risk_factors, sector_impact, price_impact_estimate, and buy_signal (boolean). The final sentiment score is derived as a weighted average: 60% LLM verdict + 40% rule-based score.

D. Price Prediction Agent

1) Feature Engineering

From the last 180 days of daily OHLCV data, 22 features are engineered for the Linear Regression (LR) model: daily return, log return, SMA-20, EMA-20, MACD, RSI, Bollinger Band percentage, volume change, high-low percentage, and lag features C(t-1) through C(t-5).

2) Linear Regression Model

An 80/20 chronological train–test split is used. The LR model is fitted on training data and iteratively unrolled for a 30-day forecast:

$$\hat{C}(t+k) = w^T x(t+k-1) + b, \quad k = 1, 2, ..., 30$$

3) ARIMA Model

An ARIMA(5,1,2) model is fitted on the last 180 days of closing prices. The order (p=5, d=1, q=2) was selected via minimisation of AIC on a rolling grid search across 20 NSE stocks.

4) Ensemble and Sentiment Adjustment

The ensemble forecast at horizon k is:

$$\hat{C}\_ens(t+k) = 0.55 \times \hat{C}\_LR(t+k) + 0.45 \times \hat{C}\_ARIMA(t+k)$$

The sentiment-adjusted final forecast:

$$\hat{C}\_final(t+k) = \hat{C}\_ens(t+k) \times (1 + \alpha \times score\_final)$$

where α = 0.015 caps the maximum sentiment-induced drift at ±1.5% per 30-day horizon to prevent runaway sentiment amplification.

5) PredictionOverrideTracker

Every forecast is persisted as a JSON record. After 30 days the actual market close is compared and the outcome is marked: Correct (directional accuracy ✓ AND error ≤ 3%), Partial (directional accuracy ✓ AND error > 3%), or Incorrect (directional accuracy ✗).

E. Master Orchestrator

The MasterAgent implements a simple yet effective voting scheme: (i) collect {Technical Buy/Hold/Sell}, {Fundamental Buy/Hold/Sell}, and {Prediction Buy/Hold/Sell}; (ii) count Buy and Sell votes — majority wins (threshold: 2 of 3); (iii) if Technical and Prediction signals conflict, output Hold with a plain-language explanation; (iv) append the LLM sentiment JSON as a contextual advisory layer rather than a direct voting agent, to avoid over-weighting noisy headline sentiment.

V. RESULTS AND EVALUATION

A. Experimental Setup

We evaluated the forecasting pipeline on a retrospective dataset of 30 NSE-listed equities spanning three sectors (Banking, IT, Pharma), using closing prices from January 2023 to December 2024. For each stock, a 30-day out-of-sample forecast was generated from each month-end as the origin, yielding approximately 720 individual forecast trials.

B. Forecasting Accuracy

Table II summarises the mean MAPE and directional accuracy across model variants. The sentiment-adjusted ensemble consistently outperforms standalone LR and ARIMA baselines.

TABLE II. Forecasting Performance — 30-Day Horizon (NSE, 2023–2024)

Model	MAPE (%)	MAE (₹)	Dir. Acc. (%)
ARIMA(5,1,2) only	4.12	62.4	61.2
Linear Regression only	3.87	57.9	63.5
Ensemble (LR + ARIMA)	3.21	49.1	67.8
Ensemble + Rule Sentiment	2.96	45.3	70.1
Ensemble + LLM Sentiment (Proposed)	2.84	43.7	72.4

The proposed sentiment-adjusted LLM ensemble records a 2.84% MAPE, marking a 31% reduction in error over standalone ARIMA and a 12% reduction over the raw ensemble. Directional accuracy improves from 61.2% (ARIMA alone) to 72.4% with LLM sentiment augmentation.

C. PredictionOverrideTracker Outcomes

Across 720 trials, the tracker classified outcomes as: Correct (direction right, error ≤3%): 52.1%; Partial (direction right, error >3%): 20.3%; Incorrect (direction wrong): 27.6%. Thus, 72.4% of forecasts achieved correct directional prediction. The average error across all trials was ₹43.7, on an average stock price of approximately ₹1,540, corresponding to a 2.84% MAPE.

D. Consensus Signal Quality

The MasterAgent's consensus signal was evaluated using a simulated paper trading strategy. Over the 24-month evaluation window, the consensus strategy produced a simulated annualised return of 18.3% versus a Nifty 50 index return of 14.7% over the same period, representing a 3.6 percentage point alpha. The Hold regime (triggered on signal conflict) correctly avoided 68% of subsequent drawdown events.

E. Sentiment Analysis Quality

The LLM-based Gemini 2.5 Flash sentiment extractor was benchmarked against manually labelled sentiment annotations on 500 NSE-related news items (two human annotators, Cohen's $\kappa = 0.81$). Gemini 2.5 Flash achieved 84.2% agreement with the human consensus label, outperforming the rule-based keyword scorer (71.5%) by 12.7 percentage points.

F. System Performance

End-to-end analysis latency (from ticker submission to full Consensus signal display) was measured at a median of 6.2 seconds on a standard cloud VM (2 vCPU, 4 GB RAM), well within the 10-second threshold deemed acceptable for retail users in usability studies [19].

VI. CONCLUSION

This paper presented Nivesh AI, a multi-agent investment decision support system that integrates technical analysis, fundamental scoring, ML-based price forecasting, and LLM-driven news sentiment into a unified Nivesh AI Consensus signal for Indian retail investors. The key contribution is a sentiment-adjusted ensemble model—combining Linear Regression and ARIMA(5,1,2) modulated by Google Gemini 2.5 Flash sentiment bounded to ±1.5% drift—that attains a 2.84% MAPE with a directional accuracy of 72.4% on 30-day NSE equity forecasts. The MasterAgent's conflict-resolution voting mechanism further demonstrated a 3.6 pp alpha over the Nifty 50 index in a paper trading simulation.

Beyond forecasting, Nivesh AI addresses accessibility barriers facing Indian retail investors through WhatsApp OTP authentication, rupee-denominated analytics, broker deep-links, and a conversational AI Q&A interface. The PredictionOverrideTracker introduces a transparent accountability mechanism rare in consumer financial applications.

Future Work. We plan to: (i) replace Linear Regression with an LSTM or Temporal Fusion Transformer to capture non-linear patterns; (ii) extend sentiment aggregation to Hindi-language news sources and social media (Twitter/X, StockTwits India) using multilingual LLMs; (iii) implement reinforcement learning-based portfolio rebalancing; and (iv) conduct a controlled user study comparing decision quality of retail investors using Nivesh AI versus unaided analysis.

ACKNOWLEDGMENT

The authors sincerely thank the faculty of the Department of Computer Science Engineering and Allied Branches, RD Engineering College, Ghaziabad, for their continuous guidance and support. We also acknowledge the open-source communities behind yfinance, finta, LangChain, and Streamlit, without which this project would not have been feasible.

REFERENCES

[1] Securities and Exchange Board of India (SEBI), "Annual Report 2023–24," SEBI Publications, Mumbai, India, 2024.

[2] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep Learning for Event-Driven Stock Prediction," in Proc. 24th Int. Jt. Conf. Artif. Intell. (IJCAI), 2015, pp. 2327–2333.

[3] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," J. Comput. Sci., vol. 2, no. 1, pp. 1–8, 2011.

[4] R. P. Schumaker and H. Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System," ACM Trans. Inf. Syst., vol. 27, no. 2, pp. 1–19, 2009.

[5] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of Technical Analysis," J. Finance, vol. 55, no. 4, pp. 1705–1765, 2000.

[6] K. Chourmouziadis and P. D. Chatzoglou, "An Intelligent Short Term Stock Trading Fuzzy System," Expert Syst. Appl., vol. 43, pp. 298–311, 2016.

[7] J. D. Piotroski, "Value Investing: The Use of Historical Financial Statement Information," J. Account. Res., vol. 38, pp. 1–41, 2000.

[8] B. Qian and K. Rasheed, "Stock Market Prediction with Multiple Classifiers," Appl. Intell., vol. 26, no. 1, pp. 25–33, 2007.

[9] K.-J. Kim, "Financial Time Series Forecasting Using Support Vector Machines," Neurocomputing, vol. 55, pp. 307–319, 2003.

[10] T. Fischer and C. Krauss, "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, 2018.

[11] H. Zhou et al., "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," in Proc. AAAI Conf. Artif. Intell., vol. 35, 2021, pp. 11106–11115.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Wiley, 2015.

[13] G. P. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," Neurocomputing, vol. 50, pp. 159–175, 2003.

[14] S. Wu et al., "BloombergGPT: A Large Language Model for Finance," arXiv preprint arXiv:2303.17564, 2023.

[15] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," arXiv preprint arXiv:1908.10063, 2019.

[16] K. Sycara et al., "Distributed Intelligent Agents," IEEE Expert, vol. 11, no. 6, pp. 36–46, 1996.

[17] A. Brabazon and M. O'Neill, Biologically Inspired Algorithms for Financial Modelling. Springer, 2006.

[18] W. Liu et al., "FinAgent: A Multimodal Foundation Agent for Financial Trading," arXiv preprint arXiv:2402.18485, 2024.

[19] J. Nielsen, Usability Engineering. Morgan Kaufmann, 1994.