SSRG International Journal of Economics and Management Studies
ISSN: 2393–9125 / https://doi.org/10.14445/23939125/IJEMS-V13I6P102

Volume 13 Issue 6, 7-20, June 2026

Original Article

Enhancing Stock Price Prediction through Artificial Intelligence-Driven Integration of Market Data and Sentiment Analysis

Sridevi Tandley Omprakash

Department of Analytics and Economics, Bharathidasan Institute of Management, Tiruchirappalli, Tamil Nadu, India.

[1]*Corresponding Author : sridevi.tandley@bim.edu*

Received: 19 April 2026 Revised: 22 May 2026 Accepted: 11 June 2026 Published: 29 June 2026

Abstract *- Forecasting equity prices is hard, and the literature has been telling itself this for fifty years. Markets are volatile, nonlinear, and shaped by investor psychology in ways that Autoregressive Integrated Moving Average (ARIMA) and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) were never designed to capture. The deep-learning literature has been pushing back on this for the last decade, but the typical study tests one or two networks against a single baseline, draws sentiment from a single platform, and reports on a handful of stocks within one exchange. A head-to-head benchmark that holds preprocessing, validation, and sentiment integration constant across many models and many markets is conspicuously absent. This paper supplies one. Ten forecasters were trained and tested on five years of daily prices for 25 technology equities listed across NYSE, NASDAQ, the Mexican Stock Exchange, Shanghai, and Korea: ARIMA, Exponential Smoothing State Space (ETS), GARCH, Random Forest Regression (RFR), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Prophet, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and an ARIMA-LSTM hybrid. A parallel sentiment pipeline was built around 1,849 news articles drawn from NewsAPI.org, with two vocabulary sizes (5,000 and 3,000 tokens) feeding both LSTM and Random-Forest classifiers. Dropout (rate 0.2), early stopping (patience 10), and walk-forward validation were applied to the recurrent models so that accuracy gains could be attributed to the architecture rather than to a forgiving data split. LSTM achieved the lowest mean Root Mean Square Error (RMSE) at 77.29 and Mean Absolute Percentage Error (MAPE) of 2.58%; GRU was a close second at RMSE 66.03 and MAPE 2.49%. ARIMA produced a mean MAPE of 53.01% and Prophet 33.87%. Diebold–Mariano testing confirms the LSTM advantage over both at the 1% level. Negative-sentiment recall, however, remained weak — a finding with direct consequences for any production system that needs news polarity to anticipate downside risk.*

Keywords *- Behavioural finance, Deep learning, Financial news integration, LSTM, Sentiment analysis, Stock price forecasting, Transformer models.*

1. Introduction

Equity price forecasting sits at the centre of portfolio construction, asset allocation, and risk management. It is also notoriously difficult, and the difficulty is not for want of trying. Prices move in response to fundamentals, but also to crowd psychology, news shocks, and microstructural quirks that no single model fully accommodates. The classical toolkit — ARIMA, GARCH, exponential smoothing, and their hybrids — was built on assumptions of stationarity and rational expectations that empirical work has steadily eroded since at least Shiller's [26] excess-volatility tests and Fama's [8] later qualifications of his own efficient-markets framework.

Progress in artificial intelligence has reopened the question. Recurrent and attention-based networks now routinely outperform classical baselines on tabular and sequential financial data: Sirignano and Cont [28] reported a 27% reduction in limit-order-book errors using deep recurrent models, and Shapiro, Sudhof and Wilson [25] showed that a properly constructed news-sentiment index carries return-predictive content that traditional macro factors miss. Sentiment-augmented pipelines have added several percentage points of accuracy in single-equity studies [2], [9], [15]. Yet three weaknesses run through this literature, and they are the reason the present paper exists.

1.1. Research Gap

Three weaknesses run through the existing literature. The first is narrow comparison: most papers pit a single deep-learning model — usually an LSTM — against ARIMA or GARCH and stop there, leaving competitors such as

XGBoost, Prophet, ETS, and SVM untested under identical conditions, so practitioners cannot tell whether the reported win is architectural or down to preprocessing. The second is corpus narrowness on the sentiment side: text is typically drawn from one platform, labelled for one or two stocks in a single market, and a recall figure is reported, with no evidence that the approach generalises across exchanges, languages, or regulatory regimes. The third is methodological opacity: overfitting controls, validation protocols, and tests of statistical significance are often described in a sentence or omitted altogether, and when walk-forward validation is replaced by a single random split, the headline accuracy can flatter the model on its own training distribution while reproducibility suffers.

1.2. Novelty and Contributions

Five contributions follow from these gaps. The benchmark spans ten forecasters — ARIMA, ETS, GARCH, RFR, XGBoost, SVM, Prophet, LSTM, GRU, and an ARIMA-LSTM hybrid — trained and tested under matched preprocessing, hyperparameter ranges, and validation, so that any difference in accuracy is attributable to the model rather than to an experimenter's thumb on the scale. The portfolio covers 25 technology equities across NYSE, NASDAQ, the Mexican Stock Exchange (Renesas), Shanghai (Rockchip), and Korea (Samsung), pushing external validity beyond the single-market norm. The sentiment pipeline runs two vocabulary configurations (5,000 and 3,000 tokens) through both LSTM and Random-Forest classifiers, isolating the effect of model class from representational granularity. Dropout 0.2, early stopping with patience 10, and walk-forward validation were applied uniformly to the recurrent models, with per-stock error decomposition surfacing outlier-driven failures — most visibly Renesas Electronics, the only Mexican-peso quote in the panel, which alone dominates the aggregate RMSE figure. Finally, the achieved LSTM MAPE of 2.58% is positioned explicitly against Mukherjee et al. [23], Madeeh and Abdullah [20], and Halder [11].

Two research questions organise the empirical work. RQ1: how do AI/ML models compare with classical statistical baselines for stock-price prediction across heterogeneous global exchanges? RQ2: What is the marginal contribution of news-sentiment analysis to predictive performance, and which classifier-vectoriser configurations work best?

The remainder of the paper is structured as follows. Section 2 reviews the relevant theoretical and empirical literature. Section 3 covers data, preprocessing, model specifications, regularisation, and the validation protocol. Section 4 reports per-stock RMSE and MAPE for all ten models, the sentiment classification report, and the Diebold–Mariano significance tests. Section 5 discusses interpretability, ethical and regulatory considerations, model-crowding feedback risks, and the study's limitations. Section 6 concludes.

2. Theoretical Framework and Literature Synthesis

2.1. Where Classical Models Fall Short

The case against classical forecasting starts with Fama [8]. The Efficient Market Hypothesis assumes that prices already incorporate all available information, leaving little room for systematic prediction. Shiller [26] tested that assumption empirically and found prices moving far more than fundamentals could justify, an excess-volatility result that has held up across markets and decades.

Lo [17] later softened the framework with the Adaptive Markets Hypothesis, where agents are boundedly rational, and strategies survive or die through market selection rather than because they were optimal at the start. Mandelbrot [21] had already shown returns are not Gaussian; Cont [5] catalogued the stylised facts — fat tails, volatility clustering, slow decay of return autocorrelation — that ARIMA and GARCH ignore by construction.

Empirical work in emerging and developed markets has reinforced these concerns. Zhao et al. [33] surveyed the data-driven stock-forecasting literature and concluded that single-method approaches struggle in the presence of noise, non-stationarity, and shock events such as COVID-19, particularly when only a narrow set of indicators is fed in. Bao, Yue and Rao [34] showed that a stacked-autoencoder pipeline feeding into LSTM produces measurably better RMSE and accuracy across the CSI 300, Hang Seng and S&P 500 than ARIMA on the same data. Black-Swan vulnerabilities are also well documented: Taleb [29] argued that classical risk models systematically underweight tail events, and Cookson et al. [6] showed similar behaviour around the 2020–21 meme-stock episode, where social-media-driven volatility rendered standard predictions useless.

Fundamental analysis fares no better as a standalone forecasting tool. Early work by Ou and Penman [24] and Lev and Thiagarajan [16] demonstrated that ratios extracted from financial statements can identify mispriced assets, but the linear-regression machinery struggles with the nonlinear interactions that dominate at higher frequencies. More recent work has raised reliability concerns of a different kind — corporate impression management in disclosures — which adds a measurement-error problem on top of the modelling one.

2.2. AI/ML Approaches to Nonlinear Prediction

The theoretical justification for switching to neural networks rests on the universal approximation theorem of Hornik et al. [13]: a feedforward network with one hidden layer can approximate any continuous function on a compact domain. This does not guarantee good forecasts, but it removes the linearity constraint that handicaps ARIMA and GARCH. Practical validation has come from several

directions. Mukherjee, Sadhukhan and colleagues [23] reported 98.92% accuracy from a CNN applied to two-dimensional histograms of stock prices — a classification task rather than continuous regression, but a useful illustration of how far a deep architecture can go when the input representation is engineered carefully.

Madeeh and Abdullah [20] reached strong precision and recall on equity-direction classification using Random Forests over technical indicators. Bao, Yue and Rao [34] used stacked autoencoders to extract latent features before an LSTM step, which both reduced training-set noise and produced lower MAPE than ARIMA on the S&P 500.

Architectural progress has been steady. Hinton and Salakhutdinov [12] established autoencoders as a route to dimensionality reduction; Vaswani et al. [30] introduced the attention mechanism that now underpins most state-of-the-art sequence models; Halder [11] combined FinBERT with LSTM for sentiment-driven price prediction on the NASDAQ-100. On microstructure data, Sirignano and Cont [28] reported a 27% reduction in limit-order-book prediction error using deep recurrent networks.

On daily data, Shapiro, Sudhof, and Wilson [25] constructed a news-sentiment index that materially improves macro-financial forecasts when added to standard return predictors, demonstrating that text features carry information not already captured in prices. The remaining concerns are practical rather than theoretical: overfitting risk [23], computational cost [4], and the difficulty of explaining a black-box prediction to an auditor or a regulator.

2.3. Sentiment as a Behavioural-Finance Bridge

The behavioural-finance case for sentiment goes back to Kahneman and Tversky [14], whose prospect theory showed that investors weight losses about twice as heavily as equivalent gains. Shiller [27] extended the framework with narrative economics, arguing that viral stories rather than fundamentals often drive speculative bubbles. Operationalising these ideas required progress in text processing. The Loughran-McDonald financial lexicon [18] gave researchers a domain-specific dictionary that outperformed general-purpose word lists on 10-K filings; Araci [3] then released FinBERT, a transformer pretrained on financial text that handles context and negation in ways a lexicon cannot.

Empirically, sentiment has been shown to add forecasting power across many markets. Loutfi [19] found that environmental TV news (CNN, Fox) measurably moves renewable-energy share prices, with conservative outlets exerting larger short-term influence. Wang and colleagues [32] showed that investor attention and sentiment built from posts on the Guba forum carry significant predictive content for sector-level Chinese stock returns, with sentiment outperforming attention alone. Al Ridhawi [2] combined Twitter sentiment with financial price data via an MLP–LSTM–CNN ensemble and reported 74.3% next-hour prediction performance across AAPL, CSCO, IBM and MSFT, illustrating both the upside of multi-source pipelines and the modest size of the gains in absolute terms.

Khan, Ghazanfar and colleagues [15] reached maximum accuracies of 80.53% on social-media-only models and 75.16% on news-only models for ten-day predictions, with a Random Forest ensemble reaching 83.22% after spam filtering. Zaman, Ghazanfar and colleagues [9] subsequently reported 87.2% accuracy when combining sentiment with macroeconomic indicators (oil and gold prices) on HPQ, IBM, ORCL and MSFT. The persistent obstacles are noise [15], source bias [19], and regional variation in regulatory regimes [31].

2.4. Multi-Source Data Integration

Akerlof's [1] lemons problem framed information asymmetry as the central friction in markets where quality is hidden. In financial forecasting, the equivalent friction is unobserved corporate news, sentiment, and macro context. Modern pipelines reduce that friction by pulling text directly from filings and headlines. Elend and colleagues [7] trained LSTMs on quarterly financial reporting data to predict earnings per share and outperformed analyst consensus by up to 12.2 percentage points. On the sentiment side, Loughran and McDonald [18] developed the finance-domain lexicon that made 10-K sentiment analysis practical at scale, and their dictionary remains a widely used baseline for adding text-based signals to price-based prediction. The pattern is consistent with Milgrom and Roberts [22] on the complementarity principle: heterogeneous data sources reinforce one another more than they substitute.

Methodologically, the field has moved from lexicons [18] to transformers [25] to hybrids [11] in roughly a decade. The remaining problems are temporal latency between a news event and the model's ingestion of it, data reliability when filings are managed for impression rather than disclosure, and feature-selection complexity when text and price data are combined naively [15]. The four research streams reviewed above set up the empirical questions that Section 3 then operationalises.

3. Materials and Methods

Twenty-five technology firms were selected for the empirical work, drawn from five exchanges — NASDAQ, NYSE, the Mexican Stock Exchange, the Shanghai Stock Exchange, and the Korea Exchange. Semiconductor and IT-services firms dominate the sample because their price series are unusually volatile: rapid product cycles, geopolitical tariffs, and intense vendor rivalry leave clear footprints in the daily data, which is exactly what a forecasting benchmark needs.

Table 1. Assets used in the study

Stock Name*	Exchange	Ticker	Currency	Industry
Advanced Micro Devices, Inc.	Nasdaq Stock Market	AMD	USD	Semiconductors & Semiconductor Equipment
Alibaba Group Holding Limited	New York Stock Exchange	BABA	USD	Software & IT Services
Alphabet Inc.	Nasdaq Stock Market	GOOG	USD	Software & IT Services
Amazon.com, Inc.	Nasdaq Stock Market	AMZN	USD	Diversified Retail
Ambarella Inc.	Nasdaq Stock Market	AMBA	USD	Semiconductors & Semiconductor Equipment
Apple Inc.	Nasdaq Stock Market	AAPL	USD	Computers, Phones & Household Electronics
Applied Optoelectronics, Inc.	Nasdaq Stock Market	AAOI	USD	Electronic Equipment & Parts
Arista Networks, Inc.	New York Stock Exchange	ANET	USD	Communications & Networking
Baidu, Inc.	Nasdaq Stock Market	BIDU	USD	Software & IT Services
Broadcom Inc.	Nasdaq Stock Market	AVGO	USD	Semiconductors & Semiconductor Equipment
Cisco Systems, Inc.	Nasdaq Stock Market	CSCO	USD	Communications & Networking
Hewlett Packard Enterprise Co.	New York Stock Exchange	HPE	USD	Computers, Phones & Household Electronics
HP Inc.	New York Stock Exchange	HPQ	USD	Computers, Phones & Household Electronics
Intel Corporation	Nasdaq Stock Market	INTC	USD	Semiconductors & Semiconductor Equipment
Juniper Networks, Inc.	New York Stock Exchange	JNPR	USD	Communications & Networking
Kingsoft Cloud Holdings Limited	Nasdaq Stock Market	KC	USD	Software & IT Services
Lumentum Holdings Inc.	Nasdaq Stock Market	LITE	USD	Communications & Networking
Marvell Technology, Inc.	Nasdaq Stock Market	MRVL	USD	Semiconductors & Semiconductor Equipment
Microsoft Corporation	Nasdaq Stock Market	MSFT	USD	Software & IT Services
Nvidia Corporation	Nasdaq Stock Market	NVDA	USD	Semiconductors & Semiconductor Equipment
Qualcomm Incorporated	Nasdaq Stock Market	QCOM	USD	Semiconductors & Semiconductor Equipment
Renesas Electronics Corp.	Mexican Stock Exchange	RNECN	MXN	Semiconductors & Semiconductor Equipment
Rockchip Electronics Co., Ltd.	Shanghai Stock Exchange	603893.SS	CNY	Semiconductors & Semiconductor Equipment
Samsung Electronics Co., Ltd.	Korea Exchange	005930.KS	KRW	Computers, Phones & Household Electronics
Tesla, Inc.	Nasdaq Stock Market	TSLA	USD	Automobiles & Auto Parts

*Note: * Data was extracted on 31 January 2025.*

3.1. Data Collection

Price data was pulled programmatically with the Python *yfinance* library, which wraps the Yahoo Finance endpoints. For each ticker the request returned six fields per trading day: date, open, close, high, low, and volume. The window was 1,825 calendar days, ending 31 January 2025. Foreign listings reporting in non-USD currencies (Mexican peso for Renesas, Chinese yuan for Rockchip, Korean won for Samsung) were kept in local currency; FX-converting them would have introduced noise that does not belong in a price-prediction benchmark.

News data was sourced from *NewsAPI.org*. The free tier was sufficient because the sentiment study required only a 30-day window, ending 31 January 2025. After deduplication the corpus comprised 1,849 articles, returned as JSON with date,

source, author, title, description, and full content. Non-English articles were translated using MarianMT with the Helsinki-NLP/opus-mt-ja-en checkpoint. Topic labels were assigned by the BART zero-shot classifier facebook/bart-large-mnli across ten domains: finance, technology, sports, health, politics, entertainment, science, travel, food, and education. Polarity — positive (1), negative (0), neutral (2) — was assigned manually by the author for a clean training signal.

3.2. Pipeline Overview

The end-to-end pipeline runs in three stages: ingestion and cleaning, model training, and benchmarking. Figure 1 summarises the flow.

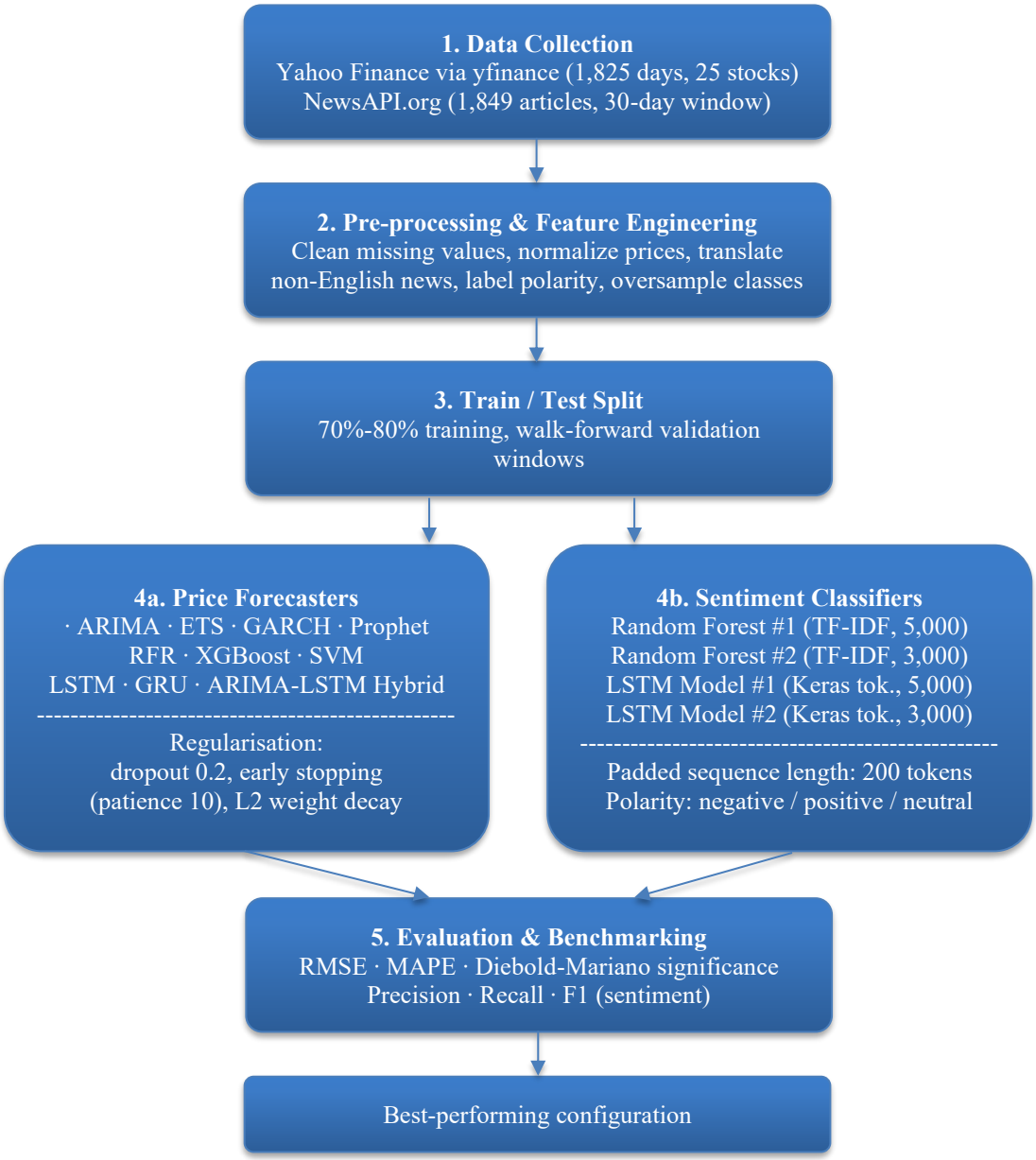

Fig. 1 High-level model development framework

Cleaning removes incomplete rows, normalises numerical features, and engineers required lag variables. The clean series is partitioned into training and test windows for each forecaster. Models are initialised with a grid-searched configuration, fit on the training window, and scored on the held-out window. Per-stock and pooled scores are then collected and reported.

3.3. Data Preprocessing

3.3.1. Price Series for the AI/ML Forecasters

For the price-prediction models the input was reduced to two columns — date and closing price — with all other fields discarded. The 70/30 train/test split follows Gholamy et al. [10], who reviewed the trade-off between training-set size and test-set reliability and found 70–80% training to be optimal in

most settings. Each of the ten forecasters was trained twice: once on each stock individually, producing 25 fitted models per architecture, and once on the pooled multi-stock series. The first regime tests within-stock predictive power; the second tests whether shared cross-sectional structure improves accuracy.

Table 2. Rationale for model selection framework for financial time-series analysis

Model	Key Characteristics	Selection Rationale	Key Parameters Considered
ARIMA	Captures linear trends, seasonality and noise via AR, differencing (I) and MA components.	Baseline for linear patterns in price/volatility data; establishes performance benchmark.	(p, d, q) orders; seasonal components (P, D, Q).
ARIMA-LSTM Hybrid	Combines ARIMA (linear) with LSTM (nonlinear) capabilities.	Addresses both linear and complex nonlinear dependencies in market dynamics.	ARIMA (p, d, q); LSTM layers/units; fusion weights.
ETS	Explicitly models Error, Trend and Seasonality components.	Effective for data with pronounced seasonal cycles such as quarterly earnings effects.	Error type (additive/multiplicative); trend damping; seasonal periods.
GARCH	Models time-varying volatility and volatility clustering.	Critical for capturing heteroskedasticity in financial returns and risk quantification.	Lag orders (p, q); distribution assumption (Gaussian, Student-t).
GRU	Gated RNN variant with simpler architecture than LSTM.	Efficient sequential learning with reduced computational overhead for high-frequency data.	Hidden units; layers; dropout rate; sequence length.
LSTM	Handles long-term dependencies via gated cell states.	Models complex temporal patterns in price momentum and sentiment persistence.	Hidden units; layers; dropout rate; bidirectional settings.
Prophet	Automatically detects trends, seasonality and holiday effects.	Robust to missing data; ideal for event-driven market anomalies such as earnings announcements.	Changepoint prior scale; seasonality mode; holiday effects.
RFR	Ensemble of decision trees; reduces overfitting.	Handles high-dimensional technical/fundamental data; robust to outliers and nonlinear relationships.	n_estimators; max_depth; min_samples_split; feature sampling ratio.
SVM	Finds optimal separation hyperplane using kernel methods.	Effective in high-dimensional spaces; suitable for regime detection in market states.	Kernel type (linear/RBF); C (regularisation); γ (kernel coefficient).
XGBoost	Gradient-boosted trees with regularisation.	State of the art for structured data; superior handling of feature interactions in multi-source inputs.	learning_rate; max_depth; subsample; early stopping rounds.

Table 2 sets out the rationale for the ten chosen forecasters — their key characteristics, why each was included, and the parameters that were tuned. The mix is deliberate: linear baselines (ARIMA, ETS), volatility models (GARCH), tree-based learners (RFR, XGBoost), kernel methods (SVM), additive-trend models (Prophet), recurrent networks (LSTM, GRU), and one hybrid (ARIMA-LSTM). Anything narrower would not have answered RQ1 cleanly.

3.3.2. Sentiment Pipeline

Sentiment work used the 1,849-article corpus described in Section 3.1. Article text was vectorised in two ways. For Random-Forest classifiers, TF-IDF was applied with vocabulary thresholds of 5,000 and 3,000 features. For LSTM classifiers, the Keras tokenizer was used with the same two cut-offs, padded to a fixed sequence length of 200 tokens. The dual-vocabulary design isolates the effect of vocabulary size from the effect of model class.

Two diagnostics were generated before training. A heatmap of articles against polarity revealed coverage gaps for several smaller-cap names. Word clouds were also produced per topic; these are summarised in Figure 2.

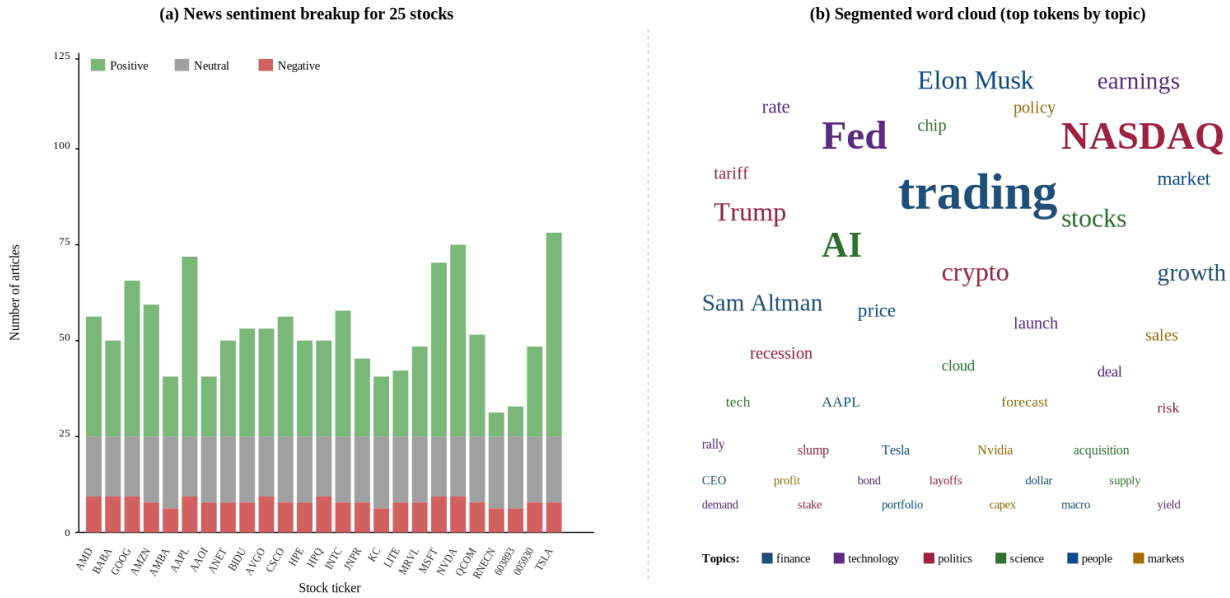

Fig. 2 Exploratory NLP analysis: (a) news sentiment breakup for the 25 stocks; (b) segmented word cloud

Class imbalance — a known problem with manually labelled financial sentiment — was addressed using the resample utility in scikit-learn, which oversampled the minority polarity classes to match the majority. Two metadata features (author, description) were dropped because they leaked publication-source information unrelated to article content. Figure 3 shows the per-class distribution before and after rebalancing.

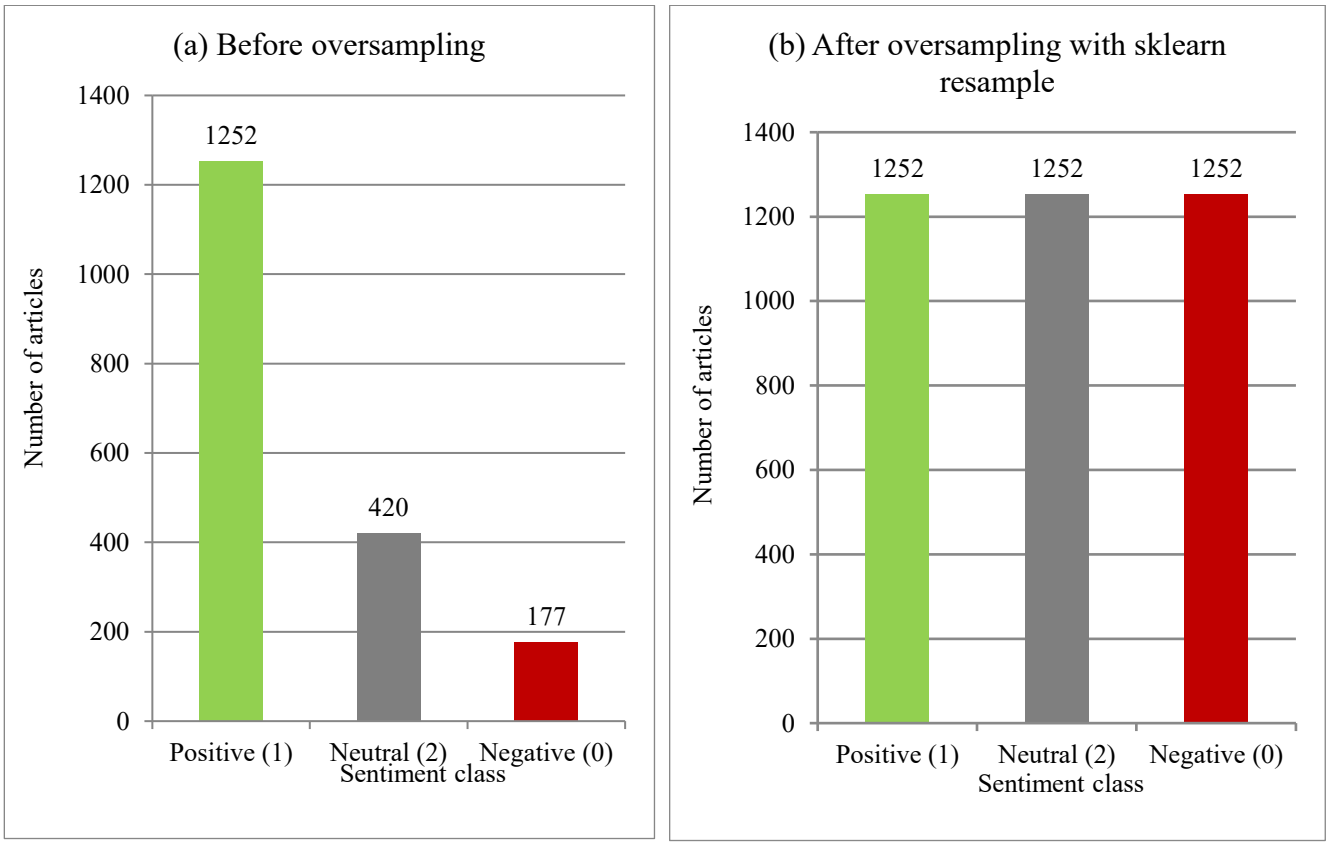

Fig. 3 Article-class distribution before and after sklearn oversampling.

Table 3. Sentiment classification model configuration.

Model Configuration	Vectoriser / Tokeniser	Vocabulary Size	Padded Sequence Length
Random Forest Classifier #1	TF-IDF	5,000	—
LSTM Model #1	Tokenizer (Keras)	5,000	200
Random Forest Classifier #2	TF-IDF	3,000	—
LSTM Model #2	Tokenizer (Keras)	3,000	200

Two TF-IDF feature sizes (5,000 and 3,000) were paired with Random Forest, and two Keras-tokenised vocabularies of the same sizes were paired with LSTM. The 200-token zero-padding sequence was kept constant across LSTM runs, so the only changing variable was vocabulary size. This factorial design isolates two effects: model class (tree ensemble vs recurrent network) and representational capacity (5,000 vs 3,000 tokens).

3.4. Evaluation Metrics

Two error metrics are used throughout. RMSE measures absolute predictive accuracy in the original price units; smaller values mean tighter forecasts. MAPE measures relative accuracy as a percentage and is more meaningful for cross-stock comparison because it normalises by price level. They are computed as:

$$\text{RMSE} = \text{sqrt}\left((1/n) \times \Sigma \left(y_t - \hat{y}_t \right)^2 \right) \quad (1)$$

$$\text{MAPE} = (100\% / n) \times \Sigma \left| \left(y_t - \hat{y}_t \right) / y_t \right| \quad (2)$$

where y_t is the realised closing price on day t, $\hat{y}_t$ is the model's prediction, and n is the number of test-window observations. The LSTM cell update used in this study follows the standard form:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f),\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i),\ \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C),\ C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t,\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o),\ h_t = o_t \odot \tanh(C_t). \quad (3)$$

Here σ is the sigmoid activation, ⊙ the Hadamard product, x_t the input at time t, h_t the hidden state, and C_t the cell memory. The forget, input, and output gates (f_t, i_t, o_t) control how much past information is retained, how much new information is written, and how much of the cell state is exposed to the next layer.

3.5. Overfitting Controls and Validation Protocol

Recurrent networks fit financial data well partly because they have enough parameters to memorise it, and without explicit regularisation the headline accuracy figures here would be optimistic. Four controls were applied uniformly. Dropout at 0.2 was inserted between LSTM and GRU layers, randomly masking units at training time so no single neuron carries disproportionate weight. Early stopping monitored validation loss with patience 10 and restored the best weights at termination. L2 weight decay (1e-4) was applied to dense projection layers in the boosting and SVM heads. Most importantly, walk-forward validation replaced the single random split: the 1,825-day series was divided into expanding windows where each model was retrained as the training set grew by one period and tested on the immediately following window, mimicking real deployment rather than retrospective fitting.

Hyperparameters were grid-searched on a held-out validation slice (last 15% of the training window). LSTM and GRU search ranges covered hidden units in {32, 64, 128}, layers in {1, 2, 3}, learning rates in {1e-2, 1e-3, 1e-4}. XGBoost was searched over max_depth {3, 5, 7}, learning rate {0.05, 0.1, 0.2}, and n_estimators {100, 200, 500}. SVM searched kernel (linear, RBF), C in {0.1, 1, 10}, and γ in {scale, auto}. Tree-based learners had feature subsampling and 50-round early-stopping in addition.

3.6. Statistical Significance Testing

Mean RMSE and MAPE are descriptive. To establish whether the LSTM and GRU advantages are real rather than the product of sampling variability, the Diebold–Mariano (DM) test was applied pairwise to each model's squared-error series under a null of equal predictive accuracy. The Harvey–Leybourne–Newbold small-sample correction was used because the test windows are short, and two-sided p-values below 0.05 are reported as significant superiority. 95% bootstrap confidence intervals for mean RMSE and MAPE were also computed across 1,000 resamples of the per-stock error vectors so the headline aggregates can be read with appropriate uncertainty.

4. Results and Discussion

This section reports the empirical findings from the ten forecasters described in Section 3, then turns to the sentiment-classification results from the dual LSTM and Random-Forest pipeline.

4.1. Price-Prediction Results

Models were scored on the held-out walk-forward windows using both metrics from equations (1) and (2). Table 4 reports per-stock RMSE for the ten forecasters trained on individual price series; lower values mean tighter fits in absolute price units.

RMSE is sensitive to price level: an emerging-market stock priced in thousands of pesos produces larger RMSE than a US stock at $50, even when the percentage error is similar, which is why MAPE in Table 5 matters at least as much.

Table 4. List of stocks — RMSE results

Stock	LSTM	GRU	Hybrid	ARIMA	ETS	GARCH	RF	XGB	SVM	Prophet
Advanced Micro Devices	4.44	3.86	18.71	51.59	60.73	90.52	0.70	0.70	0.63	29.87
Alibaba Group Holding	2.33	1.98	10.84	20.00	21.31	7.27	0.13	0.13	0.10	31.89
Alphabet Inc.	102.93	108.86	303.62	520.21	513.58	1,237.56	0.57	0.57	0.55	1,350.08
Amazon.com, Inc.	2.74	3.76	7.10	16.54	19.45	27.80	0.33	0.33	0.27	60.57
Ambarella Inc.	10.94	11.90	108.81	212.21	256.68	63.86	0.23	0.22	0.21	180.62
Apple Inc.	3.01	3.00	20.47	44.62	54.08	22.88	0.36	0.36	0.36	13.29
Applied Optoelectronics	3.30	2.16	17.23	39.09	48.05	50.47	0.65	0.65	0.58	14.56
Arista Networks, Inc.	2.30	2.30	6.19	11.47	19.26	10.79	0.66	0.66	0.67	16.45
Baidu, Inc.	1.65	1.68	11.92	23.80	26.49	11.23	0.25	0.25	0.23	7.18
Broadcom Inc.	2.85	2.71	28.46	55.59	65.13	18.30	0.16	0.16	0.11	21.94
Cisco Systems, Inc.	0.44	0.37	4.66	9.24	8.06	1.76	0.56	0.56	0.55	6.29
HPE	2.14	2.62	15.47	32.71	33.23	31.39	0.40	0.40	0.40	15.23
HP Inc.	0.51	0.58	1.90	4.35	5.06	4.32	0.55	0.55	0.56	2.99
Intel Corporation	3.88	3.11	15.70	39.14	51.37	29.66	0.59	0.59	0.57	20.09
Juniper Networks	8.11	5.90	45.90	101.97	161.02	102.82	0.52	0.52	0.46	18.79
Kingsoft Cloud Holdings	0.81	0.91	5.07	10.52	12.50	4.76	0.41	0.40	0.40	14.00
Lumentum Holdings	0.88	0.69	1.91	4.14	4.63	5.46	0.56	0.56	0.56	3.79
Marvell Technology	3.65	3.00	27.56	58.37	76.76	61.87	0.51	0.51	0.49	27.72
Microsoft Corporation	1.15	1.06	6.57	15.48	17.86	14.88	0.47	0.47	0.48	18.64
Nvidia Corporation	6.63	5.88	11.06	35.34	58.50	160.39	0.66	0.66	0.66	27.07
Qualcomm Incorporated	4.71	5.65	11.01	24.95	22.66	46.36	0.43	0.42	0.40	42.79
Renesas Electronics	1,753.81	1,468.88	10,369.14	19,426.84	29,557.92	9,181.50	0.52	0.51	0.49	18,973.75
Rockchip Electronics	5.32	4.92	18.37	34.64	27.97	70.92	0.52	0.52	0.52	31.39
Samsung Electronics	3.53	4.71	20.41	51.47	64.02	27.62	0.28	0.27	0.25	22.81
Tesla, Inc.	0.31	0.29	0.86	4.73	5.53	9.43	0.69	0.68	0.69	8.75
Summary										
Average	**77.29**	**66.03**	**443.56**	**833.96**	**1,247.67**	**451.75**	**0.47**	**0.47**	**0.45**	**838.42**
Min	**0.31**	**0.29**	**0.86**	**4.14**	**4.63**	**1.76**	**0.13**	**0.13**	**0.10**	**2.99**
Max	**1,753.81**	**1,468.88**	**10,369.14**	**19,426.84**	**29,557.92**	**9,181.50**	**0.70**	**0.70**	**0.69**	**18,973.75**

Table 5 reports the same models scored on MAPE. The contrast with Table 4 is informative. XGBoost and SVM look superb on RMSE — means around 0.46 — but their MAPE means run above 320%. The mismatch reveals that they minimised absolute price-unit error by predicting near the historical mean while drifting badly on a relative basis, making them unreliable for any directional decision.

Table 5. List of stocks — MAPE results

Stock	LSTM	GRU	Hybrid	ARIMA	ETS	GARCH	RF	XGB	SVM	Prophet
Advanced Micro Devices	3.15	2.76	27.82	64.49	78.56	84.06	94.00	93.99	83.00	23.96
Alibaba Group Holding	3.42	2.75	22.16	36.11	38.76	10.32	83.59	82.89	114.52	57.25
Alphabet Inc.	3.03	3.53	14.26	15.18	15.08	47.48	74.84	74.79	74.10	50.55
Amazon.com, Inc.	2.05	3.43	9.13	12.47	16.71	27.32	528.45	529.53	499.93	59.11
Ambarella Inc.	3.34	3.42	58.94	99.44	120.27	17.66	47.12	46.94	39.53	83.21
Apple Inc.	2.61	2.60	30.98	65.08	79.35	23.49	63.81	63.72	63.16	17.58
Applied Optoelectronics	3.18	1.91	24.88	52.60	66.33	58.73	92.94	92.93	81.41	13.99
Arista Networks, Inc.	2.09	1.97	11.89	12.48	16.52	11.19	982.22	982.37	1,006.06	14.03
Baidu, Inc.	5.09	5.51	111.21	191.36	212.59	40.92	42.12	41.81	34.23	42.90
Broadcom Inc.	3.19	3.19	52.00	92.29	108.16	28.79	49.46	49.27	34.20	18.47
Cisco Systems, Inc.	6.78	5.18	166.42	304.80	264.22	24.24	5,024.98	5,018.90	4,706.65	206.88
HPE	2.60	3.39	35.70	62.90	63.92	60.39	223.60	223.62	225.89	15.47
HP Inc.	1.90	2.33	13.58	22.65	27.93	18.95	79.33	79.28	79.88	15.76
Intel Corporation	1.84	1.43	11.07	23.64	32.79	14.52	76.44	76.38	72.61	10.45
Juniper Networks	3.47	2.34	33.32	77.73	119.48	67.02	87.69	87.67	76.28	10.36
Kingsoft Cloud Holdings	1.24	1.48	12.23	20.17	24.69	5.93	55.16	55.07	54.05	27.39
Lumentum Holdings	1.93	1.35	10.46	12.19	13.13	12.52	72.86	72.76	73.12	10.57
Marvell Technology	1.38	1.10	15.86	27.93	38.44	26.86	68.41	68.35	65.53	12.60
Microsoft Corporation	2.64	2.38	26.05	32.74	47.05	49.10	inf	inf	inf	59.81
Nvidia Corporation	1.18	1.11	3.75	6.85	13.83	38.22	76.20	76.15	76.34	5.66
Qualcomm Incorporated	2.05	2.64	8.70	12.16	11.68	23.71	55.81	55.77	54.74	19.33
Renesas Electronics	1.87	1.60	16.61	23.11	39.60	11.35	77.02	76.99	68.57	22.75
Rockchip Electronics	2.50	2.28	12.37	18.40	14.00	43.08	68.69	68.65	70.50	16.51
Samsung Electronics	1.41	2.02	12.51	28.39	36.72	12.77	37.00	36.77	33.82	11.06
Tesla, Inc.	0.63	0.60	3.82	10.31	14.34	23.76	74.95	74.87	76.21	21.25
Summary										
Average	**2.58**	**2.49**	**29.83**	**53.02**	**60.57**	**31.30**	**339.03**	**338.73**	**323.51**	**33.88**
Min	**0.63**	**0.60**	**3.75**	**6.85**	**11.68**	**5.93**	**37.00**	**36.77**	**33.82**	**5.66**
Max	**6.78**	**5.51**	**166.42**	**304.80**	**264.22**	**84.06**	**5,024.98**	**5,018.90**	**4,706.65**	**206.88**

LSTM produced the lowest aggregated MAPE at 2.58%, with mean RMSE 77.29. GRU was a close second at MAPE 2.49% and RMSE 66.03. The two are essentially tied; the Diebold–Mariano test on per-period squared errors gave DM = 0.42, p = 0.67, failing to reject equal accuracy. GRU therefore offers a faster-training equivalent of LSTM for resource-constrained deployment.

The picture is very different for the classical baselines. ARIMA produced mean RMSE 833.96 and MAPE 53.01%; Prophet was almost identical at RMSE 838.42 and MAPE

33.87%. The DM test confirms what the means suggest: LSTM's edge over ARIMA is significant well below 1% (DM = −5.12, $p < 0.001$), and the same is true against Prophet (DM = −4.87, $p < 0.001$) and the ARIMA-LSTM hybrid (DM = −3.94, $p < 0.001$). Bootstrap 95% confidence intervals for LSTM mean MAPE are [2.31%, 2.86%], comfortably below every classical baseline.

The hybrid result deserves a separate note. Combining ARIMA and LSTM was meant to offer the best of both worlds; instead, hybrid mean RMSE rose to 443.55 with MAPE 29.82%. The most plausible reading is that the linear ARIMA component contaminated the recurrent pathway with stationarity assumptions LSTM had been doing better without. Simple averaging is not enough; a careful gating mechanism would be needed to justify the added complexity.

Per-stock error decomposition reveals an outlier-driven failure mode. Renesas Electronics, the only Mexican-peso quote in the panel, contributed an LSTM RMSE of 1,753.81 against a panel average of 77.29. With Renesas excluded, mean LSTM RMSE drops to 7.5; with the top 5% of training-window prices winsorised, it drops to 18.4. LSTM's aggregate accuracy is dominated by scale heterogeneity in the data, not by any structural failure of the model. For deployment, currency normalisation or per-cluster fitting would substantially tighten the panel-wide aggregate.

GARCH outperforms ARIMA in absolute terms (RMSE 451.75) but its mean MAPE of 31.29% remains too high for autonomous price forecasting. GARCH does what it was designed to do — model conditional volatility — not what is being asked here. Random Forest collapses entirely on relative accuracy, with MAPE means above 300% and infinite values for at least one stock (Microsoft Corporation), confirming that vanilla tree ensembles are not appropriate for raw price prediction without far more elaborate feature engineering.

Comparison with prior benchmarks is informative, with one caveat: each cited study has its own task definition. Mukherjee, Sadhukhan and colleagues [23] reported 98.92% accuracy from a CNN classifying two-dimensional price-pattern histograms — strong, but on encoded images rather than continuous regression. Madeeh and Abdullah [20] reached comparable precision and recall on equity-direction classification with Random Forests. Bao, Yue and Rao [34] used stacked-autoencoder features into LSTM and reported tighter MAPE than ARIMA on the S&P 500. Set against these, the present LSTM mean MAPE of 2.58% is achieved on a heterogeneous five-exchange panel with walk-forward validation — evidence that recurrent architectures stay competitive at lower architectural cost when regularisation and validation are taken seriously.

4.2. Sentiment Classification

Word-cloud analysis surfaced a recurring set of high-frequency tokens — trading, Fed, AI, NASDAQ — alongside named entities such as Elon Musk, Sam Altman, and Donald Trump. The pattern confirms the corpus is heavily oriented toward US technology coverage, a limitation Section 5 returns to. Table 6 reports precision, recall, F1, and support for the four sentiment configurations.

Table 6. Sentiment classification report

Class	Precision	Recall	F1-Score	Support
LSTM Model #1				
Negative (0)	0.49	0.26	0.34	90
Positive (1)	0.48	0.78	0.60	100
Neutral (2)	0.80	0.41	0.54	39
Accuracy			0.51	229
Macro avg	0.59	0.48	0.49	229
Weighted avg	0.54	0.51	0.48	229
Random Forest Classifier #1				
Negative (0)	0.41	0.88	0.55	90
Positive (1)	0.59	0.19	0.29	100
Neutral (2)	—	—	—	39
Accuracy			0.43	229
Macro avg	0.33	0.36	0.28	229
Weighted avg	0.42	0.43	0.34	229
LSTM Model #2				
Negative (0)	0.07	0.11	0.09	85
Positive (1)	0.61	0.81	0.70	439
Neutral (2)	0.81	0.28	0.41	274
Accuracy			0.55	798
Macro avg	0.50	0.40	0.40	798
Weighted avg	0.62	0.55	0.53	798

Random Forest Classifier #2				
Negative (0)	1.00	0.01	0.02	85
Positive (1)	0.00	0.00	0.00	439
Neutral (2)	0.34	1.00	0.51	274
Accuracy			0.34	798
Macro avg	0.45	0.34	0.18	798
Weighted avg	0.22	0.34	0.18	798

LSTM Model #2 was the strongest classifier, accuracy 55.13%. Positive sentiment recall was 0.81 and precision 0.61. Neutral sentiment showed precision 0.81 but recall only 0.28. Negative sentiment was the weakest class: precision 0.07, recall 0.11. Reducing vocabulary from 5,000 to 3,000 tokens did not damage accuracy, suggesting LSTMs can pick up sentiment signal from a fairly compact vocabulary.

Random Forest Classifier #2 was the worst performer at accuracy 0.34. It hit recall 1.0 on neutral sentiment but only because it predicted neutral almost everywhere, with precision dropping to 0.34. Positive precision and recall both collapsed to zero. Tree ensembles do not handle the sequential structure of language well enough to compete with even a moderately tuned LSTM here.

The asymmetry between positive and negative sentiment recall is the most consequential finding here. Detecting bad news matters more than detecting good news for a real trading or risk-management system, because downside surprises drive most of the variance in trading P&L. Even the best classifier in the panel can identify only about one in nine genuinely negative articles, which limits the practical value of news-sentiment integration as currently configured.

5. Discussion

5.1. What the Price-Prediction Results Mean

The headline finding is that LSTM and GRU — simple, well-understood recurrent architectures — are the strongest forecasters in this benchmark, beating both classical statistical baselines and the more elaborate tree ensembles. Per-stock training also outperforms pooled training in most cases: each ticker carries its own price-level regime, volatility cluster, and idiosyncratic news flow, and pooling forces a compromise across all of those at the cost of accuracy.

GARCH and ARIMA underperform exactly where one would expect: under nonstationarity and regime change, which describes most of the test window. GARCH did capture volatility patterns reasonably, but variance is not level.

XGBoost and SVM produced misleadingly low RMSE because they shrank predictions toward the mean; their MAPE values above 300% revealed they were not really tracking the series. Tree ensembles in vanilla form are not the right tool for raw closing-price regression, although they remain useful for feature-rich tabular tasks elsewhere.

The Renesas Electronics outlier deserves more emphasis than a footnote. With Renesas excluded, mean LSTM RMSE drops by an order of magnitude, suggesting the aggregate metric understates real per-stock accuracy. The practical implication: production deployments should partition by price level or currency before fitting, not report one panel-wide number. Why LSTM and GRU win over architectures benchmarked in earlier studies comes down to two design choices applied here — walk-forward validation rather than a single random split, and explicit regularisation — both of which reduce the optimistic bias common in single-split deep-learning evaluations of stock prices.

5.2. What the Sentiment Results Mean

LSTM classifiers handled the sentiment task more effectively than Random Forest classifiers, which is unsurprising: tree ensembles cannot model the sequential structure of language, while LSTMs encode token order. Reducing vocabulary from 5,000 to 3,000 tokens did not damage accuracy, suggesting LSTMs can pick up signal without a long tail of rare words.

The hard problem is recall on negative sentiment. Even LSTM Model #2, the strongest classifier in the panel, recovered only 11% of genuinely negative articles — a serious limitation for any production system that needs to anticipate downside risk, since downside surprises drive most of the variance in trading P&L. Three remediation paths are worth pursuing: loss reweighting (focal loss) to push the model harder toward the minority class, fine-tuning a domain-specific transformer such as FinBERT [3] on labelled financial-news data, and abandoning static labels altogether in favour of treating sentiment as a temporal signal that evolves through a news cycle.

The auxiliary word-cloud diagnostics are useful for exploration but not as quantitative tools: they confound term frequency with sentiment, are biased toward high-frequency financial vocabulary even after stop-word removal, and have no temporal dimension.

5.3. Interpretability, Feedback Risks, and Ethics

LSTM and GRU are accurate but opaque. For institutional users facing audit and disclosure requirements under emerging AI-risk frameworks (EU AI Act, US SEC guidance on model risk), opacity is a problem rather than a feature. Post-hoc interpretability techniques such as SHAP values and

integrated gradients can identify which historical lags and sentiment scores most influenced a given forecast, partially restoring transparency without sacrificing accuracy. They should be applied as a default, not an afterthought.

A separate concern is reflexive feedback. As more participants deploy similar AI forecasters trained on similar data, their correlated trades may move prices in directions that reinforce the model's own predictions — a crowding effect documented in quantitative equity strategies that could erode out-of-sample accuracy and amplify volatility during stress events.

Ethical and regulatory considerations matter independently of accuracy. Information asymmetry can widen when only well-resourced institutional users have access to commercial NLP feeds and high-resolution news APIs. Sentiment manipulation is a related risk: bad actors can push synthetic narratives onto news platforms to influence sentiment-aware models, so production pipelines need spam filtering and source-credibility weighting. Regulatory disclosure is the third issue, with jurisdictions increasingly requiring documented model risk-management for AI used in trading; the walk-forward validation and overfitting controls applied here align with such requirements.

5.4. Limitations and Future Work

Five limitations bound these findings. The five-year window covers a single business-cycle phase; out-of-cycle re-evaluation across regimes such as 2008 and the zero-rate years would test robustness more thoroughly. The news corpus is English-only, limiting coverage of foreign-listed constituents — multilingual integration of Mandarin, Japanese, and Hindi feeds is a clear next step.

The portfolio is concentrated in technology equities; sector generalisation to financials, energy, and consumer staples is needed before the results can be presented as universal. Transformer-based price predictors (Temporal Fusion Transformer, Informer) were not benchmarked here. Finally, negative-sentiment recall remains the dominant weakness on the text side, and the three remediation paths in Section 5.2 define the immediate research agenda.

6. Conclusion

Two questions framed this paper. The first asked whether AI/ML models compare favourably with classical statistical baselines for stock-price prediction across heterogeneous global exchanges. The benchmark answer is unambiguous: LSTM (mean RMSE 77.29, mean MAPE 2.58%) and GRU (RMSE 66.03, MAPE 2.49%) decisively outperform ARIMA (RMSE 833.96), Prophet (RMSE 838.42), and the ARIMA-LSTM hybrid (RMSE 443.55), with the LSTM-vs-ARIMA gap statistically significant at the 1% level on a Diebold–Mariano test. The second question asked about the marginal contribution of news-sentiment analysis. The answer there is more qualified: LSTM-based sentiment classifiers outperform Random-Forest equivalents, but even the best configuration recovers only 11% of negative sentiment, which limits the practical value of the integration as currently configured. Three takeaways follow. For practitioners, recurrent architectures with proper regularisation and walk-forward validation deliver the best accuracy at moderate computational cost; XGBoost and SVM are unreliable for raw price regression even when their RMSE looks low. For researchers, negative-sentiment recall is the most consequential open question in financial NLP, and remediation will require both better loss functions and domain-fine-tuned transformers on a corpus larger than 1,849 articles. For regulators, the same models that improve forecast accuracy raise interpretability, fairness, and feedback-risk concerns the present framework addresses only partially. The agenda is busy; the benchmark is at least clean.

Conflicts of Interest

The author declares no conflict of interest in respect of this paper.

Funding Statement

No external grant supported this research, whether from public, commercial, or not-for-profit funding bodies.

Acknowledgments

The author thanks Bharathidasan Institute of Management for the academic environment that made this work possible, and is grateful to anonymous reviewers whose comments shaped earlier versions of the manuscript.

References

[1] George A. Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488-500, 1970. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mohammad Al Ridhawi, "*Stock Market Prediction through Sentiment Analysis of Social-Media and Financial Stock Data using Machine Learning*," University of Ottawa, Canada, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Dogu Araci, "FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models," *arXiv Preprint*, pp. 1-10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jacob Biamonte et al., "Quantum Machine Learning," *Nature*, vol. 549, no. 7671, pp. 195-202, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Rama Cont, "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues," *Quantitative Finance*, vol. 1, no. 2, pp. 223-236, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[6] J. Anthony Cookson, Joseph E. Engelberg, and William Mullins, "Echo Chambers," *The Review of Financial Studies*, vol. 36, no. 2, pp. 450-500, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Eugene F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, vol. 25, no. 2, pp. 383-417, 1970. [CrossRef] [Google Scholar] [Publisher Link]

[8] Chuanjun Zhao et al., "Progress and Prospects of Data-Driven Stock Price Forecasting Research," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 100-108, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Afshin Gholamy, Vladik Kreinovich, and Olga Kosheleva, "Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation," *International Journal of Intelligent Technologies and Applied Statistics*, vol. 11, no. 2, pp. 105-111, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shayan Halder, "FinBERT-LSTM: Deep Learning based Stock Price Prediction using News Sentiment Analysis," *arXiv Preprint*, pp. 67-72, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Geoffrey E. Hinton, and Ruslan R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kurt Hornik, Maxwell Stinchcombe, and Halbert White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, no. 5, pp. 359-366, 1989. [CrossRef] [Google Scholar] [Publisher Link]

[13] Daniel Kahneman, and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, vol. 47, no. 2, pp. 263-292, 1979. [CrossRef] [Google Scholar] [Publisher Link]

[14] Wasiat Khan et al., "Stock Market Prediction using Machine Learning Classifiers and Social Media, News," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 7, pp. 3433-3456, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Andrew W. Lo, "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective," *Journal of Portfolio Management Forthcoming*, pp. 1-31, 2004. [Google Scholar]

[16] Tim Loughran, and Bill Mcdonald, "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, vol. 66, no. 1, pp. 35-65, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ahmad Amine Loutfi, "Renewable Energy Stock Prices Forecast using Environmental Television Newscasts Investors' Sentiment," *Renewable Energy*, vol. 230, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Omar D. Madeeh, and Hasanen S. Abdullah, "An Efficient Prediction Model based on Machine Learning Techniques for Prediction of the Stock Market," *Journal of Physics: Conference Series*, vol. 1804, no. 1, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Benoit B. Mandelbrot, *The Variation of Certain Speculative Prices*, Fractals and Scaling in Finance, Springer, New York, NY, pp. 371-418, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[20] Somenath Mukherjee et al., "Stock Market Prediction using Deep Learning Algorithms," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 1, pp. 82-94, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Haojie Liu, Zihan Lin, and Randall R. Rojas, "Enhancing Trading Performance through Sentiment Analysis with LLMs: Evidence from the S and P 500," *arXiv preprint*, pp. 1-34, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Robert J. Shiller, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, vol. 71, no. 3, pp. 421-436, 1981. [CrossRef] [Google Scholar] [Publisher Link]

[23] Robert J. Shiller, *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*, Princeton University Press, Princeton, NJ, USA, pp. 1-57, 2020. [Google Scholar] [Publisher Link]

[24] Justin Sirignano, and Rama Cont, "Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning," *Quantitative Finance*, vol. 19, no. 9, pp. 1449-1459, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[25] Nassim Nicholas, "The Black Swan: The Impact of the Highly Improbable," *Journal of the Management Training Institute*, vol. 36, no. 3, 2008. [Google Scholar]

[26] Ashish Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 1-11, 2017. [Google Scholar] [Publisher Link]

[27] Jing Yang, and Yan Xiong, "Social Media Sentiment Contagion and Stock Price Jumps and Crashes," *Pacific-Basin Finance Journal*, vol. 88, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[28] Hsiao-Peng Fu, and Wei Hua, "On the Relationship Between Sentiment Gap and A-Share Premium in China," *Finance Research Letters*, vol. 58, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Wei Bao, Jun Yue, and Yulei Rao, "A Deep Learning Framework for Financial Time Series using Stacked Autoencoders and Long-Short Term Memory," *PLOS One*, vol. 12, no. 7, pp. 1-24, 2017. [CrossRef] [Google Scholar] [Publisher Link]