Data Science in Finance and Economics

https://www.aimspress.com/journal/dsfe

DSFE, 6(2): 250–276.
DOI:10.3934/DSFE.2026009
Received: 21 August 2025
Revised: 25 February 2026
Accepted: 13 March 2026
Published: 06 May 2026

Research article

Sentiment-enhanced rice price forecasting under sparse social-media coverage: Evidence from Saudi rice imports

Saad Alqithami* **and Musaad Alzahrani**

Department of Computer Science, Al-Baha University, Albaha 65779, Saudi Arabia

* **Correspondence:** Email: salqithami@bu.edu.sa.

This supplementary document accompanies the accepted manuscript . To keep the supplement tightly aligned with the final camera-ready paper, it focuses on four reproducibility items referenced in the main text: (i) a detailed literature table that expands the compact reviews in Tables 1–2, (ii) the GPT-4 prompt template, inference settings, and label-audit protocol used for tweet-level sentiment labeling, (iii) the month-level tweet-count and sentiment aggregates from which the lagged sentiment features were constructed, and (iv) the model-selection spaces used for the five reported learners. Consistent with the paper's data-availability statement, raw tweet text is not redistributed here; reproducibility instead relies on tweet IDs and collection scripts in the public repository together with the prompt details and aggregate tables documented below.

S1. Detailed literature review supporting Tables 1–2 in the main paper

Table S1 expands the compact literature summaries in Tables 1 and 2 of the main paper. Preference was given to agricultural or commodity applications that use exogenous variables, sentiment, or multimodal data fusion, while a small number of foundational method references were retained only when they directly support the modeling and evaluation choices used in the final manuscript. The table is illustrative rather than exhaustive, but it is aligned to the studies explicitly emphasized in the accepted paper.

Table S1. Detailed literature table supporting the focused reviews in Tables 1 and 2 of the main paper. Foundational references are retained only when they directly underpin the models or evaluation protocol used in this study.

Study	Market / context	Method	Data modalities	Key takeaway
Panel A: Statistical and econometric foundations				
Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.	General time-series forecasting	ARIMA, SARIMA	Univariate time series	Foundational reference for autoregressive and seasonal modeling; provides the classical statistical framework that motivates the SARIMAX benchmark used in the main paper.
Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts, Melbourne, Australia. `OTexts.com/fpp3`.	General forecasting	Forecast evaluation; rolling and expanding-window validation	Time series	Establishes best practices for time-ordered validation, forecast accuracy assessment, and regression models with ARIMA errors.
Putra, A. W., Supriatna, J., Koestoer, R. H., & Soesilo, T. E. B. (2021). Differences in local rice price volatility, climate, and macroeconomic determinants in the Indonesian market. *Sustainability*, 13(8), 4465. doi: 10.3390/su13084465.	Indonesia (rice)	Volatility modeling	Prices, climate, and macroeconomic variables	Shows that climate and macroeconomic drivers of rice-price volatility differ across local markets, highlighting the importance of context-specific exogenous variables.
Yadav, A. (2024). A comparative study of time series, machine learning, and deep learning models for forecasting global price of wheat. *SN Operations Research Forum*, 5, 113. doi: 10.1007/s43069-024-00395-9.	Global wheat	ARIMA/ARIMAX, ML, and DL comparison	Wheat prices plus exogenous commodity and macroeconomic variables	Finds no single universal winner, but models that incorporate external variables outperform simpler univariate baselines; among purely univariate time-series models, SARIMA performs best.
Panel B: Machine-learning and deep-learning forecasting studies				
Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi: 10.1023/A:1010933404324.	General predictive modeling	Random Forest	Tabular predictors	Introduces bagged tree ensembles that capture nonlinear interactions and provide variable-importance measures with comparatively weak parametric assumptions.
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). doi: 10.1145/2939672.2939785.	General predictive modeling	Gradient boosting (XGBoost)	Tabular predictors	Provides a scalable and regularized boosting framework that remains a strong benchmark for structured forecasting data.
Sari, M., Duran, S., Kutlu, H., Guloglu, B., & Atik, Z. (2024). Various optimized machine learning techniques to predict agricultural commodity prices. *Neural Computing and Applications*, 36(19), 11439–11459. doi: 10.1007/s00521-024-09679-x.	Multiple agricultural commodities	GA-ELM, GA-LSTM, and ARIMA comparison	Historical price series for 11 agricultural commodities	Reports that the GA-ELM model outperforms both GA-LSTM and ARIMA across the studied commodities and remains robust during the large price disruptions observed in 2022.
Wang, Z., French, N., James, T., Schillaci, C., Chan, F., Feng, M., & Lipani, A. (2023). Climate and environmental data contribute to the prediction of grain commodity prices using deep learning. *Journal of Sustainable Agriculture and Environment*, 2(3), 251–265. doi: 10.1002/sae2.12041.	U.S. grain markets	CNN-LSTM hybrid	Price, climate, and environmental variables	Shows that adding climate and environmental covariates improves deep-learning forecasts of grain commodity prices relative to price-only baselines.
Panel C: Sentiment analysis and multimodal fusion				

Continued on next page

Study	Market / context	Method	Data modalities	Key takeaway
Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168. doi: 10.1111/j.1540-6261.2007.01232.x.	Equity markets	Media-tone regressions	News tone and market activity	Shows that media pessimism contains predictive information for short-run market outcomes, providing an early foundation for text-based sentiment as an economic signal.
Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. doi: 10.1016/j.jocs.2010.12.007.	Equity markets	Mood indices and predictive modeling	Twitter sentiment and market data	Demonstrates that aggregated social-media mood dimensions can improve short-horizon market prediction, motivating the use of behavioral signals in forecasting.
Xu, J.-L., & Hsu, Y.-L. (2022). The impact of news sentiment indicators on agricultural product prices. *Computational Economics*, 59(4), 1645–1657. doi: 10.1007/s10614-021-10189-4.	Agricultural products	Linear and quantile regression	News sentiment, weather, oil prices, and product prices	Shows that adding sentiment scores and oil prices improves price prediction; linear regression and recursive-window evaluation outperform the alternatives studied.
Ewald, C. O., & Li, Y. (2024). The role of news sentiment in salmon price prediction using deep learning. *Journal of Commodity Markets*, 36, 100438. doi: 10.1016/j.jcomm.2024.100438.	Salmon market	CNN-LSTM and other deep-learning models	Historical prices and news sentiment	Finds that deep-learning models outperform traditional methods in salmon-price prediction and that sentiment features reduce prediction errors further.
Bonato, M., Cepni, O., Gupta, R., & Pierdzioch, C. (2024). Forecasting the realized volatility of agricultural commodity prices: Does sentiment matter? *Journal of Forecasting*, 43(6), 2088–2125. doi: 10.1002/for.3106.	Agricultural commodities	HAR-type realized-volatility forecasting	Realized volatility and sentiment indicators	Evaluates the out-of-sample predictive power of sentiment for agricultural commodity volatility forecasting and shows that sentiment can add incremental information.
An, W., Wang, L., & Zeng, Y.-R. (2024). Social media-based multi-modal ensemble framework for forecasting soybean futures price. *Computers and Electronics in Agriculture*, 226, 109439. doi: 10.1016/j.compag.2024.109439.	Soybean futures	Multimodal ensemble forecasting	News text, search intensity, market variables, and historical prices	Constructs an investor-concern sentiment index and shows that a multimodal ensemble improves short- and medium-term soybean-futures forecasts.
Reis Filho, I. J., Ricardino, M. M., & Rezende, S. O. (2022). On the enrichment of time series with textual data for forecasting agricultural commodity prices. *MethodsX*, 9, 101758. doi: 10.1016/j.mex.2022.101758.	Agricultural commodities	RF and SVR with textual enrichment	Price series and keyword-derived text features	Shows that enriching agricultural commodity time series with text-derived indicators can improve commodity-price forecasting, supporting the use of textual side information in corn and soy settings.
Wang, W., & Liu, Y. (2025). A novel framework for agricultural futures price prediction with BERT-based topic identification and sentiment analysis. *Journal of Forecasting*, 44(6), 1969–1992. doi: 10.1002/for.3278.	Soybean futures	BERTopic, FinBERT, LSTM, and HMM-LSTM	Market data and agricultural news headlines	Shows that topic-aware sentiment features provide substantial incremental information for price prediction across lags, with HMM-LSTM performing especially well at medium and long horizons.

S2. GPT-4 labeling protocol

S2.1. Prompt template

The verbatim prompt template below was used to assign tweet-level sentiment labels. It is reproduced as used in the analysis so that the labeling step can be audited and replicated.

```
As an expert financial analyst specializing in agricultural commodities,
please classify the sentiment of the following tweet regarding rice in Saudi Arabia.
The sentiment should reflect the likely impact on consumer demand for rice.
The tweet is in Arabic.

Tweet: "[Tweet Text]"

Categories:
1. Positive: Expresses positive sentiment, such as satisfaction with quality,
   good harvest news, or excitement about new availability. This would likely
   increase demand.
2. Negative: Expresses negative sentiment, such as complaints about price, quality,
   or availability, or news of a bad harvest. This would likely decrease demand.
3. Neutral: The tweet is informational, a question, or does not express a clear
   sentiment that would impact demand.

Sentiment:
```

S2.2. Inference settings

- Model identifier recorded in the analysis scripts: `gpt-4`
- One tweet submitted per API request
- Temperature: 0.2
- Maximum tokens: 5
- Output constrained to exactly one of three labels: `positive`, `neutral`, or `negative`

S2.3. Label-audit protocol

A separate 500-tweet audit sample was drawn from the pre-filter query corpus (i.e., before the engagement filter was applied). The two authors independently annotated that holdout without access to GPT-4 outputs, after which disagreements were resolved by discussion to form an adjudicated human reference. The reported 89% agreement and Cohen's $\kappa = 0.81$ therefore refer to agreement between GPT-4 and the adjudicated human labels. The audit sample was used only for validation and was not included in the monthly aggregates or forecasting models. After classification, the three returned labels were mapped to numeric polarity values $\{-1, 0, +1\}$ for monthly aggregation.

S3. Sentiment data coverage and monthly tweet counts

Table S2 reports the raw month-end tweet counts N_t and sentiment aggregates used to construct the forecasting features for the full January 2015–January 2024 panel (109 months). The table is intentionally *pre-feature-engineering*: the four-month shift, the additional 1- and 2-month lags, and the six-month trailing mean described in the main paper are all applied downstream to these monthly aggregates. A month is marked as having observed sentiment signal if the binary availability indicator I_t equals 1, meaning $N_t > 0$. Months with $N_t = 0$ have no retained tweets after the relevance and engagement filters; in those months the displayed AS_t and WS_t values are placeholders (set to zero in the released file) and should be interpreted as missing rather than neutral. The forecasting models therefore include both I_t and $\log(1 + N_t)$ so that months with no observed text signal are distinguishable from months with genuinely near-neutral sentiment. Across the panel, 39 months contain at least one retained tweet (total retained tweets = 69; max $N_t = 4$; median $N_t = 0$).

Table S2. Monthly tweet counts and month-end sentiment aggregates used to construct the forecasting features.

Month	N_t	I_t	AS_t	WS_t
2015-01	0	0	0.000	0.000
2015-02	0	0	0.000	0.000
2015-03	0	0	0.000	0.000
2015-04	0	0	0.000	0.000
2015-05	0	0	0.000	0.000
2015-06	1	1	-1.000	-8.114
2015-07	0	0	0.000	0.000
2015-08	0	0	0.000	0.000
2015-09	1	1	0.000	0.000
2015-10	0	0	0.000	0.000
2015-11	1	1	0.000	0.000
2015-12	2	1	0.000	-0.994
2016-01	0	0	0.000	0.000
2016-02	0	0	0.000	0.000
2016-03	3	1	-0.333	-2.245
2016-04	1	1	1.000	7.219
2016-05	2	1	1.000	7.259
2016-06	0	0	0.000	0.000
2016-07	2	1	0.000	0.011
2016-08	1	1	-1.000	-7.347
2016-09	0	0	0.000	0.000
2016-10	0	0	0.000	0.000
2016-11	0	0	0.000	0.000
2016-12	1	1	-1.000	-7.040

Continued on next page

Month	N_t	I_t	AS_t	WS_t
2017-01	0	0	0.000	0.000
2017-02	1	1	1.000	6.513
2017-03	0	0	0.000	0.000
2017-04	0	0	0.000	0.000
2017-05	2	1	1.000	6.082
2017-06	0	0	0.000	0.000
2017-07	0	0	0.000	0.000
2017-08	3	1	0.333	2.042
2017-09	1	1	0.000	0.000
2017-10	0	0	0.000	0.000
2017-11	0	0	0.000	0.000
2017-12	0	0	0.000	0.000
2018-01	0	0	0.000	0.000
2018-02	0	0	0.000	0.000
2018-03	0	0	0.000	0.000
2018-04	0	0	0.000	0.000
2018-05	0	0	0.000	0.000
2018-06	0	0	0.000	0.000
2018-07	1	1	1.000	6.914
2018-08	0	0	0.000	0.000
2018-09	0	0	0.000	0.000
2018-10	0	0	0.000	0.000
2018-11	0	0	0.000	0.000
2018-12	0	0	0.000	0.000
2019-01	0	0	0.000	0.000
2019-02	0	0	0.000	0.000
2019-03	0	0	0.000	0.000
2019-04	0	0	0.000	0.000
2019-05	0	0	0.000	0.000
2019-06	0	0	0.000	0.000
2019-07	0	0	0.000	0.000
2019-08	0	0	0.000	0.000
2019-09	0	0	0.000	0.000
2019-10	2	1	0.000	-0.090
2019-11	0	0	0.000	0.000
2019-12	0	0	0.000	0.000
2020-01	2	1	0.000	0.242
2020-02	0	0	0.000	0.000
2020-03	4	1	0.750	5.918
2020-04	4	1	0.750	5.319

Continued on next page

Month	N_t	I_t	AS_t	WS_t
2020-05	1	1	1.000	6.429
2020-06	1	1	0.000	0.000
2020-07	0	0	0.000	0.000
2020-08	0	0	0.000	0.000
2020-09	0	0	0.000	0.000
2020-10	3	1	-1.000	-8.290
2020-11	2	1	0.500	3.675
2020-12	0	0	0.000	0.000
2021-01	0	0	0.000	0.000
2021-02	0	0	0.000	0.000
2021-03	0	0	0.000	0.000
2021-04	1	1	1.000	6.953
2021-05	0	0	0.000	0.000
2021-06	0	0	0.000	0.000
2021-07	0	0	0.000	0.000
2021-08	1	1	-1.000	-11.166
2021-09	0	0	0.000	0.000
2021-10	0	0	0.000	0.000
2021-11	0	0	0.000	0.000
2021-12	1	1	1.000	8.345
2022-01	0	0	0.000	0.000
2022-02	1	1	0.000	0.000
2022-03	0	0	0.000	0.000
2022-04	3	1	1.000	8.184
2022-05	3	1	-0.667	-5.538
2022-06	1	1	1.000	7.111
2022-07	0	0	0.000	0.000
2022-08	0	0	0.000	0.000
2022-09	1	1	0.000	0.000
2022-10	2	1	0.000	-0.019
2022-11	1	1	-1.000	-9.648
2022-12	2	1	0.500	3.238
2023-01	0	0	0.000	0.000
2023-02	0	0	0.000	0.000
2023-03	0	0	0.000	0.000
2023-04	0	0	0.000	0.000
2023-05	0	0	0.000	0.000
2023-06	0	0	0.000	0.000
2023-07	0	0	0.000	0.000
2023-08	0	0	0.000	0.000

Continued on next page

Month	N_t	I_t	AS_t	WS_t
2023-09	3	1	1.000	7.210
2023-10	1	1	1.000	7.868
2023-11	2	1	0.500	4.157
2023-12	2	1	-0.500	-3.744
2024-01	2	1	0.500	3.727

S4. Hyperparameter settings used in the experiments

Table S3 documents the search spaces used for model selection in the final paper. To match the camera-ready main text, OLS, Ridge, Gradient Boosting, and Random Forest were tuned on the common robust-scaled non-SARIMAX design matrix, whereas SARIMAX used the reduced exogenous block described in Section 3.3.1 of the manuscript. Candidate configurations were evaluated with time-ordered validation so that every training fold preceded its corresponding validation fold.

These search settings correspond to the five model families reported in Table 4 of the main paper. The rolling-origin robustness checks discussed in Section 5.1 of the manuscript re-estimate a matched climate-only versus sentiment-augmented SARIMAX pair rather than conducting a new cross-family search.

Table S3. Hyperparameter search spaces and model-selection protocol used in the experiments. Unless otherwise noted, model selection minimized validation RMSE under expanding-window time-series cross-validation.

Model	Search space / selection details
SARIMAX (with exogenous regressors)	Candidate non-seasonal orders $p, q \leq 3$ and seasonal orders $P, Q \leq 2$ (season length $s = 12$) were evaluated in a stepwise search. Final specifications were selected by minimizing AIC subject to residual adequacy checks. The reduced exogenous block comprised contemporaneous temperature levels, the 12-month trailing mean of TAVG, and—for the sentiment-augmented specification—the four-month-lagged weighted-sentiment series and its first difference. Estimation used `statsmodels` with `fit(disp=False)`.
Gradient Boosting (GBM)	Randomized search over 300 draws on the common robust-scaled non-SARIMAX design matrix with learning rate $\in \{0.01, 0.05, 0.10, 0.20\}$, max depth $\in \{3, 4, 5, 6\}$, subsample $\in \{0.6, 0.8, 1.0\}$, and number of estimators up to 600. Each draw was evaluated with five-fold expanding-window time-series validation. Early stopping (patience = 50 rounds) was used to limit unnecessary complexity.
Random Forest (RF)	Randomized search over 200 draws on the common robust-scaled non-SARIMAX design matrix, tuning max depth (5–20), minimum samples per split (2–10), and maximum number of leaf nodes (50–300). Forest size was fixed at 1,000 trees, and the number of predictors sampled at each split followed the standard $\sqrt{p}$ heuristic. Validation used the same five-fold expanding-window scheme as GBM.
Ridge Regression	The penalty parameter λ was selected on a logarithmic grid, $10^{-3} \leq \lambda \leq 10^{2}$, using leave-one-year-out time-series cross-validation on the common robust-scaled non-SARIMAX design matrix.
Linear Regression (OLS)	No hyperparameters were tuned beyond the common preprocessing pipeline. The model was estimated in closed form on the same robust-scaled non-SARIMAX design matrix used for Ridge, GBM, and RF.

AIMS Press