Finance Research Letters 57 (2023) 104251

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Impactful messaging: Elite sentiment in Chinese new energy vehicle vs machine learning perspective

Xingyue Gong, Guozhu Jia *

College of physical and Electronics Engineering, Sichuan Normal University, Chengdu 610000, China

ARTICLE INFO

Keywords:
Elite sentiment
Machine learning
New energy vehicle
CEEMDAN
Prediction

ABSTRACT

Elite messaging is sensitive to framing decisions and shaping public sentiment. This study examines the influence of elite sentiment on the pricing of the stock index of China's new energy vehicles (NEVs). To construct elite sentiment indexes for China's NEVs, data from a highly active online elite forum were collected and a series of machine learning models were utilized to generate predictions. The results demonstrate that the elite sentiment index has a pronounced predictive impact on stock index prices. Furthermore, incorporating the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method can further enhance the baseline model's predictive ability.

1. Introduction

Sentiment analysis has proven to be a valuable tool for asset pricing in financial research (Nyakurukwa and Seetharam, 2023). Many studies have analyzed sentiment in news and social media through text mining with respect to, for example, the crude oil market (Akhtar et al., 2013; Li et al., 2019), the gold futures market (Smales, 2014), and the stock market (AlZaabi, 2021; Nyakurukwa and Seetharam, 2022). These investigations provide a holistic examination of public sentiment and viewpoints, underscoring the pivotal influence of psychological aspects in the financial market. However, the growing popularity and use of internet technologies, especially social networking platforms, have presented multiple challenges in the construction of sentiment indices; the primary limitation stems from the potential noise and lack of representativeness in sentiment data collected through large-scale online measurements from channels such as social media. Efficiently extracting useful information from the pool of data is crucial for addressing this issue. The information promulgated by the elites on social media is commonly deemed to possess greater value due to their elevated levels of social influence and credibility (Box-Steffensmeier and Moses, 2021; Sun et al., 2020). Compared to regular social media users, sentiment data from elite users exhibit higher accuracy and representativeness (Deng et al., 2018; Sun et al., 2020).

New energy vehicles (NEVs) have emerged as a prominent topic that consistently captures the attention of the elite group, which includes industry experts in politics, business, and media, owing to NEVs environmentally friendly nature and innovative technologies. First, the elite group generally emphasizes environmental conservation and sustainable development, and NEVs offer zero emissions, which aligns with their environmental ideals and values. Second, NEVs are equipped with innovative and forward-looking electric propulsion systems and related technologies, setting them apart from conventional fuel-powered vehicles. Various members of the elite group hold a keen interest and curiosity in these novel technologies and products. Third, the NEV industry introduces massive potential economic opportunities and market growth. The elites' focus on NEVs and their dissemination of pertinent information on social media

* Corresponding author.
E-mail address: 309739124@qq.com (G. Jia).

https://doi.org/10.1016/j.frl.2023.104251
Received 20 June 2023; Received in revised form 22 July 2023; Accepted 23 July 2023
Available online 24 July 2023
1544-6123/

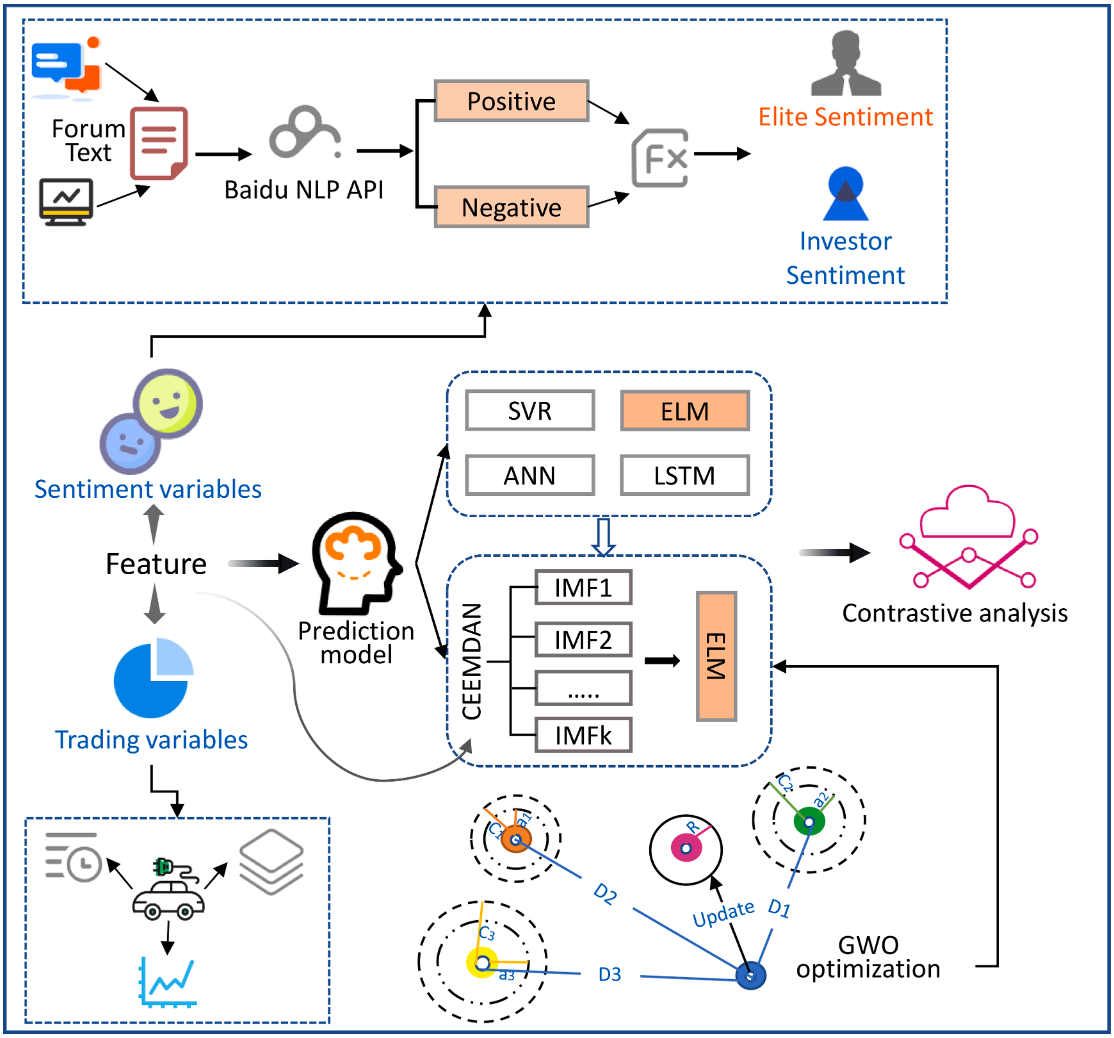

Fig. 1. Framework of main methodologies.

also influence the public's interest and engagement (Box-Steffensmeier and Moses, 2021). This study presents a straightforward and practical sentiment strategy that integrates the baseline model (hybrid machine learning) and demonstrates its applicability in the Chinese NEV stock market. By utilizing the mentioned methodology, we evaluate the impact of the elite sentiment index on the Chinese NEV stock price index. Furthermore, we contrast it with text-based investor sentiment, providing further empirical evidence to support the effectiveness of the elite sentiment index in stock index forecasting.

This study makes three contributions. First, it constructs the elite sentiment index for the Chinese NEV market. Second, the study provides the first empirical validation of the effectiveness of elite sentiment in stock price index forecasting. The elite sentiment index integrates the comprehensive features of investor, public, and user sentiments. The construction method of this index is more streamlined and efficient compared to traditional sentiment indices. Third, this study proposes the optimized complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)-ELM model that incorporates elite sentiment to enhance the forecasting accuracy of the Chinese NEV stock index.

The study is structured as follows. Section 2 presents the data, variables, and models; Section 3 examines the main empirical results; and Section 4 provides the conclusion. Fig. 1 visually depicts the study's complete framework.

Table 1
Variable definitions.

Variables	Definitions
Trading variables	
Open	Daily opening price of New Energy Vehicle (NEV)
High	Daily highest price of NEV
Low	Daily lowest price of NEV
Close	Daily closing price of NEV
Change	Rise or fall in price of NEV
Change (%)	Percentage change of NEV
Volume shares	Daily trading volume of NEV
Turnover	Daily trading value of NEV
Sentiment index variables	
Investor sentiment	Sentiment of investors on the stock price of NEV stocks in the Dong-Fang Wealth Stock Forum
Elite sentiment	Sentiment of elite users on the NEV forum in "SMTH"

Table 2
Descriptive statistics.

Variable	Obs	Mean	Std	Min	Max	Skewness	Kurtosis	Jarque–Bera
Open	1156	3144.894	1646.711	1263.74	6643.26	0.635	1.931	132.726*
High	1156	3190.836	1675.359	1310.79	6695.39	0.629	1.916	132.709*
Low	1156	3097.191	1613.826	1261.76	6563.37	0.636	1.937	132.367*
Close	1156	3145.860	1645.479	1289.75	6632.51	0.631	1.926	111.596*
Change	1156	1.889	84.035	−351.49	476.62	0.174	6.379	555.901*
Change (%)	1156	0.083	2.208	−7.62	8.23	0.073	4.022	51.426*
Volume shares	1156	742.717	393.175	160.17	2396.91	0.757	3.166	111.596*
Turnover	1156	297.179	241.653	30.51	1129.21	0.895	2.827	155.795*
Investor sentiment	1156	−0.391	0.808	−1.54	3.21	0.597	4.944	250.851*
Elite sentiment	1156	−0.202	0.436	−3.09	1.79	0.238	3.584	27.393*

Note. * denotes significance at the 1% level.

2. Data and method

2.1. Data

This study utilizes the NEV stock price data obtained from Csindex,[1] specifically the CSI New Energy Vehicle Index (399,976), covering the period from January 2, 2018, to September 30, 2022. This dataset includes various variables such as Open, Close, High, Low, Change, Change (%), Volume Shares, and Turnover of NEV. These variables serve as the trading variables in the analysis (Table 1).

2.1.1. Constructing the sentiment index

The analysis includes two sentiment indices, the elite sentiment index and investor sentiment index. The elite sentiment index is the focus of our research. The Chinese internet forum "SMTH" serves as a widely utilized source of elite information due to its status as one of China's earliest and most representative online communities (Wu, 2014). Attracting a considerable number of technical elites and intellectuals, it is a community predominantly composed of highly skilled and knowledgeable individuals (Yuan, 2018). We developed a web crawler to download the titles of NEV posts from "SMTH".[2] To measure investor sentiment regarding the Chinese NEV stock market, investor comments were collected from the Dong-Fang Wealth Stock Forum[3] for the GF CSI New Energy Vehicle Index. From January 2, 2018, to September 30, 2022, we filtered out comments from collected elite and investor reviews that contained only punctuation marks, emoticons, or single words. This resulted in 55,719 and 39,357 pieces of valid data, respectively. We calculated the sentiment scores of the elite users and investors using Baidu Intelligent Cloud sentiment analysis and calculated the daily sentiment index using the following formula.

$$MI_i = \ln \frac{1 + M_i^{pos}}{1 + M_i^{neg}} \tag{1}$$

where MI_i refers to the sentiment index at time t and M_i^{pos} and M_i^{neg} are the number of positive and negative comments in day i, respectively.

[1] Source of NEV data: https://www.csindex.com.cn/
[2] https://www.newsmth.net/
[3] https://so.eastmoney.com/

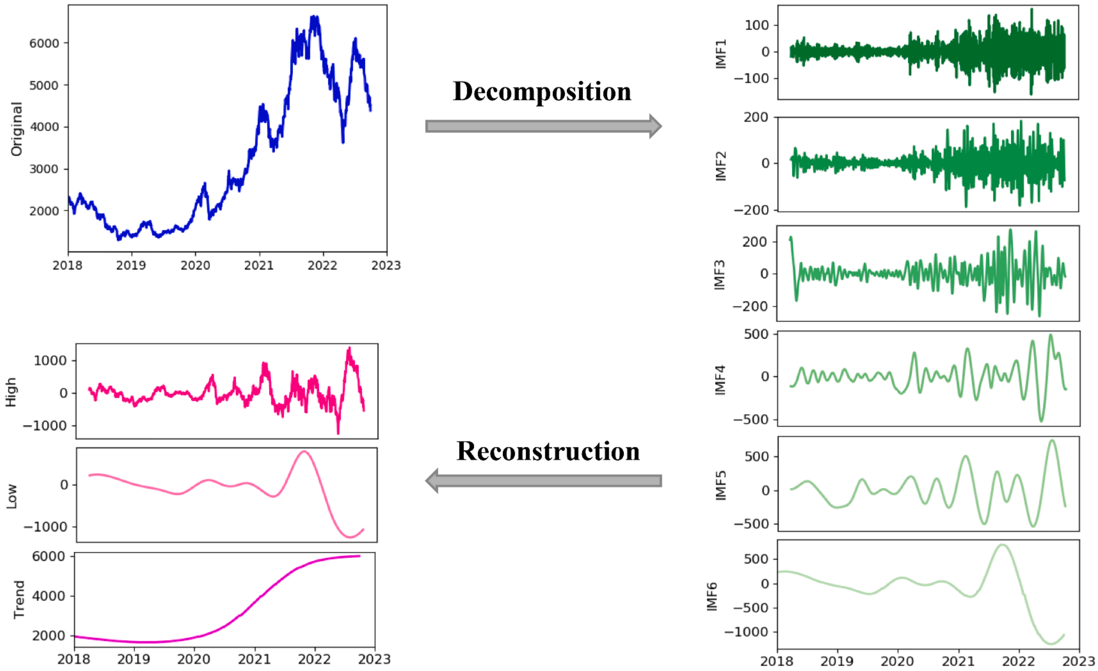

Fig. 2. The decomposition and reconstruction process for China's NEV stock index.

2.1.2. *Summary statistics*

As shown in Table 2, the mean of all the sentiment indices is negative. The Jarque–Bera test indicates that these variables follow a normal distribution, rejecting the assumption of normality at a 1% significance level.

2.2. *Methodology*

With the development of artificial intelligence, machine learning has been extensively applied in the prediction of financial time series (Chen et al., 2017; Hsu et al., 2016; Rather et al., 2015). To evaluate the effectiveness of the elite sentiment-driven approach, we selected a set of representative machine learning models for comparison. SVR and ELM were chosen due to their wide applicability (Ghoddusi et al., 2019), while ANN and LSTM were selected due to having been extensively used in financial time series forecasting over the past decade (Leippold et al., 2022; Sezer et al., 2020).

To enhance the predication accuracy and further analyze the impact of elite sentiment indexes on stock prices of NEVs from different time domains and frequencies we integrated the aforementioned high-performing baseline models with the CEEMDAN method to establish a hybrid machine learning model. CEEMDAN is based on the traditional empirical mode decomposition method (Huang et al., 1998) and was introduced by Torres (Torres et al., 2011) and decomposes complex sequences into subsequences of different time scales and subsequently reconstructs these subsequences into high-frequency, low-frequency, and trend components (Zhang et al., 2008) that exhibit more regular and stable fluctuations (Fig.2). For more details, refer to Appendix A.

2.3. *Evaluating the models*

To assess the models' predictive performance, mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) were used. The formulae are as follows:

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left|\mathrm{y}_i^1 - \mathrm{y}_i^2\right| \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left(\mathrm{y}_i^1 - \mathrm{y}_i^2\right)^2} \tag{7}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\frac{\left|\mathrm{y}_i^1 - \mathrm{y}_i^2\right|}{\mathrm{y}_i^2} \tag{8}$$

Table 3
Input data settings description.

Title	Input	Prediction
First group	Trading variables	Close
Second group	Trading variables + Investor sentiment	
Third group	Trading variables + Elite sentiment	
Fourth group	Trading variables +Investor sentiment+ Elite sentiment	

Table 4
One-day-ahead forecasting performance comparisons by different models and input indexes.

Model	Evaluation indicators	First group	Second group	Third group	Fourth group
SVR	MAE	115.4814	114.4066	109.0799	110.3854
	RMSE	142.4050	140.2822	135.0652	137.4268
	MAPE	0.1499	0.1493	0.1505	0.1504
	R^2	0.9575	0.9588	0.9618	0.9604
ELM	MAE	101.2469	101.3895	99.3157	99.1176
	RMSE	127.8230	126.4622	124.9894	124.5888
	MAPE	0.0195	0.0195	0.0191	0.0190
	R^2	0.9653	0.9661	0.9669	0.9671
ANN	MAE	116.8917	112.6133	105.3632	116.5136
	RMSE	145.2015	140.0555	132.1916	144.7222
	MAPE	0.0222	0.0215	0.0201	0.0222
	R^2	0.9549	0.9582	0.9629	0.9553
LSTM	MAE	128.5373	141.2814	122.2213	119.7824
	RMSE	159.0250	173.4702	152.3665	148.5141
	MAPE	0.0246	0.0268	0.0234	0.0229
	R^2	0.9457	0.9326	0.9509	0.9530

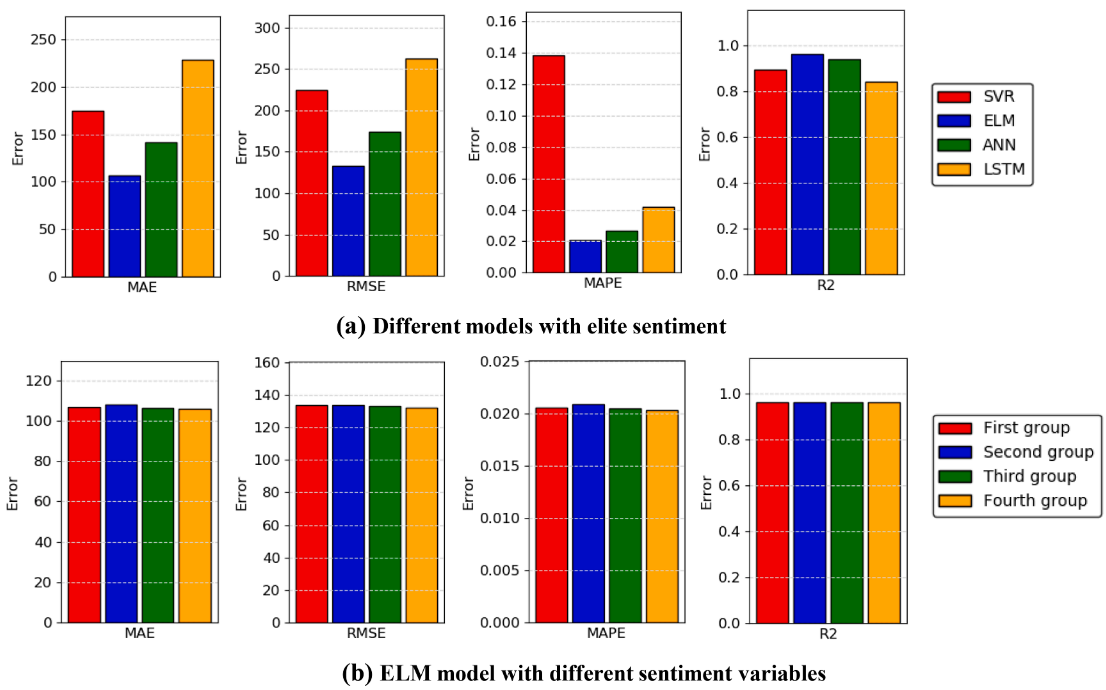

Fig. 3. Comparison results of one-day-ahead NEV stock index forecasting performance.

Table 5
High-frequency component sequence prediction performance comparisons by different input indexes.

Evaluation indicators	MAE	RMSE	MAPE	R^2
First group	84.6372	105.0778	0.7530	0.9642
Second group	83.1721	103.6884	0.7180	0.9651
Third group	80.5991	101.2150	0.7587	0.9668
Fourth group	81.4074	103.1997	0.7432	0.9655

Table 6
Low-frequency component sequence prediction performance comparisons by different input indexes.

Evaluation indicators	MAE	RMSE	MAPE	R^2
First group	65.8278	83.4490	0.1402	0.9838
Second group	31.3006	40.8716	0.0727	0.9961
Third group	17.1039	20.8674	0.0985	0.9990
Fourth group	17.8010	21.9666	0.0338	0.9989

Table 7
Trend component sequence prediction performance comparisons by different input indexes.

Evaluation indicators	MAE	RMSE	MAPE	R^2
First group	3.0365	4.7624	0.0005	0.9987
Second group	9.0394	11.7179	0.0015	0.9926
Third group	2.0237	2.5909	0.0003	0.9996
Fourth group	5.8562	6.6533	0.0010	0.9976

Table 8
Comparison of predictive performance using ELM and the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)-ELM models under elite sentiment index.

Evaluation indicators	MAE	RMSE	MAPE	R^2
ELM	99.3157	124.9894	0.0191	0.9669
CEEMDAN-ELM	83.3660	105.4968	0.0161	0.9764

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i^1 - y_i^2\right)^2}{\sum_{i=1}^{N}\left(y_i^1 - y_m^1\right)^2} \tag{9}$$

The sample size is denoted by N, with y_i^1 being the real closing price of NEV, y_i^2 being the predicted values generated by the model, and y_m^1 denoting the mean of the real values. A smaller value for the error metrics (MAE, RMSE, and MAPE) indicates better performance of the forecasting model. TheR^2 value is used to assess the fit of the predicted value to the true value.

3. Empirical results and discussion

The gray wolf optimization algorithm (Mirjalili et al., 2014) is used to select the relevant parameters of these models. For the algorithm's flowchart, refer to Appendix B.

We conduct comparison experiments with four sets of inputs for different models to evaluate the predictive performance with different input variables. The four sets of input settings are shown in Table 3.

Table 4 presents performance measures of the baseline model for different groups of input variable for MAE, RMSE, MAPE, and R^2. Among all the models that include the elite sentiment index, only the ELM model exhibits a significant performance improvement. For example, the model's performance in terms of MAE is improved by 39.03%, 24.98%, and 53.44% compared to the SVR, ANN, and LSTM models, respectively. Other evaluation metrics also showed enhancements (Fig. 3(a)). The forecasting results using the ELM model with various sentiment variables are depicted in Fig. 3(b). The pictures show that the predictive performance of ELM with the elite sentiment index is similar to ELM with the investor sentiment index, which means that the former index holds predictive potential. To further validate the results, the Wilcoxon signed-rank test was performed to evaluate the significance of the performance metrics (refer to **Appendix C**).

To verify the predictive accuracy and stability of the Chinese NEV stock price index, we constructed ELM prediction models for various time series components, such as high-frequency, low-frequency, and trend components. The prediction results are presented in Tables 5–7, revealing that the predictive impact of the elite sentiment index on the Chinese NEV stock index is significantly stronger than that of the investor sentiment index across different time scales. This may be because the elite sentiment index serves as a comprehensive reflection of various aspects, including investor, public, and user sentiments. Compared to the sole consideration of investor sentiment index as a model input, the elite sentiment index provides a more comprehensive and multidimensional view of emotional information. It effectively captures the emotions, social sentiments, and consumer attitudes of market participants, thus offering richer incremental information for investment decision-making and market forecasting.

Table 8 displays the predictive effects of the mixture and single models, revealing that the CEEMDAN-ELM model outperforms the ELM model. These findings suggest that by combining the elite sentiment index with CEEMDAN-ELM, the hybrid model can effectively reduce forecasting errors and enhance the overall prediction accuracy of the NEV stock price index.

4. Conclusions

This study investigates the impact of elite sentiment on the Chinese NEV stock index by utilizing the elite sentiment index as input for a series of machine learning models, and the findings reveal a significant improvement in prediction accuracy when incorporating the elite sentiment index into machine learning models, particularly the CEEMDAN-ELM model. This improvement can be attributed to the index's focus on the sentiment of influential users, which helps filter out noise and irrelevant information during large-scale data collection. Therefore, we consider the elite sentiment index to be a crucial factor in predicting behavioral finance. The integration of the elite sentiment index into the CEEMDAN-ELM model presents a highly promising framework for forecasting NEV stock index accurately.

Funding

This work has been supported by the Major Cultivation Project of Education Department in Sichuan Province, China (18CZ0006).

CRediT authorship contribution statement

Xingyue Gong: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Guozhu Jia:** Resources, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2023.104251.

References

Akhtar, S., Faff, R., Oliver, B., Subrahmanyam, A., 2013. Reprint of: stock salience and the asymmetric market effect of consumer sentiment news. J. Bank. Finance 37, 4488–4500. https://doi.org/10.1016/j.jbankfin.2013.07.032.

AlZaabi, S., 2021. Correlating sentiment in Reddit's Wallstreetbets with the stock market using machine learning techniques. Theses.

Box-Steffensmeier, J.M., Moses, L., 2021. Meaningful messaging: sentiment in elite social media communication with the public on the COVID-19 pandemic. Sci. Adv. 7 https://doi.org/10.1126/sciadv.abg2898 eabg2898.

Chen, W., Pourghasemi, H.R., Kornejady, A., Zhang, N., 2017. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305, 314–327. https://doi.org/10.1016/j.geoderma.2017.06.020.

Deng, S., Huang, Z.(James), Sinha, A.P., Zhao, H., 2018. The Interaction between Microblog Sentiment and Stock Return: an Empirical Examination.

Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: a review. Energy Econ. 81, 709–727. https://doi.org/10.1016/j.eneco.2019.05.006.

Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., Johnson, J.E.V., 2016. Bridging the divide in financial market forecasting: machine learners vs. financial economists. Expert Syst. Appl. 61, 215–234. https://doi.org/10.1016/j.eswa.2016.05.033.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci. 454, 903–995. https://doi.org/10.1098/rspa.1998.0193.

Leippold, M., Wang, Q., Zhou, W., 2022. Machine learning in the Chinese stock market. J. Financ. Econ. 145, 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017.

Li, X., Shang, W., Wang, S., 2019. Text-based crude oil price forecasting: a deep learning approach. Int. J. Forecast. 35, 1548–1560. https://doi.org/10.1016/j.ijforecast.2018.07.006.

Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv. Eng. Softw. 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.

Nyakurukwa, K., Seetharam, Y., 2023. The evolution of studies on social media sentiment in the stock market: insights from bibliometric analysis. Sci. Afr. 20, e01596. https://doi.org/10.1016/j.sciaf.2023.e01596.

Nyakurukwa, K., Seetharam, Y., 2022. The wisdom of the Twitter crowd in the stock market: evidence from a fragile state. Afr. Rev. Econ. Finance 14, 203–228. https://doi.org/10.10520/ejc-aref_v14_n1_a7.

Rather, A.M., Agarwal, A., Sastry, V.N., 2015. Recurrent neural network and a hybrid model for prediction of stock returns. Expert Syst. Appl. 42, 3234–3241. https://doi.org/10.1016/j.eswa.2014.12.003.

Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning : a systematic literature review: 2005–2019. Appl. Soft Comput. 90, 106181 https://doi.org/10.1016/j.asoc.2020.106181.

Smales, L.A., 2014. News sentiment in the gold futures market. J. Bank. Finance 49, 275–286. https://doi.org/10.1016/j.jbankfin.2014.09.006.
Sun, Y., Liu, X., Chen, G., Hao, Y., Zhang, Z.(Justin), 2020. How mood affects the stock market: empirical evidence from microblogs. Inf. Manage. 57, 103181 https://doi.org/10.1016/j.im.2019.103181.
Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2011 - 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Prague, Czech Republic, pp. 4144–4147. https://doi.org/10.1109/ICASSP.2011.5947265.
Wu, A.X., 2014. Ideological polarization over a China-as-superpower mind-set: an exploratory charting of belief systems among Chinese internet users, 2008–2011. Int. J. Commun. 8, 30. https://ijoc.org/index.php/ijoc/article/view/2346.
Yuan, Y., 2018. Deviation from publicity":topic preference of intellectual communities- an investigation based on SMTH forum and popular topics on NetEase (in Chinese). Contemp. Youth Res. 24–30. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2018&filename=QING201803004&uniplatform=OVERSEA&v=u56s8qjtUsjklAuzr7rK9V_8Efe-BlEUwAG9djP5TxJgHDFrGTt0bwBmjXPYchjP.
Zhang, X., Lai, K.K., Wang, S.-Y., 2008. A new approach for crude oil price analysis based on empirical mode decomposition. Energy Econ. 30, 905–918. https://doi.org/10.1016/j.eneco.2007.02.012.