A BLENDED SOFT-COMPUTING MODEL FOR STOCK-VALUE PREDICTION

N. Usha Devi and R. Mohan

ABSTRACT

Stock investments play a crucial role in deciding the global economic growth of the country. Investors can optimize profit and avoid risk through accurate stock-value prediction models, which motivates researchers to work on various aspects of correlated features and predictive models for stock-value prediction. The existing stock-value prediction models used data like Twitter, microblogs, price history and Google trends. On the other hand, domain-specific dictionary-based deep learning evolved as a competitive model for alternative models in stock value prediction. But, the accuracy of these models depends on the quality of the input, the correlation among the features and the correctness of the sentiment scores generated for the dictionary terms. Financial-news sentiment analysis for stock-value prediction with dictionary-based learning needs attention in improving the quality of the input and dictionary terms' sentiment score generation. The present research aims to develop a blended soft-computing model for stock-value prediction (BSCM) with cooperative fusion and dictionary-based deep learning. In the current work, six Indian stocks that cover uptrend, sideways and downtrend characteristics are considered with stock-price histories and news headlines from 8th August 2016 to 31st March 2023, i.e., 2427 days. The number of records in price-history dataset is 14,562 and in the news headlines dataset is 46,213. The performance of the stock-value prediction can be improved by taking advantage of multi-source information and context-aware learning. The present research aims to achieve three objectives: 1. Applying cooperative fusion to combine the news headlines and price history of stocks collected from multiple sources to improve the quality of the input with correlated features. 2. Building a dictionary, FNSentiment, with a novel strategy. 3. Predicting stock values using FNSentiment and News Sentiment Prediction Model (NSPM) integration. In the experimentation, the proposed model outperformed the state-of-the-art models with an accuracy of 91.11%, RMSE of 10.35, MAPE of 0.02 and MAE of 2.74.

KEYWORDS

1. INTRODUCTION

The stock-value prediction models gained attention in recent times. The stock value varies with the variety of features. The major categories of these features are internal and external features. Internal features include close price, the price-to-earnings ratio (P/E), the price-to-book ratio (P/B) and the like [1]. The news, Twitter, Google trends and other social media reveal external features: currency value, political decisions, organization profit, loss, the relation between employees and chief executive officer and the like, as described by [2]. The selection of the features dramatically impacts the stock-value prediction accuracy. The popular stock-value prediction models considered price history, news, Twitter [3] and microblogs as sources of stock information. But, social groups or microblogs cover a limited number of people's opinions that cannot be trusted compared to the news media reports.

In recent research, news financial sentiment analysis models evolved to predict the stock price [4]. An interesting correlation was found between the news features and the stock price. In the proposed model, to join these correlated features, cooperative fusion [5] is used. Cooperative fusion is a methodology to combine correlated features from various sources to improve the quality of the input. Through this fusion methodology, the performance of the predictive model can be improved.

Financial news sentiment analysis involves generating sentiment scores for the words in the input news. The existing English-term dictionaries were insufficient to capture the meaning of the business context statements. The Natural Language Processing (NLP) problem, polysemy [6] means that multiple interpretations are possible for the same word depending on the context. Hence, domain-specific dictionaries are required to improve the accuracy of sentiment prediction.

N. Usha Devi and R. Mohan are with NIT Tiruchirappalli, Tuvakudi, Tamil Nadu, India. Emails: 406915052@nitt.edu and rmohan@nitt.edu

Domain–specific dictionaries for stock value prediction exist for various languages. But, these dictionaries need improvement in considering correlated features of terms to generate accurate sentiment scores in dictionary learning. In the proposed system, the novel dictionary FNSentiment is developed with fused information by combining correlated features, news and close price.

Ahmad et al. [7] expressed that the accuracy of the sentiments depends on the learning model's performance. Deep-learning models exhibit great computation power in fields like image processing and text analytics. In specific, Deep Neural Networks (DNNs) [8], like Convolutional Neural networks (CNNs) [9], Long Short Term Memory (LSTM) [10] and Gated Recurrent units (GRUs) [11] are dynamic classification and predictive models. The proposed model uses the NSPM model, which combines CNN and LSTM to compute the stock value. The objectives of the proposed work are as follows.

1) Developing a cooperative fusion methodology to combine the news headlines and price history to improve the quality of the input data to the stock-value prediction model.
2) Generating domain–based dictionary, FNSentiment to with dictionary-based deep learning using the fused information.
3) Developing a stock-value prediction model combining FNSentiment and NSPM.

The remaining sections of the paper are outlined as follows: Section 2 presents the related work. Section 3 explains the theory and implementations of the proposed model. Section 4, interprets the results and discussion. Section 5 presents the conclusion and some suggestions for future work.

2. RELATED WORK

This section presents the various methods to predict stock prices and the importance of dictionary-based deep learning for stock-value prediction. Further, the section discusses the advantages and limitations and steps to overcome the boundaries of the existing models and give a clear vision of the proposed model.

2.1 Stock-value Prediction

Stock trading is a business investment technique that results in drastic variations in profit or loss in a short span with a quick change in stock value. In recent years, the following methods have evolved to address the stock-market analysis and prediction; Auto Regression (AR), Auto Regressive Integrated Moving Average (ARIMA), Auto Regressive Moving Average (ARMA) models, linear regression, Support Vector Regression (SVR), Bayes neural network, Hybrid Network Adaptive Time-series recommendation framework (HNATS) [12], Long Short Term Memory Cellular Automata (LSTMCA) [13], Fuzzy time series analysis [14], GRU [15]. Stock-value prediction is a time-series problem involving statistical or textual data analysis.

The researchers used various statistical models to address stock-value prediction based on internal features. Kumar et al. [16] have proposed an SVR with the fuzzy model and a genetic algorithm with SVR [17] to perform time-series analysis on statistical data stock's close price. Tunisian stock data analysis using a hierarchical deep neural network [18] showed a considerable performance. Even though the time-series prediction analyzes the stock's close price and other statistical factors like the P/E ratio, this analysis ignored external features that dynamically decide stock variations.

The time-series text analysis involves investigating the stock-market data collected from several media. Long et.al [19] have used the news media analysis for stock-value prediction using SVM with S&S kernel and obtained good performance. In this study, the authors expressed the importance of considering the news data for stock-value prediction. Many researchers have experimented with mass media, news, Twitter, microblogs [20], online financial comments [21], behavioural finance [22] and the like and found that the media creates hype and touches user emotions in stock trading [23]. However, in social-media analysis for stock-value prediction, all the traders must be members of a specific social network with active communication to capture data, which is impossible. In addition, the users might be irrational; hence, the correctness of the social-media data is questionable.

To overcome the existing models' limitations, we considered news data and price history for stock-value prediction in the present study. News is quickly captured and reachable to traders *via* newspapers and electronic media. Moreover, the news data is trustworthy when compared with social-media data. We

251

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 03, September 2023.

have developed a cooperative fusion method to use the price history and news headlines to obtain input datasets for stock-value prediction.

Table 1. The classification of the news headlines into positive, negative and neutral.

Stock	No. of Sentences	The polarity of news sentence		
		Positive	Negative	Neutral
WIPRO	6262	2993	2273	996
TCS	9114	6414	1580	1120
BHARTIARTL	5812	2490	1812	1510
SBIN	12855	7795	3234	1826
NXTDIGITAL	5013	2907	1488	618
PNB	7157	2980	2887	1290

2.2 Dictionary-based Sentiment Analysis

In recent research, many dictionaries have evolved to analyze text belonging to various domains. These dictionaries have been developed to be object-specific, category-specific, language-specific and the like. SentiDomain [24] was introduced as a rule-based sentiment dictionary developed for particular domain objects. This work calculates the sentiments for each object cluster using cosine similarity. This method analyzed the user review to measure product rankings and user satisfaction. Loughran and Mc- Donald manually developed a financial news dictionary, LMFinance for Hong Kong news. This dictionary outperformed the existing dictionaries, SentiWordNet and Senticnet [25].

A Korean-language dictionary [26] was developed to extract nouns from news statements and sentiments of positive and negative words obtained by calculating the average frequency of all positive and negative words in the Korean language. Most of the dictionaries were built in Japanese and Chinese rather than in English. We studied various methods to create dictionaries for deriving a domain-specific dictionary. Then, we analyzed the advantages of manual and automated dictionaries. Consequently, we proposed a semi-automated FNSentiment dictionary to take the benefits of both manual and automated dictionaries.

2.3 Sentiment Analysis with Deep Learning

The challenges in the sentiment analysis made researchers tend towards using dynamic computing models. In recent studies, deep-learning models have shown exemplary performance in sentiment analysis. Abdi et.al. [27] have introduced RNSA, which uses RNN and LSTM combinations for sentiment analysis. This model finds the sentiments of the users' emotions in social-media data. The authors found that in word-level and sentence-level features with pre-trained embedding vectors, Word2vec showed promising results.

Chen et.al. [28] have used GRU to analyze a dictionary created from Chinese social networks. This model classifies the user's emotions as positive or negative. Then, these sentiments are used to predict the stock price. Later, the authors proposed another deep-learning model, RNNboost [29], for stock prediction. The authors believed deep learning efficiently finds user emotions and is reliable for stock-value prediction. In addition, they showed that dictionary-based knowledge is suitable for analyzing domain-specific data. The hybrid CNN-LSTM dynamically classifies the statements positively and negatively [30]. In the present work, we have designed a deep-learning model, News sentiment-prediction model (NSPM) that employs CNN and LSTM with Word2Vec embedding in the proposed system.

3. PROPOSED SYSTEM

This section presents the theory and implementations of blended soft-computing model for stock-value prediction (BSCM) for the stock-value prediction for news updates using a dictionary-based deep learning approach. Figure 1 shows the BSCM architecture. The design and development of BSCM are as follows:

1) The news headlines are collected from the National Stock Exchange (NSE) and Times of India (ToI). Price history is collected from the NSE.
2) The cooperative fusion method combines price history and news headlines information from

multiple sources.
3) The FNSentiment dictionary was developed that consists of significant bigram terms with close price and sentiment scores.
4) NSPM is modeled using the deep-learning approach for stock-value prediction when integrated with FNSentiment.
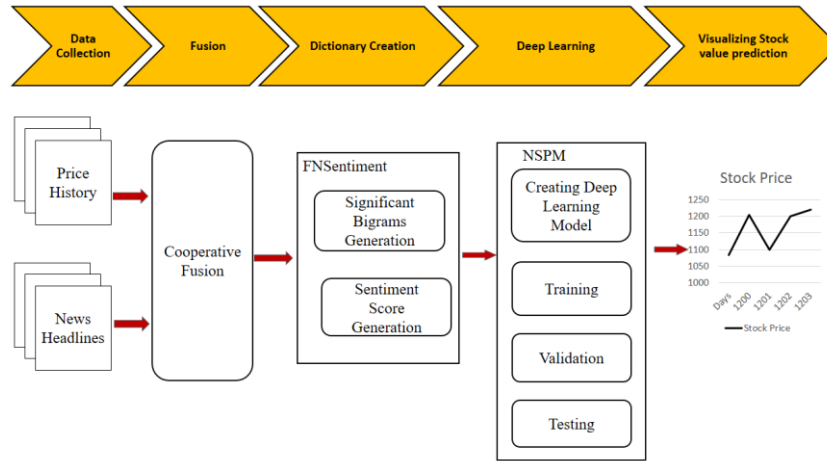

Figure 1. Architecture of the proposed model.

Figure 2. The workflow of cooperative fusion model.

Table 2. Sample terms with sentiment score in FNSentiment dictionary.

Significant bigram	Sentiment score	Close price (R s)
bad debt	0.034	490
crore loss	0.126	517
delays cheque	0.474	693
hits low	0.042	642
clearing debt	0.712	1100
more growth	0.854	1322
profit drops	0.038	786

3.1 Data Collection

The experimentation in this research considers six stocks WIPRO, TCS, BHARIARTL, SBIN, NXTDIG- ITAL and PNB to cover three possible trends in the stocks. Figure 4 illustrates the price-history input datasets with three dataset characteristics; upward, sideways and down trends. The stock datasets are of two types; one is price-history data and the other one is news headlines data from 8th August 2016 to 31st March 2023. The price-history data represents day to day transactions of trading.

253

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 03, September 2023.

The data has been collected from the NSE. The dataset contains everyday transaction details: Date, Adjacent close price, High, Low, Close price and Volume. The dataset size for each stock is 2427 for the trading days from 8th August 2016 to 31st March 2023. The total size of the price-history data is 14,562 for the six stocks. The news headlines dataset is framed by extracting data from two business news sources for the six stocks: 1. ToI, India's primary new media in English and 2. NSE press-release descriptions. The news database contains 46,213 sentences. This dataset consists of news headlines and the date of the news dissemination. Now the datasets are ready for fusion to obtain the quality inputs for stock-value prediction.
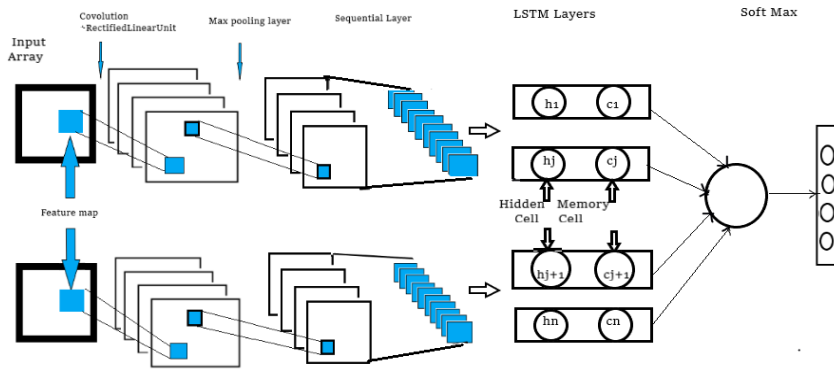

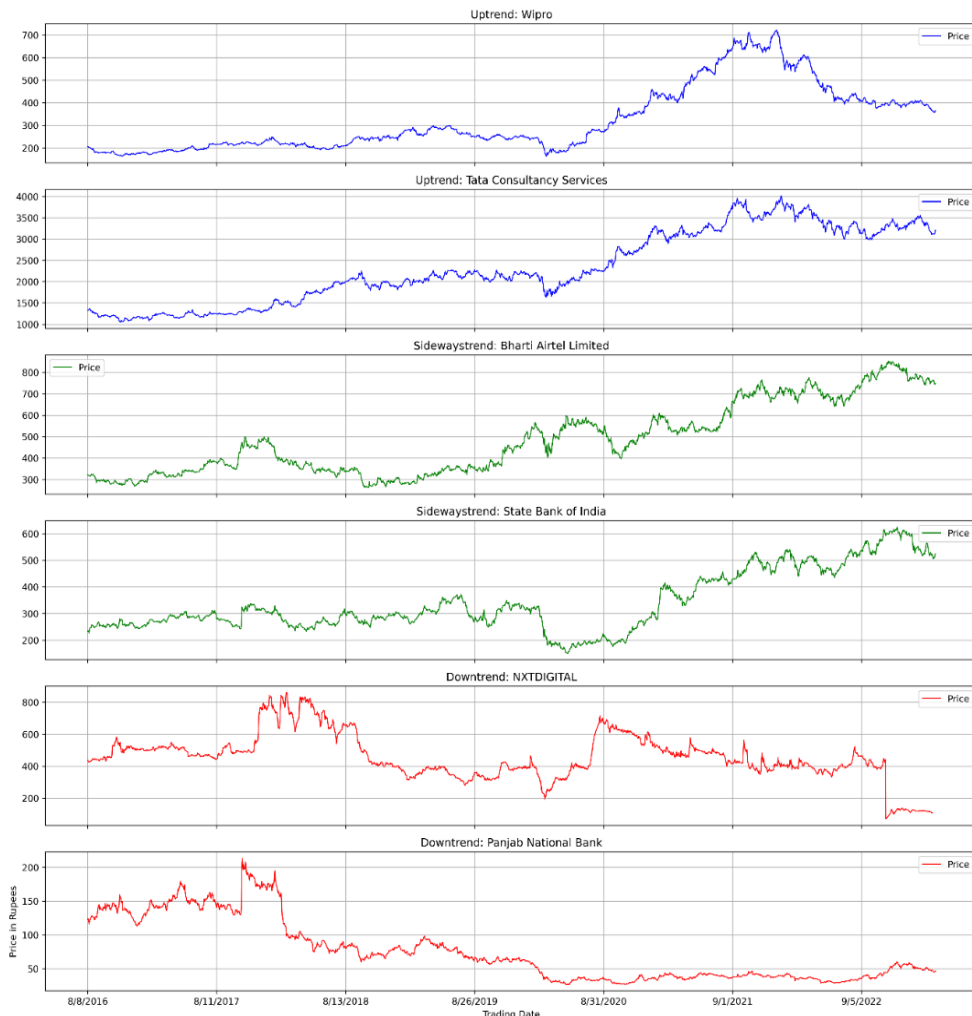Figure 3. The layered architecture of news sentiment prediction model.

Figure 4. The stock input datasets with three characteristics; uptrend, sideways and down trends.

3.2 Fusion Method

The input datasets to build the dictionary were obtained by cooperative fusion, as illustrated in Figure 2. After data collection, the pre-processing and feature-selection methods were used to find the essential elements of the price-history data. In the pre-processing step, the null value of data will be substituted by the previous day's instance value. The fusion method was applied to the datasets as follows:

1) Price-history pre-processing and feature selection.

a) Pre-processing: The stop words and the like were eliminated from the news sentences.

b) Feature engineering: For each sentence in the news headlines:

(i) A sequence of bigrams (two consecutive words) from the pre-processed news sentence is generated with a semicolon as a separator. In the next step, a polarity feature ranging from -1 to 1 will be added to each headline entry based on domain knowledge. Table 1 summarizes the domain knowledge about the news headlines.

(ii) Finally, the bigrams with the same date and stock entries are joined with semicolon as a separator and the new polarity value is considered as the sum of polarities in the respective entries.

2) In the rule-based merge step, the close-price data is mapped with news headlines data and *vice versa* based on date and stock features.

a) For the available entry in news headlines data, if the corresponding date entry is missing in the price history, a new entry is created with date and close price by considering the 1 to n steps available close price for the stock.

b) If the corresponding date entry is missing in the news headlines data, then move back to the 1 to n steps to find the most recent news for the stock.

c) For common date and stock, the stock's price-history is merged with the news headlines data.
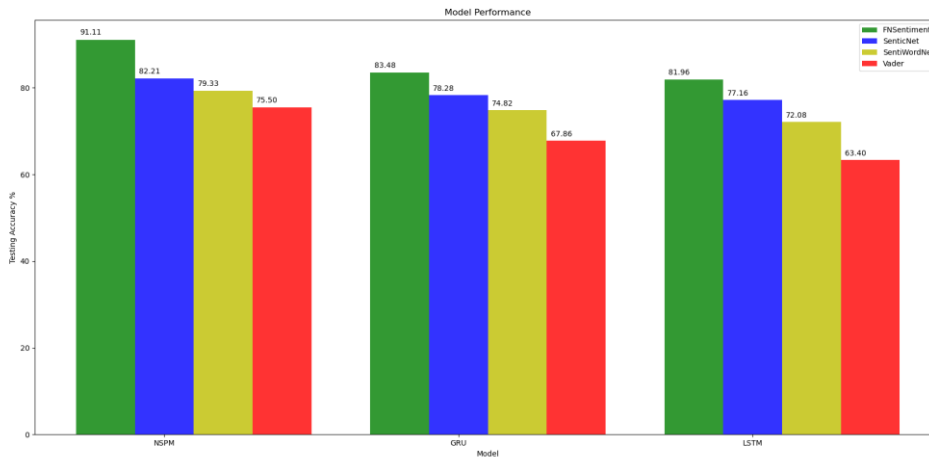
Figure 5. Performance of the news sentiment-prediction model with various standard dictionaries.

$$S_i = \frac{S}{(n-1)} \tag{1}$$

$$ST_i = \sum S_j \ for \ j = 1 \ to \ k \tag{2}$$

$$NST_i = \frac{ST_i - Min}{Max - Min} \tag{3}$$

3.3 Financial News Term Sentiment Dictionary (FNSentiment)

The fabrication of FNSentiment consists of two steps 1. Computing significant bigrams, 2. Computing bigram polarities. In the first step, significant bigram generation produces the critical terms for news sentiment analysis. The significant bigrams were obtained by the two consecutive terms generated by combining a noun, adjective, adverb and verb by discarding the other terms in each news sentence. The

255

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 03, September 2023.

polarity computation finds the polarity values for all sentences. The sentence with word length n+1 and score S generates n terms. The S gives the sentiment score for each significant term in the sentence. A sentiment score S_i for i=1 to n was computed using (1), contributing to the sentiment score; hence, the output was a set of terms with sentiment for each sentence. Next, the sentence's sentiment denoted by STi was computed using (2). After this step, The FNSentiment was updated with the pairs of terms and their scores. Next, the Normalized Sentiment Score (NST) was computed using (3) to convert the sentiment scores on the scale (0, 1). Thus, NSTi generates polarity values ranging from (0, 1). The NSTi value zero represents the highly negative sentiment term. The NSTi value one defines the extreme positive sentiment term and 0.5 means the terms with a neutral sentiment. In (3), min and max represent ST's minimum and maximum values, respectively. Table 3 shows an instance of the FNSentiment dictionary.

3.4 Stock-value Prediction Using NSPM

The objective of the NSPM is to predict the news sentiments using deep learning, as shown in Figure 3. The model consists of three layers; embedding, CNN and LSTM. First, the embedding layer creates input vectors to train the model. In this step, the layer generates equivalent word vectors for the inputs. In the next step, the embedded vectors were passed from the convolution layer to the max pooling layer to capture the most significant features from the information. Finally, the input was passed through the sequential layers of LSTM for further learning. The ci indicates an internal cell that collects input in LSTM. The hj represents the hidden state that produces output. The final layer softmax outputs a value from zero to one, indicating the sentence sentiment value. The hyper-parameters of NSPM were found through the Bayesian optimization tuner of the Keras tuner.

Table 3. An instance of fused information obtained from the news and price-history datasets.

Date	Stock	Close price (R s)	Significant bigrams	Sentiment score
17-Apr-21	SBIN	339.9	statebankofindia private; private bank; bank seen; seen race; race card; card business	1
18-Apr-21	SBIN	339.9	statebankofindia pharmacist; pharmacist recruitment apply; apply online	0
20-Apr-21	SBIN	329.5	statebankofindia say; say charge; charge zero; zero balance; balance account; account prior; prior reasonable	1
24-Apr-21	SBIN	336.45	statebankofindia cut; cut growth; growth vijay; vijay mallya; mallya say; say money; money owes; owes indian; indian bank; bank public; public money; money cannot; cannot made; made bankrupt	-1
25-Apr-21	SBIN	336.45	bihar suffer; suffer crore; crore financial; financial loss; loss covid	-1
27-Apr-21	SBIN	353.05	statebankofindia clerk; clerk exam; exam registration; registration begin	0

Table 4. Significant bigram and sentiment-score generation for the sample fused information to develop FNSentiment.

Significant bigrams	Sentiment score	Normalized sentiment score	Close price
statebankofindia private	0.17	1	339.9
private bank	0.17	1	339.9
bank seen	0.17	1	339.9
seen race	0.17	1	339.9
race card	0.17	1	339.9
card business	0.17	1	339.9

statebankofindia pharmacist	0	0.54	339.9
pharmacist recruitment	0	0.54	339.9
recruitment apply	0	0.54	339.9
apply online	0	0.54	339.9
bihar suffer	-0.2	0	336.4
suffer crore	-0.2	0	336.4
crore financial	-0.2	0	336.4
financial loss	-0.2	0	336.4
loss covid	-0.2	0	336.4

Table 5. Performance comparison of news sentiment-prediction model with baseline models.

Model	Phase	Accuracy (%)						Avg.
		WIPRO	TCS	BHARTI	SBIN	NXT	PNB	
NSPM with FNSentiment	Testing	91.33	92.51	90.35	91.06	90.14	91.25	91.11
	Training	92.32	93.63	91.08	93.14	91.72	92.19	92.35
	Validation	89.45	89.87	88.66	88.91	88.32	89.52	89.12
NSPM	Testing	86.11	84.34	83.47	85.17	83.79	80.14	83.84
	Training	88.99	86.44	88.67	87.91	86.90	82.39	86.88
	Validation	85.59	83.44	82.15	84.10	82.84	79.49	82.94
LSTM With FNSentiment	Testing	86.62	85.14	85.38	82.43	81.06	80.24	83.48
	Training	87.99	87.74	86.55	83.93	82.41	81.32	84.99
	Validation	84.10	83.56	83.15	80.21	80.41	79.49	81.82
LSTM	Testing	82.12	81.74	80.11	78.26	78.79	78.18	79.87
	Training	85.16	83.45	81.16	81.06	81.16	80.12	82.02
	Validation	80.28	80.14	79.36	75.13	77.54	75.52	78.00
GRU With FNSentiment	Testing	85.12	81.24	84.18	80.17	81.79	79.24	81.96
	Training	86.56	84.71	85.27	82.06	82.16	80.59	83.56
	Validation	85.59	83.44	82.15	84.10	82.84	79.49	82.94
GRU	Testing	81.84	81.24	83.28	78.18	72.93	71.33	78.13
	Training	84.27	83.82	84.17	80.04	76.62	74.59	80.59
	Validation	79.39	80.29	80.11	75.86	69.93	70.84	76.07
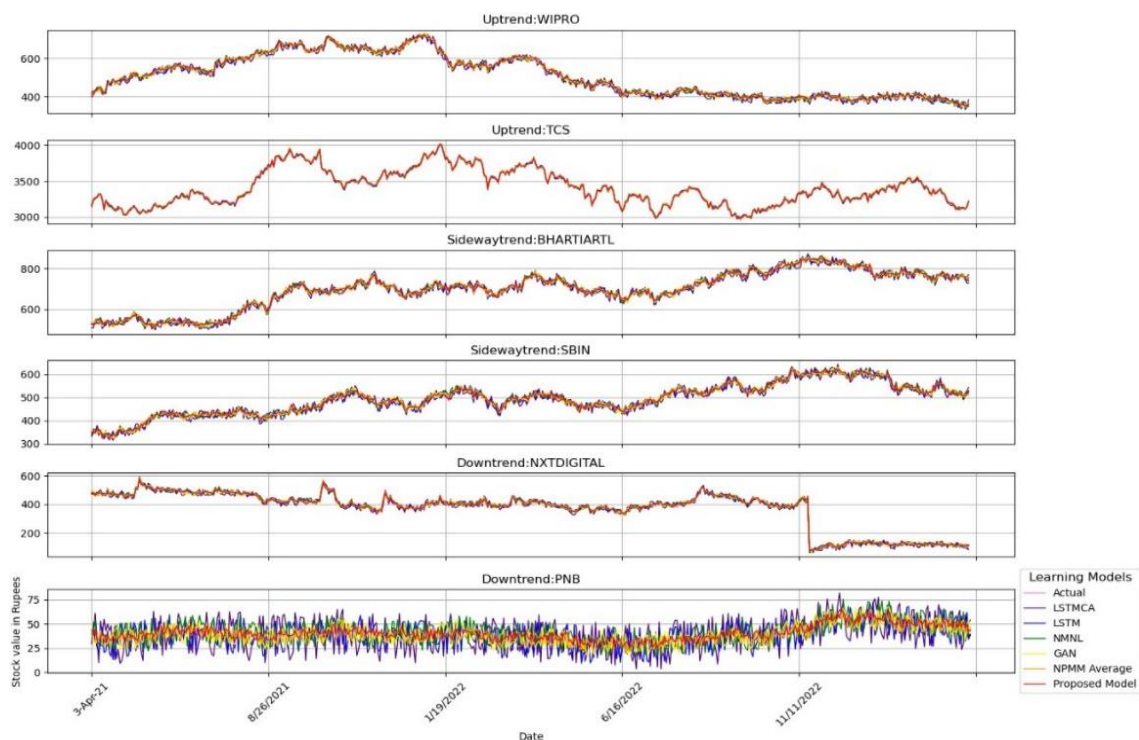
Figure 6. Stock-value prediction using the proposed model and existing models.

257

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 03, September 2023.

4. RESULTS AND DISCUSSION

The dataset is divided into 70% for training data from 8th Aug 2016 to 3rd March 2021 (1699 days) and 30% testing data from 4th March 2021 to 31st March 2023 (728 days). The results, performance analysis of the proposed model, evaluation and comparison with the models in recent literature are explained in the following sub-sections.

Table 6. Comparison of proposed and baseline models using the RMSE metric for each input stock dataset.

Model/ Stock	WIPRO	TCS	BHARTIARTL	SBIN	NXT DIGITA L	PNB
Proposed Model	9.53	10.00	10.88	10.86	10.24	10.58
NPMM Average[31]	24.36	24.26	27.84	24.74	29.44	26.02
GAN [32]	51.23	51.06	47.11	50.34	49.04	50.72
NMNL[33]	76.70	83.83	84.46	86.73	90.42	89.07
LSTM [11]	125.24	127.87	110.64	127.58	125.27	123.01
LSTMCA [13]	200.36	209.52	219.52	218.89	224.35	221.20

Table 7. Comparison of proposed and baseline models using the MAE metric for each input stock dataset.

Model/ Stock	WIPRO	TCS	BHARTIART L	SBIN	N XT DIGITAL	PNB
Proposed Model	2.54	2.70	2.84	2.80	2.76	2.78
NPMM Average[31]	4.25	4.22	4.56	4.32	4.66	4.34
GAN [32]	6.22	6.12	5.92	6.09	6.10	6.19
NMNL[33]	7.51	8.03	8.00	8.05	8.32	8.27
LSTM [11]	9.58	9.73	8.85	9.74	9.73	9.62
LSTMCA [13]	12.05	12.53	12.78	12.86	12.87	12.89

Table 8. Comparison of proposed and baseline models using the MAPE metric for each input stock dataset.

Model/ Stock	WIPRO	TCS	BHARTIARTL	SBIN	N XT DIGITAL	PNB
Proposed Model	0.005	0.001	0.004	0.006	0.010	0.071
NPMM Average[31]	0.009	0.001	0.007	0.009	0.016	0.110
GAN [32]	0.013	0.002	0.009	0.013	0.022	0.159
NMNL[33]	0.016	0.002	0.012	0.017	0.029	0.210
LSTM [11]	0.020	0.003	0.013	0.020	0.034	0.249
LSTMCA [13]	0.025	0.004	0.019	0.027	0.044	0.331

Table 9. Comparison of proposed and baseline models using Accuracy, RMSE, MAE and MAPE.

Model/ Metric	Accuracy (%)	RMSE	MAE	MAPE
Proposed Model	91.11	10.35	2.74	0.02
NPMM Average[31]	86.57	26.11	4.39	0.03
GAN [32]	84.4	49.92	6.11	0.04
NMNL[33]	81.89	85.20	8.03	0.05
LSTM [11]	80.77	123.27	9.54	0.06
LSTMCA [13]	78.16	215.64	12.66	0.07

The experimentation of the present research is described as follows. The collected experimental data, news headlines and close prices, was initially joined using cooperative fusion to generate quality input. In the next step, the FNSentiment dictionary was built by computing significant bigrams and their polarities. As a final step, NSPM is used to predict the stock value with the FNSentiment using dictionary–based deep learning.

4.1 Metrics for Evaluation

The metrics used for model evaluation are Accuracy, RMSE, MAE and MAPE. The accuracy determines the percentage of the number of sentences recognized correctly among the total tested

sentences. The RMSE gives the square root of averaged squared error. The error represents the difference between the actual and predicted values. The MAE gives the absolute difference between the predicted and actual values. MAPE is the mean absolute percentage error that determines the relative error. The proposed work aims to optimize these metrics.

$$Accuracy = \frac{Number\ of\ correctly\ categorized\ news\ sentences}{Total\ number\ of\ sentences} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Actual_i - Predicted_i)}{N}} \tag{5}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(Actual_i - Predicted_i) \tag{6}$$

$$MAPE = \sum_{i=1}^{N}\left|\frac{(Actual_i - Predicted_i)}{Actual_i}\right| \tag{7}$$

4.2 Results of Cooperative Fusion

The news and price-history datasets were combined to create the fused information. In this process, feature selection is applied to price-history datasets to select the prominent features 'close price' and 'Date'. Feature engineering is applied to pre-processed news data sentences to obtain bigrams. Further, a sentiment score between -1 and 1 is appended to each news sentence based on the domain knowledge. In the next step, the close price and bigrams of specific stock were mapped with Date. This step results in datasets with (Date, Stock, Close price, Bigrams, Sentiment score). Table 3 shows information fusion for the sample input dataset.

4.3 FNSentiment Dictionary

The formulation of the FNSentiment dictionary starts by collecting the fused information from the previous step. The next step generates significant bigrams by considering the noun, adjective, adverb and verb combinations of the bigrams obtained from the fused information. In the next step, formulae (1) and (2) are applied to compute the sentiment score for each bigram. Then, normalized sentiment score is calculated for the sentiment score field to convert the range of the values from 0 to 1. For the duplicate terms in the dictionary, sentiment scores were summed up and the close prices were averaged. Now, the FNSentiment contains triplets (significant bigrams, sentiment score, close price). An instance of these results is shown in Table 4.

4.4 Performance of News-sentiment Prediction Model (NSPM)

The NSPM and alternate deep-learning models are integrated with various dictionaries and compared for the analysis. The hyper-parameters of NSPM are done through Bayesian optimization. The hyper-parameters are as follows: The learning rate for the generator and discriminator is 0.01, the suitable optimizer is Adam. We considered the number of epochs as 100 throughout the experimentation. The NSPM with FNSentiment was evaluated and compared with standard dictionaries, SenticNet, SentiWordNet and Vader. Figure 5 illustrates the comparison results. These results demonstrate that context-aware learning is possible through the integration of FNSentiment and NSPM with an accuracy of 91.11%.

Table 5 shows the experiment summary on the six stocks. The NSPM with FNSentiment is a promising approach compared with NSPM alone, with improved accuracy by 3.33% from the experimental results. Figure 5 illustrates the accuracy of NSPM integrating with the FNSentiment and existing dictionaries. The FNSentiment with NSPM shows an accuracy of 91.11%; thus the results concluded that the NSPM with FNSentiment outperformed the recent literature models.

4.5 Performance Analysis of Existing and Proposed Models

BSCM is evaluated and compared with the baseline models with the metrics: Accuracy, RMSE, MAE and MAPE. The metrics were computed using formulae (5), (6) and (7). Tables 6, 7 and 8 illustrate the model evaluation results using the metrics. The summary of the results is shown in Table 9. The BSCM model outperformed all the baseline models with an accuracy of 91.11%. The evaluation of other metrics, RMSE of 10.35, MAPE of 0.002 and MAE of 2.74, showed that the BSCM is a reliable stock-

value prediction system. Figure 6 shows the stock values predicted by the proposed model BSCM and the existing models for the six stocks: WIPRO, TCS, BHARIARTL, SBIN, NXTDIGITAL and PNB.

5. CONCLUSION AND FUTURE WORK

In the present work, we computed the futuristic stock values for six stock datasets with the proposed model and the models in the literature. The proposed model achieved an accuracy of 91.11%, RMSE of 10.35, MAPE of 0.002 and MAE of 2.74. BSCM improved stock-value prediction with a rise in accuracy by 4.54% and with a fall of the MAE by 1.65, MAPE by 0.01 and RMSE by 15.76 compared with the existing models. BSCM outperformed the models in the literature. The NSPM for news sentiment prediction improved accuracy by 3.72% after integrating with a novel dictionary FNSentiment. The results showed that the cooperative fusion method and dictionary-based deep learning models improved the stock-value prediction accuracy. In future studies, we want to incorporate context-based clustering to refine the significant bigrams in predicting the stock value. The BSCM can be enhanced by integrating with clustering to analyze the critical news features for various stocks like oil, bank, software stocks and the like to develop sector-wise dictionaries to optimize the runtime.

REFERENCES

[1] S. Mittal and C. Nagpal, "Predicting a Reliable Stock for Mid and Long Term Investment," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 10, pp. 8440–8448, 2022.

[2] Q. Li, Y. Chen, J. Wang et al., "Web Media and Stock Markets: A Survey and Future Directions from a Big Data Perspective," IEEE Trans. Knowl. Data Eng., vol. 30, no. 2, pp. 381–399, 2018.

[3] Z. Berradi, M. Lazaar, O. Mahboub, H. Berradi and H. Omara, "Combination of Deep-learning Models to Forecast Stock Price of AAPL and TSLA," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 8, no. 4, pp. 345 – 356, 2022.

[4] S. García-Méndez et al., "Automatic Detection of Relevant Information, Predictions and Forecasts in Financial News through Topic Modeling with Latent Dirichlet Allocation," Applied Intelligence, vol. 2023, pp. 1–19, 2023.

[5] R. A. Mokhtar, R. A. Saeed and H. Alhumyani, "Cooperative Fusion Architecture-based Distributed Spectrum Sensing under Rayleigh Fading Channel," Wireless Personal Communications, vol. 124, no. 1, pp. 839–865, 2022.

[6] A. Moldovan, "Descriptions and Tests for Polysemy," Axiomathes, vol. 31, pp. 229–249, 2021.

[7] S. Ahmad, S. Mishra, F. J. Zareen and S. Jabin, "Sensor-enabled Biometric Signature-based Authentication Method for Smartphone Users," Int. J. of Biometrics, vol. 15, no. 2, pp. 212–232, 2023.

[8] P. Mohanty, J. P. Sahoo and A. K. Nayak, "Application of Deep Learning Approach for Recognition of Voiced Odia Digits," Int. J. of Computational Science and Engineering, vol. 25, no. 5, pp. 513–522, 2022.

[9] Y. Guo, S. Han, C. Shen, Y. Li, X. Yin and Y. Bai, "An Adaptive SVR for High-frequency Stock Price Forecasting," IEEE Access, vol. 6, pp. 11397–11404, 2018.

[10] W. Fang, T. Jiang, K. Jiang et al., "A Method of Automatic Text Summarization Based on Long Short-term Memory," Int. J. of Computational Science and Engineering, vol. 22, no. 1, pp. 39–49, 2020.

[11] S. Yang, "A Novel Study on Deep Learning Framework to Predict and Analyze the Financial Time Series Information," Future Generation Computer Systems, vol. 125, pp. 812–819, 2021.

[12] W. Wang, Y. Shi and R. Luo, "Sparse Representation-based Approach to Prediction for Economic Time Series," IEEE Access, vol. 7, pp. 20614–20618, 2019.

[13] N. Usha Devi and R. Mohan, "Long Short-term Memory with Cellular Automata (LSTMCA) for Stock Value Prediction," Proc. of 3rd Int. Conf. ICDECT-2K19, pp. 841–848, DOI: 10.1007/978-981-15-1097-7_70, 2020.

[14] Y. Wang and L. Han, "Adaptive Time Series Prediction and Recommendation," Information Processing & Management, vol. 58, no. 3, p. 102494, 2021.

[15] J. Hu, Q. Chang and S. Yan, "A GRU-based Hybrid Global Stock Price Index Forecasting Model with Group Decision-making," Int. J. of Computational Science and Eng., vol. 26, no. 1, pp. 12–19, 2023.

[16] G. Kumar, S. Jain and U. P. Singh, "Stock Market Forecasting Using Computational Intelligence: A Survey," Archives of Computational Methods in Eng., vol. 28, pp. 1069-1101, 2020.

[17] F. Iqbal, J. M. Hashmi, B. C. M. Fung et al., "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," IEEE Access, vol. 7, pp. 14 637–14 652, 2019.

[18] O. Lachiheb and M. S. Gouider, "A Hierarchical Deep Neural Network Design for Stock Returns Prediction," Procedia Computer Science, vol. 126, pp. 264–272, 2018.

[19] W. Long, L. Song and Y. Tian, "A New Graphic Kernel Method of Stock Price Trend Prediction Based on Financial News Semantic and Structural Similarity," Expert Systems with Applications, vol. 118, pp. 411–424, 2019.

"A Blended Soft-computing Model for Stock-value Prediction", N. Usha Devi and R. Mohan.

[20] Y. Chang, S. Li, A. Liu and J. Jin, "Quality Assessment of Screen Content Images Based on Multi-stage Dictionary Learning," J. of Visual Comm. and Image Representation, vol. 79, no. July, p. 103248, 2021.

[21] Y. Qian, Z. Li and H. Yuan, "On Exploring the Impact of Users' Bullish-bearish Tendencies in Online Community on the Stock Market," Infor. Processing & Management, vol. 57, no. 5, p. 102209, 2020.

[22] Y. Ge et al., "Beyond Negative and Positive: Exploring the Effects of Emotions in Social Media during the Stock Market Crash," Information Processing & Management, vol. 57, no. 4, p. 102218, 2020.

[23] D. Cabrera, C. Cubillos, A. Cubillos et al., "Affective Algorithm for Controlling Emotional Fluctuation of Artificial Investors in Stock Markets," IEEE Access, vol. 6, pp. 7610–7624, 2018.

[24] M. Ahmed, Q. Chen and Z. Li, "Constructing Domain-dependent Sentiment Dictionary for Sentiment Analysis," Neural Computing and Applications, vol. 32, pp. 14719–14732, 2020.

[25] X. Li, P. Wu and W. Wang, "Incorporating Stock Prices and News Sentiments for Stock Market Prediction: A Case of Hong Kong," Infor. Processing & Management, vol. 57, no. 5, p. 102212, 2020.

[26] J. Kim et al., "Stock Price Prediction through the Sentimental Analysis of News Articles," Proc. of the IEEE 11th Int. Conf. on Ubiquitous and Future Networks (ICUFN), pp. 700–702, Zagreb, Croatia, 2019.

[27] A. Abdi, M. Rahmati and M. M. Ebadzadeh, "Entropy Based Dictionary Learning for Image Classification," Pattern Recognition, vol. 110, p. 107634, 2021.

[28] W. Chen, Y. Zhang et al., "Stock Market Prediction Using Neural Network through News on Online Social Networks," Proc. of the 2017 IEEE Int. Smart Cities Conf. (ISC2), pp. 1–6, Wuxi, China, 2017.

[29] W. Chen, C. K. Yeo, C. T. Lau and B. S. Lee, "Leveraging Social Media News to Predict Stock Index Movement Using RNN-boost," Data & Knowledge Engineering, vol. 118, pp. 14–24, Nov. 2018.

[30] P. K. Jain, V. Saravanan and R. Pamula, "A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-generated Contents," Transactions on Asian and Low-resource Language Information Processing, vol. 20, no. 5, pp. 1–15, 2021.

[31] Y. Han, J. Kim and D. Enke, "A Machine Learning Trading System for the Stock Market Based on N-period Min-max Labeling Using XGBoost," Expert Systems with Applications, vol. 211, p. 118581, 2023.

[32] A. Kumar, A. Alsadoon, P. Prasad et al., "Generative Adversarial Network (GAN) and Enhanced Root Mean Square Error (ERMSE): Deep Learning for Stock Price Movement Prediction," Multimedia Tools and Applications, vol. 81, pp. 3995–4013, 2022.

[33] X. Chen, X. Ma, H. Wang, X. Li and C. Zhang, "A Hierarchical Attention Network for Stock Prediction Based on Attentive Multi-view News Learning," Neurocomputing, vol. 504, pp. 1–15, 2022.

ملخص البحث:

تلعــب الاســتثمارات فــي الأســهم دوْراً حاســماً فــي النّمـوّ الاقتصــادي للــدّولِ ويمكــن للمســتثمرين تعظيــم أربــاحهم وتجنِّـب المخــاطرة لــدى اســتخدامهم نمــاذج توقُّـع قيــم الأســهم، الأمــر الــذي حــدا بالبــاحثين للعمــل علــى تطويــر نمــاذج لتوقُّـع أسعــار الأسهــم. وقــد ظهــر الــتّعلُّم العميــق المرتكــز علــى اســتخدام القــواميس الخاصّــة بمجــالات معينــة كتقنيــة منافســة بديلــة لتوقّــع أســعار الأسهــم فــي الأســواق الماليــة. إلّا أنّ دِقّــة النّمــاذج المســتخدمة لهــذا الغــرض تعتمــد علــى جــودة المــدخلات، والارتبــاط بــين السِّــمات، وصحّــة المشــاعر المتولِّــدة مــن المصــطلحات القاموسية. مــن هُنــا فــإنّ تحليــل البيانــات الماليــة مــن أجْـل توقُّـع أســعار الأســهم باســتخدام الــتّعلُّم القائــم علــى القــواميس يجــب أن يُركِّــز علــى تحســين جــودة المدخلات وتوليد نتائج عالية بالمشاعر المتضمَّنة في المصطلحات القاموسية.

يهــدف هــذا البحــث الــى تطويــر نمــوذج مخــتلط للحوســبة النّاعمــة لتوقُّــع قيمــة الأســهم باســتخدام مــا يُعــرف بالانــدماج المقــارن والــتّعلُّم العميــق المســتند الــى القــواميس. ويــدرس النّمــوذج المقتــرح (6) أســواق للأســهم مــن حيــث تــواريخ أســعار الأســهم وعنــاوين الأخبــار فــي تلــك الأســواق. والجــدير بالــذّكر أنّ الأداء المتعلِّــق بتوقُّـع أســعار الأســهم فــي النّمــاذج المســتخدمة يمكــن تحســينُهُ بالاســتفادة مــن مزايــا المعلومــات متعــدّدة المصــادر والــتّعلُّم الــواعي للسّــياق. ولــدى تجريــب النّمــوذج المقتــرح، تفــوّق علــى مثيلاتــه مــن التّقنيــات الــواردة فــي أدبِيّــات الموضــوع؛ فقــد حقَّــق دِقّــة بلغــت 91.11 % الــى جانــب تمُيُّــزه فــي مقــاييس التقيــيم الأخــرى (RMSE=10.35؛ MAPE=0.02؛ MAE=2.74).