



News Text Mining-Based Business Sentiment Analysis and Its Significance in Economy

Ming Yang, Binghan Jiang*, Yimin Wang, Tianyu Hao and Yuankun Liu

Faculty of Business and Economics, The University of Hong Kong, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:

Baiyuan Ding,
National University of Defense
Technology, China

Reviewed by:

Ling Shen,
Jiangsu Vocational College of Finance
and Economics, China
Xianhui Gu,
Henan Agricultural University, China

*Correspondence:

Binghan Jiang
u3587762@connect.hku.hk

Specialty section:

This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: 12 April 2022

Accepted: 17 May 2022

Published: 14 July 2022

Citation:

Yang M, Jiang B, Wang Y, Hao T
and Liu Y (2022) News Text
Mining-Based Business Sentiment
Analysis and Its Significance
in Economy.
Front. Psychol. 13:918447.
doi: 10.3389/fpsyg.2022.918447

The purpose of business sentiment analysis is to determine the emotions or attitudes expressed toward the company, products, services, personnel, or events. Text analysis are the simplest and most developed types of sentiment analysis so far. The text-based business sentiment analysis still has some unresolved challenges. For example, the machine learning algorithms are unable to recognize double meanings, jokes and allusions. The regional differences between language and non-native speech structures cannot be explained. To solve this problem, an undirected weighted graph is constructed for news topics. The sentences in an article are modeled as nodes, and the normalized sentence similarity is used as the link of the nodes, which can help avoid the influence of sentence length on the summary results. In the topic extraction process, the keywords are not limited to a single word, to achieve the purpose of improving the readability of the abstract. To improve the accuracy of sentiment classification, this work proposes a robust news mining-based business sentiment analysis framework, called BuSeD. It contains two main stages: (1) news collection and preprocessing, and (2) feature extraction and sentiment classification. In the first stage, the news is collected by using crawler tools. The news dataset is then preprocessed by reducing noises. In the second stage, topics in each article is extracted by using traditional topic extraction tools. And then a convolutional neural network (CNN)-based text analyzing model is designed to analyze news from sentence level. We conduct comprehensive experiments to evaluate the performance of BuSeD for sentiment classification. Compared with four classical classification algorithms, the proposed CNN-based classification model of BuSeD achieves the highest F1 scores. We also present a quantitative trading application based on sentiment analysis to validate BuSeD, which indicates that the news-based business sentiment analysis has high economic application value.

Keywords: business sentiment analysis, text mining, convolutional neural network, economy, artificial intelligence

INTRODUCTION

Business sentiment analysis is to capture and track the opinions, emotions or feelings expressed by consumers when they are participating in various business interactions (Sun et al., 2020). Sentiment analysis involves the analysis and processing of text, natural language, computational linguistics, or biometrics. The purpose is to determine the emotions or attitudes expressed toward the company, products, services, personnel, or events. With the rapid development of the Internet,

the boundary between information providers and users has become more and more blurred. Under the background of the Internet financial era, the sources and channels of financial information are becoming more and more rich and diverse. Investors have more channels to obtain relevant financial information. In the past century, financial research is often limited to the research of data in digital form, while ignoring the research of unstructured financial “data”. In the current Internet era, analysts’ research reports, stock bar forum posts the news of news media and unstructured text information such as microblog and Facebook can often reflect the investment sentiment of investors on the stock market. This information often plays a subtle role in investors’ investment decisions (Provoost et al., 2019).

Text-based sentiment analysis is a hot research topic in both business and artificial intelligence (Alantari et al., 2022). Text analysis are the simplest and most developed types of sentiment analysis so far. They are in great demand and have a long history of development. They are also the technology most used by enterprises and the public sector. Text-based sentiment analysis, especially for business purposes, can be summarized as dividing sentences, paragraphs, posts or documents into negative, neutral or positive categories. Among them, more complex emotion and attitude processing, meaning extraction, intention classification and emotion analysis based on linguistics are becoming more and more popular. Automatic emotion analysis is usually realized through supervised machine learning, dictionary based unsupervised mining or their combination (Veltmeijer et al., 2021). Crawling to popular public websites to extract news data is an efficient way of collecting data for business sentiment analysis.

A common concept of machine learning in sentiment analysis is that the success of sentiment artificial intelligence “training” always depends on the quality of the input data. Larger, better, and cleaner data sets are necessary to avoid the “garbage in, garbage out” situation (Sun et al., 2022). However, even we have high quality dataset, the text-based business sentiment analysis still has some unresolved challenges. For example, the machine learning algorithms are unable to recognize double meanings, jokes, and allusions. The regional differences between language and non-native speech structures cannot be explained. For sentiment intelligence, dealing with irony in written speech can be a difficult task, and there may be a distorted understanding of meaning and intention (Kim et al., 2021). Although social media is usually the source of opinions and intentions mined by machine learning algorithms, the language is undeniably specific and not necessarily real speech in real life (Zhang et al., 2022).

To improve the accuracy of business sentiment prediction, this work proposed a robust news mining-based business sentiment analysis framework, called BuSeD. It contains two main stages: (1) news collection and preprocessing, and (2) feature extraction and sentiment classification. In the first stage, the news is collected by using crawler tools. The news dataset is then preprocessed by reducing noises. In the second stage, topics in each article is extracted by using a classical topic extraction tool. An undirected weighted graph is constructed for news

topics. The sentences in an article are modeled as nodes, and the normalized sentence similarity is used as the link of the nodes, which can help avoid the influence of sentence length on the summary results. In the topic extraction process, the keywords are not limited to a single word, to achieve the purpose of improving the readability of the abstract. And then a convolutional neural network (CNN)-based text analyzing model is designed to analyze news from sentence level. We conduct comprehensive experiments to evaluate the performance of BuSeD for sentiment prediction. Compared with state-of-the-art classification algorithms, the proposed CNN-based prediction model of BuSeD achieves the highest F1 scores. We also present a case study to validate the economic application of business sentiment analysis, which indicates that the news-based business sentiment analysis has high economic application value. The main contribution of this paper is as follows:

(1) We propose a CNN-based robust news mining-based business sentiment analysis framework. It can enhance the explainability of the regional differences between language and non-native speech structures by using the undirected weighted graph to model topics in sentences.

(2) We propose a CNN-based text sentiment analyzing model to improve the sentiment classification accuracy.

(3) We conduct comprehensive experiments to prove the efficiency of the proposed BuSeD.

The structure of the rest paper is as follows: Section 2 reviews the state-of-the-art works in text mining and business sentiment analysis. Section 3 presents the details of the proposed BuSeD framework. Section 4 introduces the experiment process and analysis the experiment results. Section 5 concludes this work and gives future research directions.

RELATED WORK

Using web crawling techniques and text mining methods, work (Pan et al., 2020) explored the impacts and mitigation measures of COVID-19 on China’s agricultural economy in a study of over 750,000 words on two media channels, WeChat and Sina Weibo, on COVID-19 and agriculture in China. Jo et al. (Jo and You, 2019) used text mining methods, relevant newspapers, and government documents to interpret South Korea’s announcement of a new measure. The purpose is to expand the National Health Insurance (NHI) coverage strategies and to explore the conflicts between them and the more active role the government should play in policy development. Work (Pejić Bach et al., 2019) examines what techniques are used in text mining, what is at the heart of the field and what data sources are commonly used for text mining in the financial sector’s operations, particularly in the age of the internet, big data and social media. The aim of work (Lee and Hong, 2020) is to explore trends in blockchain technology through text mining analysis of patents and news articles, and to propose a blockchain policy agenda by capturing societal interests.

By using a case study research approach, work (Hu and Wang, 2019) analyses its synergistic approach to teaching and learning through a journalism translation module, while

suggesting that through the integration of academic and professional communities, trainers of translators may have their own reflections on innovative teaching and learning strategies. Work (Kovalev, 2021) explores the dual influence of market and political pressures on journalists and the resulting characteristics of censorship and self-censorship in Russia, through which the conditions of the Russian media market are explored. The purpose of work (Schmeller, 2022) is to share a content analysis of the usability of a business news article for teaching business model concepts in a capstone business course. They suggested that professors must use innovative, empirical approaches to improve learning.

In work (Souma et al., 2019), deep learning approach using recurrent neural networks combined with long-short term memory units was used to train Thompson Reuters news archive data from 2003 to 2012 and to test predictive power on news archive data from 2013. In work (Mohan et al., 2019), the accuracy of stock price prediction is improved by collecting a large amount of time series data and analyzing it using a deep learning model in conjunction with relevant news articles. In work (Alonso et al., 2021), the different uses of sentiment analysis in fake news detection are examined, the most relevant elements and drawbacks are discussed, as well as the requirements which should be met soon. Work (Štrimaitis et al., 2021) uses three supervised machine learning models for sentiment analysis of financial background news collected from Lithuanian language websites. Hyperparameter optimization is also performed using grid search to obtain the optimal hyperparameters for each classifier.

The above-mentioned methods do not solve the problem that it is not easy to explain the regional differences between language and non-native speech structures. In addition, the sentiment classification accuracy should be further improved. Therefore, in this paper, we propose a new method to enhance the regional differences between language and non-native speech structures, and to improve the classification accuracy.

BUSINESS SENTIMENT ANALYSIS BY ANALYZING NEWS

We propose a news mining and sentiment analysis framework, called BuSeD, to analyze news business sentiment, which can help make economic decisions. It contains two main parts. At first, the news is collected by using crawler tools. The news dataset is preprocessed by reducing noises. Topics in each article is extracted by using traditional topic extraction tools. And then a CNN-based text analyzing model is designed to analyze news from sentence level. The workflow of BuSeD is shown in **Figure 1**.

Noise Reduction and News Topics Extraction

The news is first extracted by using the general-purpose web crawler (GPWC). GPWC is the whole web crawler. The implementation process is as follows. First, get the initial URL. The initial URL address can be manually specified by the user or determined by one or more initial crawling web pages specified

by the user. Second, crawl the page according to the initial URL and get a new URL. After obtaining the initial URL address, you need to crawl the web page in the corresponding URL address first, then store the web page in the original database, find a new URL address while crawling the web page, and store the crawled URL address in a URL list for de duplication and judging the crawling process. Third, put the new URL into the URL queue. After obtaining the next new URL address in the second step, the new URL address will be put into the URL queue. Fourth, read the new URL from the URL queue and crawl the web page according to the new URL. At the same time, get the new URL from the new web page and repeat the above crawling process. Fifth, stop crawling when the stop conditions set by the crawler system are met. When writing crawlers, corresponding stop conditions are usually set. If the stop condition is not set, the crawler will crawl until the new URL address cannot be obtained. If the stop condition is set, the crawler will stop crawling when the stop condition is met.

Gensim is used to extract news topics. The input of Gensim is original and unstructured digital text (plain text). The semantic structure of the document is automatically found by calculating the statistical cooccurrence mode in the training corpus. These algorithms are unsupervised, which means that there is no need for manual input - just a set of plain text corpus. Once these statistical patterns are found, any plain text (sentence, phrase, word) can be expressed concisely by semantic representation. An undirected weighted graph is constructed for news topics. The sentences in an article are modeled as nodes, and the normalized sentence similarity is used as the link of the nodes, which can help avoid the influence of sentence length on the summary results. In the topic extraction process, the keywords are not limited to a single word, to achieve the purpose of improving the readability of the abstract.

Although business news text has the advantages of strong professionalism and low noise, the “noise interference” caused by massive text data will seriously affect the accuracy of the model. Noise interference means that noises in speech signals can disturb the communication connection between people. The reason is noises can cover up the acoustical components to hide information from listeners. The way to reduce the impact is to remove non-technical nouns and some unimportant prepositions in the original news text. In addition, the research collects common stop words for English and some meaningless digital symbols that may appear in news texts. It also introduces filtering processing to increase the impact of emotional words and verbs, which represent the emotions and views of financial stakeholders on the model.

We use the method of k-means clustering to filter noises in the news texts. Clustering is to group a data set into multiple groups or clusters. The data in the same cluster are highly similar, while the data in different clusters are quite different. Typical clustering algorithms include k-means (K-mean) and k-medoids (k-center point) algorithms. K-means-based clustering can realize noise cleaning. It treats the data grouped in smaller clusters as noise. It mainly uses the distance-based noise recognition method. Cluster analysis is carried out on the data set based on the distance between two data.

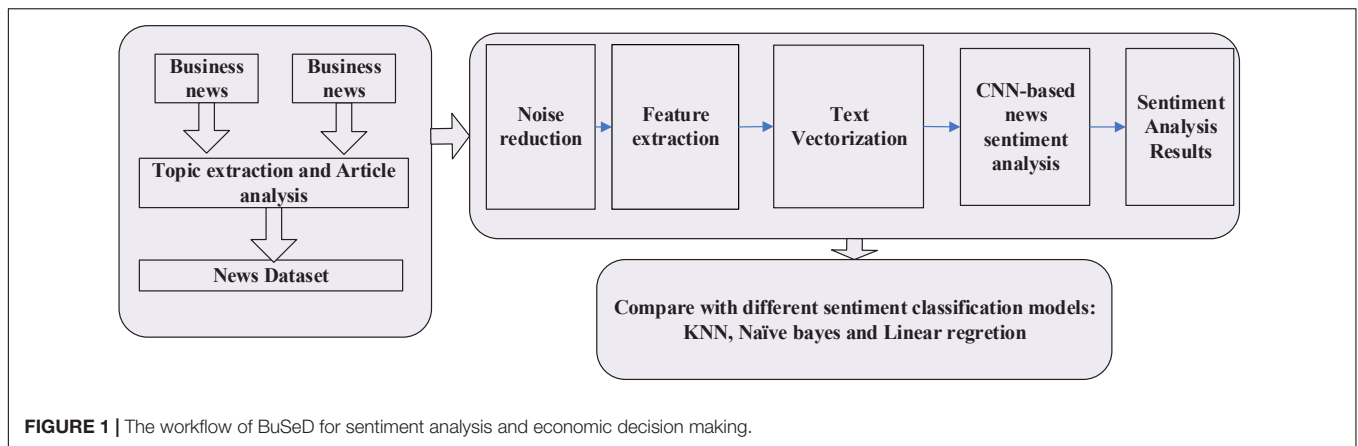


FIGURE 1 | The workflow of BuSeD for sentiment analysis and economic decision making.

Convolutional Neural Network-Based Fine-Grained Sentiment Analysis

A CNN-based text analyzing model is designed to analyze text sentiment from fine-grained sentence levels. The CNN is composed of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The input is a convolution layer with only one feature surface. Its special network structure enables it to capture small feature information. This technique has been widely used in the field of image recognition. The parameter setting of the proposed CNN model is shown in **Table 1**.

The input (embedding) layer converts the text into a vector and expands the dimension to meet the parameter requirements of the convolutional neural network. Then the convolutional layer and pooling layer are used to obtain local features and get the most important features. The convolutional layer is connected to the upper-layer feature surface through local connection and uses the feature of weight sharing to reduce the complexity of the model. The pooling layer uses max pooling to extract the most important features. During the training process, the Rectified Linear Unit (ReLU) is used as the activation function to transform the linear neural network into a nonlinear neural network. The output result is changed from $h(x)$ in Formula 1 to $\Gamma_k(x)$ in Formula 2, which can efficiently speed up the convergence.

$$h(x) = w^T x + b \tag{1}$$

$$\Gamma_k(x) = \max(w^T x + b, 0) \tag{2}$$

TABLE 1 | Structure of the CNN model for business sentiment analysis.

Layers	Cell number	Filter size
Convolution layer 1	150	(150,4)
Convolution layer 2	80	(80,4)
Pooling layer	80	-
Dense	10	Unit(10)

where $h(x)$ represents the output of the linear output layer. $\Gamma_k(x)$ is the ReLU activation function to transform $h(x)$ to non-linear form. k represents the k th convolution window, w^T is the weight matrix, and b is the bias value.

The corresponding output feature surface size after convolution and pooling operations is done by Formula 3.

$$\phi_i = \frac{1}{W_{sz}} \left(\frac{\gamma_{sz} - \mu'}{\tau} + 1 \right) \tag{3}$$

where ϕ_i is the size of the output feature surface. γ_{sz} represents the size of each input feature surface. $\mu' \in [6, 7, 10]$ is the size of the convolution kernel. τ represents the sliding step size of the convolution kernel on the upper layer, and W_{sz} is the size of the pooling window. The model is trained by adjusting the number of parameters of the convolutional layers so that the size of ϕ_i is an integer.

The features trained by the convolutional layer and the pooling layer are used as the input and output classification results of the fully connected layer. That is, according to the probability distribution of sentences in different categories, output sentiment level labels for each sentence. $o(i)$ is the output of the text on the i th emotional tendency, which represents the probability that the text is classified as the i th emotional tendency. After normalization by softmax, $o(i)$ is expressed as:

$$o(\sigma_i) = \frac{\exp(\theta_i \cdot v + b_i)}{\sum_{j=1}^n \exp(\theta_j \cdot v + b_j)} \tag{4}$$

where θ_i and b_i are the parameters and biases of the fully connected layer corresponding to the output σ_i , respectively, and n is the total number of output categories.

To avoid over-fitting, dropout is used to make the features learned by neurons more robust. In addition, the model also limits the two-normal form size of the weight vector to achieve better classification results. In terms of measuring the prediction effect, this model uses cross entropy as the loss function. We apply CNN to text sentiment analysis, so the CNN model finally outputs sentiment classification results according to four rating scales, which are extremely negative, negative, positive and extremely positive.

TABLE 2 | Topics extracted from the crawled news.

Words	#	Words	#	Words	#	Words	#
Money	1200	Profitability	563	National	197	Countries	124
Opportunity	987	Political	338	Entrepreneur	188	Investment	130
Economic	933	International	372	Business plan	244	Power	270
Trade	670	Marketing	468	Business daily	278	Government	256
Income	489	Global	270	Financial news	345	Policy	148

TABLE 3 | Comparison of sentiment classification results.

Methods	Precision	Recall	F1-Score
BuSeD	0.968	0.924	0.933
LR	0.877	0.862	0.843
SVM	0.923	0.894	0.903
KNN	0.912	0.894	0.908
DCNN	0.872	0.826	0.855
1D-CNN	0.928	0.915	0.920

EXPERIMENT ANALYSIS

Experiment Data and Settings

We crawl more than 10,000 business news in recent five years from the Web. The training set data consists of 10,000 randomly selected finance and economics related news and 500 automatic abstracts of news retrieved with “business” as the keyword. The news from January 2018 to December 2021 were randomly selected according to the same proportion to ensure that the style of the media reports did not affect the training results. Each sentence was manually annotated, and a total of 80,000 sentences were annotated. The test set consists of 3,000 articles, which are formed by excluding 7,000 news articles selected as the training set from the collected 10,000 news articles and adopt the same labeling method as the training set. The evaluation criteria for sentiment analysis results in this paper are Precision, Recall and F1-Score (Sun et al., 2021).

Compared Sentiment Analysis Models

To accurately analyze the sentiment in news data, other machine learning methods have also been tried in this work. Naive Bayes is a classification method based on Bayes’ theorem and the assumption of independence of feature conditions, and it is also one of the popular and popular classification models. Because NB needs to estimate few parameters, it is not sensitive to missing data, and the algorithm is relatively simple. In theory, the NB model has the smallest error rate compared with other classification methods, so it can be used as one of the candidate models for comparison of results.

Logistic Regression (LR) (Mehroliya et al., 2021), also known as logistic regression analysis, is a generalized linear regression analysis model, which is often used in data mining, automatic disease diagnosis, economic forecasting, and other fields. The independent variables can be either continuous or discrete. Through Logistic regression analysis, the weights of independent variables can be obtained to automatically perform features choose.

Support Vector Machine (SVM) (Zhou et al., 2021) is a commonly used discrimination method. In the field of machine learning, it is a supervised learning model that is usually used for pattern recognition, classification, and regression analysis. After researching many literatures, we found that many scholars now prefer to choose SVM for modeling analysis. Based on this, it is also used as one of the candidate models here.

KNN neighbor algorithm, or K-Nearest Neighbor (KNN) classification algorithm (Khorshid and Abdulazeez, 2021). It is one of the most effective and simple methods in data mining classification technology. The so-called K nearest neighbors is the k nearest neighbors, which means that each sample can be characterized and characterized by its nearest k neighbors. The core idea of the KNN algorithm is that if most of the k nearest neighbors of a sample in the feature space belong to a certain category, then the sample also belongs to this category and has the characteristics of the samples in this category. This method only decides the category of the sample to be classified according to the category of the nearest one or several samples in the decision of outputting the classification result. The KNN method is only related to a very small number of adjacent samples when making class decisions. Moreover, this method mainly uses the limited surrounding samples, rather than the method based on the discriminative class domain, to determine the class to which it belongs. Therefore, the KNN method is more suitable than other methods for the sample set to be divided with more overlapping or overlapping of class domains.

Dynamic convolutional neural network (DCNN) (Kalchbrenner et al., 2014). It integrates the dynamic k-max pooling.

One dimension CNN (1D-CNN) (Kim, 2014). It pretrains the word embeddings based on large amounts of news texts. It also adopts optimization strategy for fine-tuning.

Methods of processing text into word vectors include TF-IDF, Out-of-Dict, Word-Frequency, One-Hot, Word-Frequency and Google’s Word2Vec.

Performance Comparison Between BuSeD and Classical Algorithms

We first analyze the high-frequency words of 10,000 news articles. After the stop words are removed, the top 20 words in the number of occurrences are shown in **Table 2**. As can be seen from **Table 2**, the high-frequency words are concentrated on the objects of the business and economics.

Table 3 shows the comparison of BuSeD, LR, SVM, KNN, DCNN, and 1D-CNN in terms of precision, recall and F1-score. Compared with the four classical models, the proposed BuSeD achieves the optimal performance. The precision, recall and F1 of BuSeD are 0.968, 0.924, and 0.933 respectively, which are much higher than the values of LR, SVM and KNN.

Validity Analysis of BuSeD

To further verify the validity of the sentiment model, a quantitative trading application based on sentiment analysis is analyzed. The application shows the correspondence between the real-time stock price data trends and the subject sentiment

classified by BuSeD. Our study shows that when the stock price keeps rising, the sentiment of the financial news corresponding to the market should be positive. When the stock price keeps falling, the sentiment of the news corresponding to the market should be negative. Based on this hypothesis, we will try to verify the validity of the sentiment model BuSeD by the trend change of stock price and the synchronous change of news sentiment.

The stock price trend is mainly analyzed and predicted in terms of the stock price trend, news popularity and public opinion analysis at runtime. In terms of stock price trends, the study obtained the daily high/low and opening/closing prices of official stock prices through multiple channels and customized the corresponding stock price curve and box charts. In terms of news popularity, the research counted the news popularity of Hang Seng Index constituent stocks within 8 years and set up real-time follow-up on each day. In terms of public opinion analysis, the opinions, and sentiments of financial stakeholders on the stock price trend are extracted based on the BuSeD model.

The research uses the financial public opinion analysis module to analyze the relationship between news sentiment and stock price. Financial public opinion analysis refers to using BuSeD to analyze the sentiment polarity of news, and then infer the state of the market and the future stock price trends based on the direct relationship between the number of positive and negative news and stock price fluctuations. Through analysis, there is a strong correlation between news public opinion and stock price. When the positive news increases sharply or the negative news accumulates sharply, the market is in a turning period, and the stock price also moves on the corresponding direction.

CONCLUSION

To improve the accuracy of sentiment classification, we proposed a robust news mining-based business sentiment

analysis framework, called BuSeD. It has two main components: news collection and preprocessing, and feature extraction and sentiment classification. In news collection and preprocessing component, the news is collected by using crawler tools. The news dataset is then preprocessed by reducing noises. In feature extraction and sentiment classification component, topics in each article is extracted by using traditional topic extraction tools. And then a convolutional neural network (CNN)-based text analyzing model is designed to analyze news from sentence level. We conducted comprehensive experiments to evaluate the performance of BuSeD for sentiment classification. Compared with four classical classification algorithms, the proposed CNN-based classification model of BuSeD achieves the highest F1 scores. We also presented a quantitative trading application based on sentiment analysis to validate BuSeD, which indicates that the news-based business sentiment analysis has high economic application value.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MY proposed the idea and wrote the manuscript. BJ discussed with MY for proposing the idea and writing the manuscript. YW collected dataset and conducted the experiment. TH helped to collect dataset and provided suggestions in revision. YL helped to coding. All authors contributed to the article and approved the submitted version.

REFERENCES

- Alantari, H. J., Currim, I. S., Deng, Y., and Singh, S. (2022). An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. *Int. J. Res. Mark.* 39, 1–19.
- Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., and Vilares, J. (2021). Sentiment analysis for fake news detection. *Electronics* 10:1348.
- Hu, W., and Wang, X. (2019). Balancing university teaching and media industry needs: a case study of teaching finance and economics news translation. *Int. J. High. Educ.* 8, 247–256.
- Jo, W., and You, M. (2019). News media's framing of health policy and its implications for government communication: a text mining analysis of news coverage on a policy to expand health insurance coverage in South Korea. *Health Policy* 123, 1116–1124. doi: 10.1016/j.healthpol.2019.07.011
- Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Association for Computational Linguistics*, Baltimore, MD, 655–665.
- Khorshid, S. F., and Abdulazez, A. M. (2021). Breast cancer diagnosis based on k-nearest neighbors: a review. *PalArchs J. Archaeol. Egypt Egyptol.* 18, 1927–1951.
- Kim, Y. (2014). "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 1746–1751. doi: 10.1371/journal.pone.0253308
- Kim, Y., Sidtis, D. V. L., and Sidtis, J. J. (2021). Emotional nuance enhances verbatim retention of written materials. *Front. Psychol.* 12:2273. doi: 10.3389/fpsyg.2021.519729
- Kovalev, A. (2021). The political economics of news making in Russian media: ownership, clickbait and censorship. *Journalism* 22, 2906–2918.
- Lee, S. M., and Hong, S. G. (2020). Policy agenda proposals from text mining analysis of patents and news articles. *J. Digit. Converg.* 18, 1–12.
- Mehroliya, S., Alagarsamy, S., and Solaikutty, V. M. (2021). Customers response to online food delivery services during COVID-19 outbreak using binary logistic regression. *Int. J. Consum. Stud.* 45, 396–408. doi: 10.1111/ijcs.12630
- Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., and Anastasiu, D. C. (2019). "Stock price prediction using news sentiment analysis," in *Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, (Newark, CA: IEEE), 205–208.
- Pan, D., Yang, J., Zhou, G., and Kong, F. (2020). The influence of COVID-19 on agricultural economy and emergency mitigation measures in China: a text mining analysis. *PLoS One* 15:e0241167. doi: 10.1371/journal.pone.0241167
- Pejić Bach, M., Krstić, Ž, Seljan, S., and Turulja, L. (2019). Text mining for big data analysis in financial sector: a literature review. *Sustainability* 11: 1277.

- Provoost, S., Ruwaard, J., Van Breda, W., Riper, H., and Bosse, T. (2019). Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study. *Front. Psychol.* 10:1065. doi: 10.3389/fpsyg.2019.01065
- Schmeller, R. (2022). Can business news articles be used to teach business model analysis? *Dev. Bus. Simulat. Exp. Learn.* 49, 129–134.
- Souma, W., Vodenska, I., and Aoyama, H. (2019). Enhanced news sentiment analysis using deep learning methods. *J. Comput. Soc. Sci.* 2, 33–46. doi: 10.7717/peerj-cs.476
- Štrimaitis, R., Stefanovič, P., Ramanauskaitė, S., and Slotkiene, A. (2021). Financial context news sentiment analysis for the Lithuanian language. *Appl. Sci.* 11:4443.
- Sun, B., Mao, H., and Yin, C. (2020). Male and female users' differences in online technology community based on text mining. *Front. Psychol.* 11:806. doi: 10.3389/fpsyg.2020.00806
- Sun, L., Wang, Y., Qu, Z., and Xiong, N. N. (2021). BeatClass: a sustainable ECG classification system in IoT-based eHealth. *IEEE Internet Things J.* 9, 7178–7195. doi: 10.1109/JIOT.2021.3108792
- Sun, L., Zhong, Z., Qu, Z., and Xiong, N. (2022). PerAE: an effective personalized AutoEncoder for ECG-based biometric in augmented reality system. *IEEE J. Biomed. Health Inform.* doi: 10.1109/JBHI.2022.3145999
- Veltmeijer, E. A., Gerritsen, C., and Hindriks, K. (2021). Automatic emotion recognition for groups: a review. *IEEE Trans. Affect. Comput.* doi: 10.1109/TAFFC.2021.3065726
- Zhang, Z., Wei, X., Zheng, X., Li, Q., and Zeng, D. D. (2022). Detecting product adoption intentions via multiview deep learning. *INFORMS J. Comput.* 34, 541–556.
- Zhou, J., Qiu, Y., Zhu, S., Armaghani, D. J., Li, C., Nguyen, H., et al. (2021). Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. *Eng. Appl. Artif. Intell.* 97: 104015.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Jiang, Wang, Hao and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.