1. **A hybrid approach for stock trend prediction based on tweets embedding and historical prices**
    1. **Code:** No
    2. **Year:** 2021
    3. **Journal/Conference:** *World Wide Web*
    4. **Abstract:** Recently, the development of data mining and natural language processing techniques enable the relationship probe between social media and stock market volatility. The integration of natural language processing, deep learning and the financial field is irresistible. This paper proposes a hybrid approach for stock market prediction based on tweets embedding and historical prices. Different from the traditional text embedding methods, our approach takes the internal semantic features and external structural characteristics of Twitter data into account, such that the generated tweet vectors can contain more effective information. Specifically, we develop a Tweet Node algorithm for describing potential connection in Twitter data through constructing the tweet node network. Further, our model supplements emotional attributes to the Twitter representations, which are input into a deep learning model based on attention mechanism together with historical stock price. In addition, we designed a visual interactive stock prediction tool to display the result of the prediction.
    5. **Link:** https://link.springer.com/article/10.1007/s11280-021-00880-9

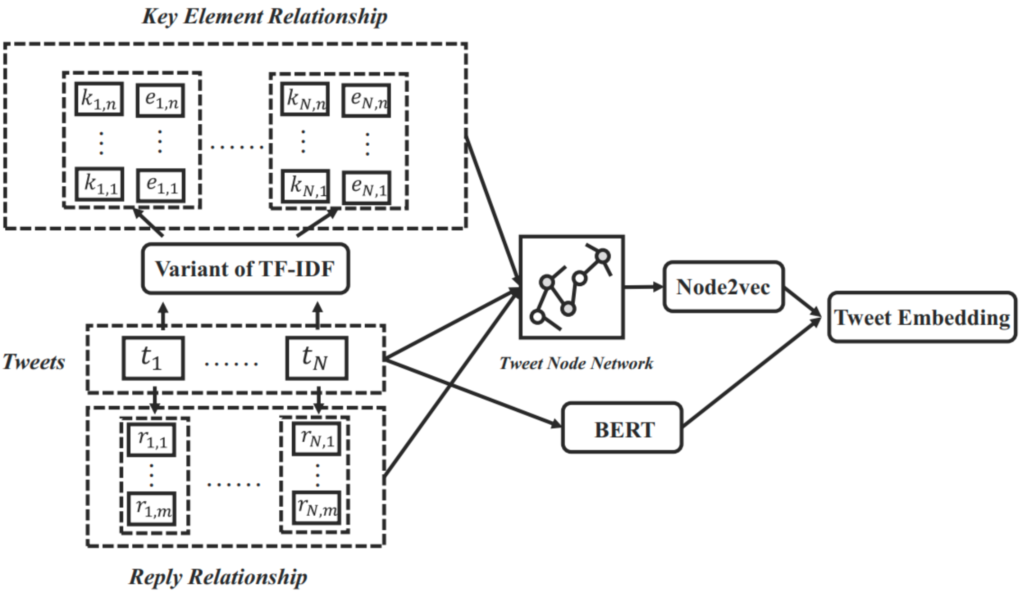


Figure 1 The Tweet Node Model. The model extracts the reply relations r and the key elements e of each tweet, then calculates the weight k of each element in tweets. Then the tweet node network is constructed according to the above three. Finally, the model introduces Node2vec and BERT to extract semantic factors and emotional factors respectively and generates the tweet embeddings

2. **A self-regulated generative adversarial network for stock price movement prediction based on the historical price and tweets**
    1. **Code:** NO
    2. **Year:** 2022
    3. **Journal/Conference:** Knowledge-Based Systems
    4. **Abstract:** Stock price movement prediction is an important task of the financial prediction field. The current mainstream approaches usually apply financial texts and

some corresponding stock price information to predict the stock price movement. However, the current methods usually suffer from two shortcomings: (1) To reduce the stochasticity in the stock price and financial text information, some researchers adopt generative models to better treat the stochasticity while enduring the overfitting problem during training. (2) Although the current state-of-the-art methods based on the generative adversarial network have been proposed to reduce the overfitting, they only concentrate on the overfitting problem of the stock price information and neglect the above problem of financial text information with higher stochasticity. In this paper, we propose a self-regulated generative adversarial network by combining the generative adversarial network and cooperative network for the stock price movement prediction. Furthermore, the proposed model can effectively reduce the stochasticity and overfitting problems simultaneously for the stock price and the financial text information. The experimental results on the currently commonly used stock dataset based on tweets confirm that the proposed method can achieve the novelly state-of-the-art performance compared with some current advances.

5. **Link:** https://www.sciencedirect.com/science/article/pii/S0950705122003288

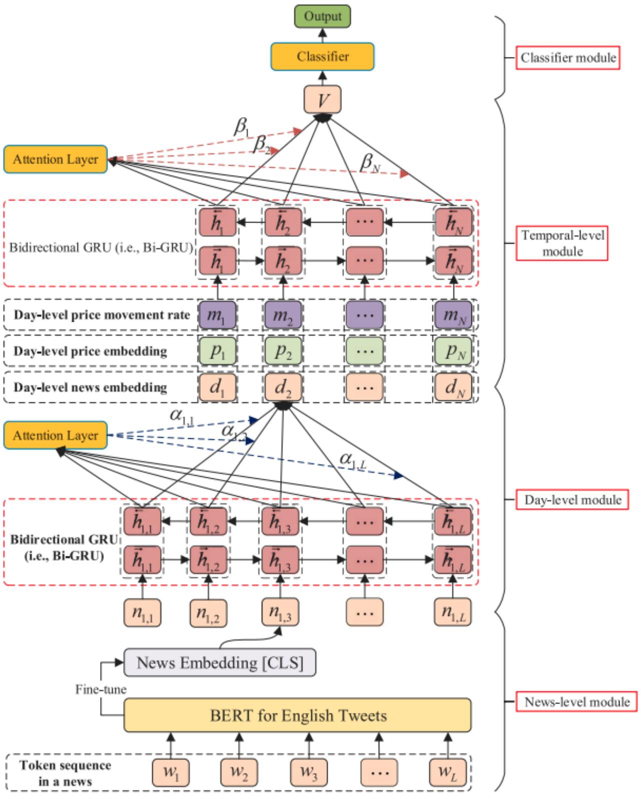

Output
Classifier
V
Classifier module
β_1
β_2
β_N
Attention Layer
Bidirectional GRU (i.e., Bi-GRU)
$\overleftarrow{h}_1$ $\overleftarrow{h}_2$ ··· $\overleftarrow{h}_N$
$\overrightarrow{h}_1$ $\overrightarrow{h}_2$ ··· $\overrightarrow{h}_N$
Temporal-level module
Day-level price movement rate
m_1 m_2 ··· m_N
Day-level price embedding
p_1 p_2 ··· p_N
Day-level news embedding
d_1 d_2 ··· d_N
$\alpha_{1,1}$
$\alpha_{1,2}$
$\alpha_{1,L}$
Attention Layer
Bidirectional GRU (i.e., Bi-GRU)
$\overleftarrow{h}_{1,1}$ $\overleftarrow{h}_{1,2}$ $\overleftarrow{h}_{1,3}$ ··· $\overleftarrow{h}_{1,L}$
$\overrightarrow{h}_{1,1}$ $\overrightarrow{h}_{1,2}$ $\overrightarrow{h}_{1,3}$ ··· $\overrightarrow{h}_{1,L}$
Day-level module
$n_{1,1}$ $n_{1,2}$ $n_{1,3}$ ··· $n_{1,L}$
News Embedding [CLS]
Fine-tune
BERT for English Tweets
News-level module
Token sequence in a news
w_1 w_2 w_3 ··· w_L

Fig. 3. The components of the cooperative network in the proposed method.

3. **FAST: Financial News and Tweet Based Time Aware Network for Stock Trading**

    1. Code: No
    2. Year: 2021
    3. **Journal/Conference:** EACL 2021
    4. **Abstract:** Designing profitable trading strategies is complex as stock movements are highly stochastic; the market is influenced by large volumes of noisy data across diverse information sources like news and social media. Prior work mostly treats stock movement prediction as a regression or classification task and is not directly optimized towards profit-making. Further, they do not model the fine-grain temporal irregularities in the release of vast volumes of text that the market responds to quickly. Building on these limitations, we propose a novel hierarchical, learning to rank approach that uses textual data to make time-aware predictions for ranking stocks

based on expected profit. Our approach outperforms state-of-the-art methods by over 8% in terms of cumulative profit and risk-adjusted returns in trading simulations on two benchmarks: English tweets and Chinese financial news spanning two major stock indexes and four global markets. Through ablative and qualitative analyses, we build the case for our method as a tool for daily stock trading.

5. **Link:** https://aclanthology.org/2021.eacl-main.185/

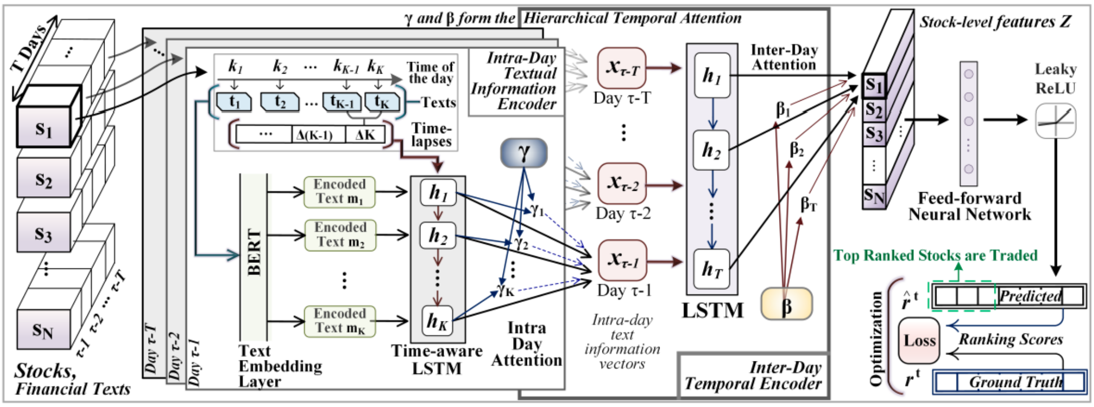


Figure 3: FAST: Time-aware modeling, hierarchical temporal attention, joint optimization for ranking.

4. **Accurate Stock Movement Prediction with Self-supervised Learning from Sparse Noisy Tweets**
    1. **Code:** [GitHub - deeptrade-public/slot: repository for accepted paper in BigData 2022 conference](https://github.com/deeptrade-public/slot)
    2. **Year:** 2022
    3. **Journal/Conference:** [2022 IEEE International Conference on Big Data (Big Data)]()
    4. **Abstract:** Given historical stock prices and sparse tweets, how can we accurately predict stock price movement? Many market analysts strive to use a large amount of information for stock price prediction, and Twitter is one of the richest sources of information presenting real-time opinions of people. However, previous works that use tweet data in stock movement prediction have suffered from two limitations. First, the number of tweets is heavily biased towards only a few popular stocks, and most stocks have insufficient evidence for accurate price prediction. Second, many tweets provide noisy information irrelevant of actual price movement, and extracting reliable information from tweets is as challenging as predicting stock prices.In this paper, we propose SLOT (Self-supervised Learning of Tweets for Capturing Multi-level Price Trends), an accurate method for stock movement prediction. SLOT has two main ideas to address the limitations of previous tweet-based models. First, SLOT learns embedding vectors of stocks and tweets in the same semantic space through self-supervised learning. The embeddings allow us to use all available tweets to improve the prediction for even unpopular stocks, addressing the sparsity problem. Second, SLOT learns multi-level relationships between stocks from tweets, rather than using them as direct evidence for prediction, making it robust to the unreliability of tweets. Extensive experiments on real world datasets show that SLOT provides the state-of-the-art accuracy of stock movement prediction.
    5. **Link:** https://ieeexplore.ieee.org/abstract/document/10020720

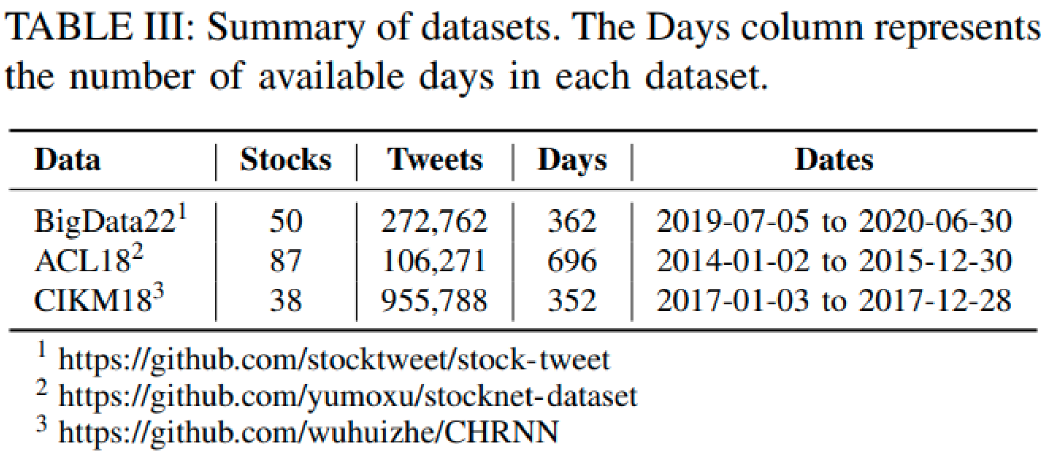
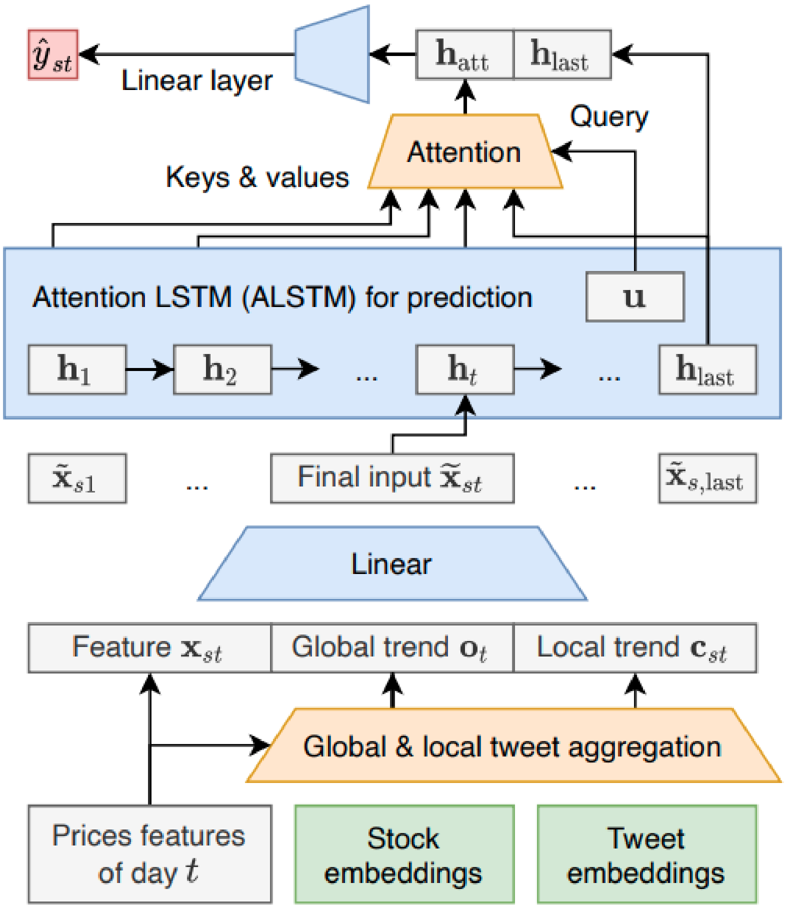


TABLE III: Summary of datasets. The Days column represents the number of available days in each dataset.

Data	Stocks	Tweets	Days	Dates
BigData22[1]	50	272,762	362	2019-07-05 to 2020-06-30
ACL18[2]	87	106,271	696	2014-01-02 to 2015-12-30
CIKM18[3]	38	955,788	352	2017-01-03 to 2017-12-28

[1] https://github.com/stocktweet/stock-tweet
[2] https://github.com/yumoxu/stocknet-dataset
[3] https://github.com/wuhuizhe/CHRNN



$\hat{y}_{st}$
Linear layer
$\mathbf{h}_{\text{att}}$
$\mathbf{h}_{\text{last}}$
Query
Attention
Keys & values
Attention LSTM (ALSTM) for prediction
$\mathbf{u}$
$\mathbf{h}_1$
$\mathbf{h}_2$
...
$\mathbf{h}_t$
...
$\mathbf{h}_{\text{last}}$
$\tilde{\mathbf{x}}_{s1}$
...
Final input $\tilde{\mathbf{x}}_{st}$
...
$\tilde{\mathbf{x}}_{s,\text{last}}$
Linear
Feature $\mathbf{x}_{st}$
Global trend $\mathbf{o}_t$
Local trend $\mathbf{c}_{st}$
Global & local tweet aggregation
Prices features of day t
Stock embeddings
Tweet embeddings

Fig. 4: The overall structure of SLOT for making a prediction $\hat{y}_{st}$ for stock s at day t. SLOT learns stock and tweet embeddings by self-supervised learning (Sec. III-C) and creates two types of trend features (Sec. III-D and III-E). The ALSTM model combines the three types of features for prediction.

5. **Using Google Trends and Baidu Index to analyze the impacts of disaster events on company stock prices**
    1. **Code:** No
    2. **Year:** 2020

3. **Journal/Conference:** [Industrial Management & Data Systems]
4. **Abstract:**

Purpose

With the ascendance of information technology, particularly through the internet, external information sources and their impacts can be readily transferred to influence the performance of financial markets within a short period of time. The purpose of this paper is to investigate how incidents affect stock prices and volatility using vector error correction and autoregressive-generalized auto regressive conditional Heteroskedasticity models, respectively.

Design/methodology/approach

To characterize the investors' responses to incidents, the authors introduce indices derived using search volumes from Google Trends and the Baidu Index.

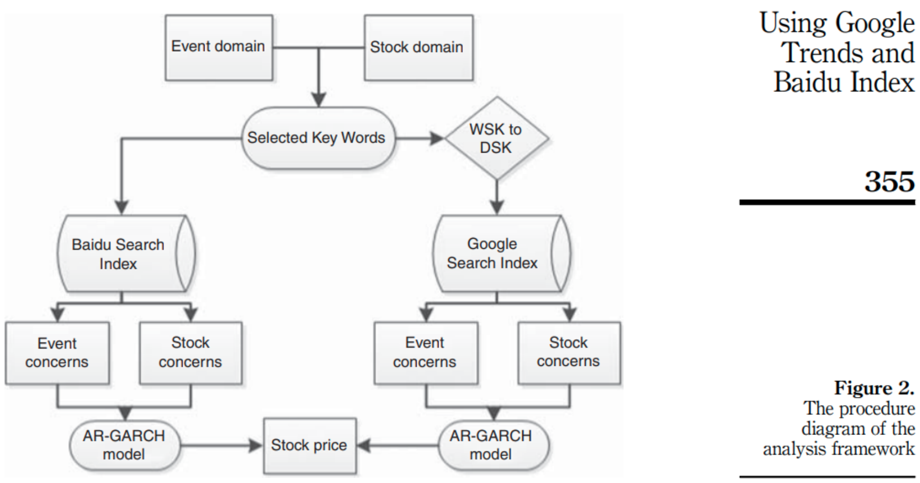
Findings

The empirical results indicate that an outbreak of disasters can increase volatility temporarily, and exert significant negative effects on stock prices in a relatively long time. In addition, indices derived from different search engines show differentiation, with the Google Trends search index mainly representing international investors and appearing more significant and persistent.

Originality/value

This study contributes to the existing literature by incorporating open-source data to analyze how catastrophic events affect financial markets and effect persistence.

5. **Link:** https://www.emerald.com/insight/content/doi/10.1108/IMDS-03-2019-0190/full/html



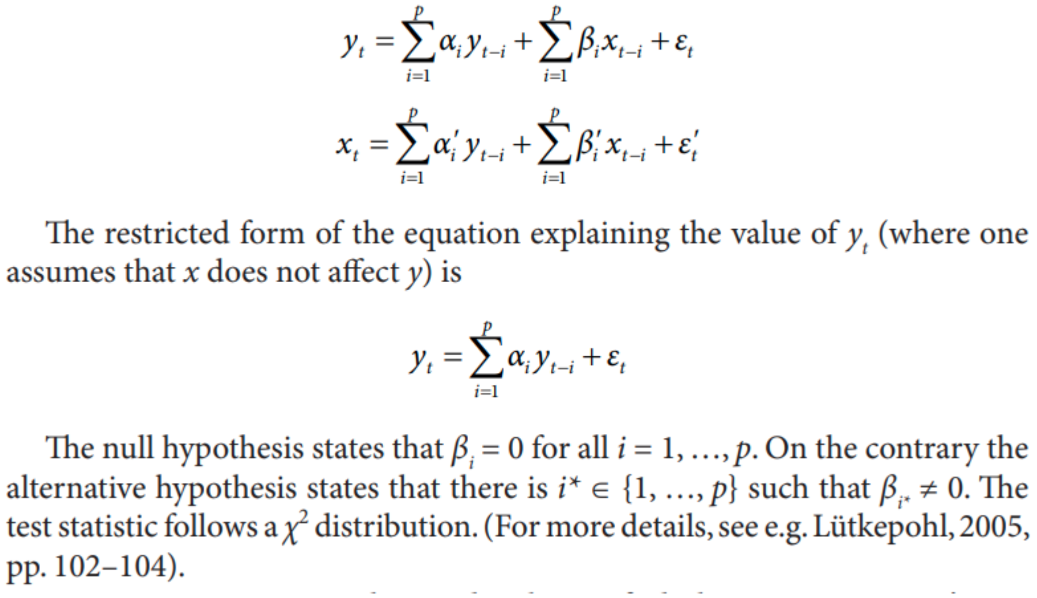
Using Google Trends and Baidu Index

355

Figure 2. The procedure diagram of the analysis framework

6. **How Google Trends can improve market predictions— the case of the Warsaw Stock Exchange**
   1. **Code:** No
   2. **Year:** 2022
   3. **Journal/Conference:** Economics and Business Review, 2022

4. **Abstract:** The aim of this paper is to investigate interdependencies between the WIG20 index and economic policy uncertainty (EPU) related keywords quantified by a Google Trends search index. Tests for two periods from January 2015 till December 2019 and from June 2016 till May 2021 have been performed. This allowed the period of relative stability from the time of economic shock caused by the COVID-19 pandemics followed by various restrictions imposed by the governments to be distinguished. A bivariate VAR model to selected search terms and the value of the WIG20 index was applied. After using AIC to establish the optimal number of lags the Granger causality test was performed. The increased empirical relationship has been confirmed between twelve EPU related terms and changes in the WIG20 index in the second period versus six terms for the pre-COVID period. It was also found that in the post-COVID period the intensity of reverse relations increased.
5. **Link:** https://sciendo.com/article/10.18559/ebr.2022.2.2



$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t$$

$$x_t = \sum_{i=1}^{p} \alpha_i' y_{t-i} + \sum_{i=1}^{p} \beta_i' x_{t-i} + \varepsilon_t'$$

The restricted form of the equation explaining the value of y_t (where one assumes that x does not affect y) is

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t$$

The null hypothesis states that $\beta_i = 0$ for all $i = 1, \dots, p$. On the contrary the alternative hypothesis states that there is $i^* \in \{1, \dots, p\}$ such that $\beta_{i^*} \neq 0$. The test statistic follows a χ^2 distribution. (For more details, see e.g. Lütkepohl, 2005, pp. 102–104).

7. **Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong**
   1. **Code:** No
   2. **Year:** 2020
   3. **Journal/Conference:** Information Processing & Management
   4. **Abstract:** Stock prediction via market data analysis is an attractive research topic. Both stock prices and news articles have been employed in the prediction processes. However, how to combine technical indicators from stock prices and news sentiments from textual news articles, and make the prediction model be able to learn sequential information within [time series](#) in an intelligent way, is still an [unsolved problem](#). In this paper, we build up a stock prediction system and propose an approach that 1) represents numerical price data by technical indicators via technical analysis, and represents textual news articles by sentiment vectors via [sentiment analysis](#), 2) setup a layered [deep learning model](#) to learn the sequential information within market snapshot series which is constructed by the technical indicators and news sentiments, 3) setup a fully connected [neural network](#) to make stock predictions. Experiments have been conducted on more than five years of Hong Kong Stock Exchange data using four different sentiment dictionaries, and results show that 1) the proposed approach outperforms the baselines in both validation and test sets using two different [evaluation metrics](#), 2) models incorporating prices and news sentiments outperform models that only use either technical indicators or news sentiments, in both individual

stock level and sector level, 3) among the four sentiment dictionaries, finance domain-specific sentiment dictionary (Loughran–McDonald Financial Dictionary) models the news sentiments better, which brings more prediction performance improvements than the other three dictionaries.

5. **Link:** https://www.sciencedirect.com/science/article/pii/S0306457319307952

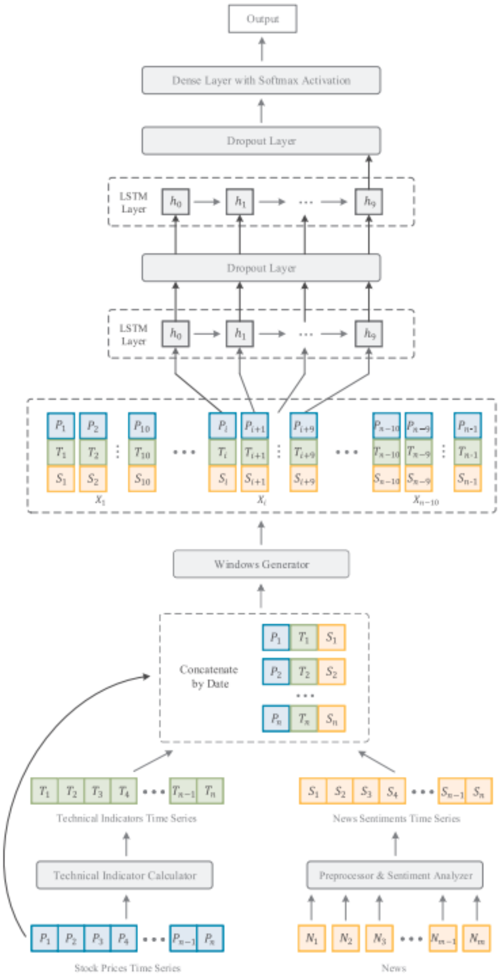


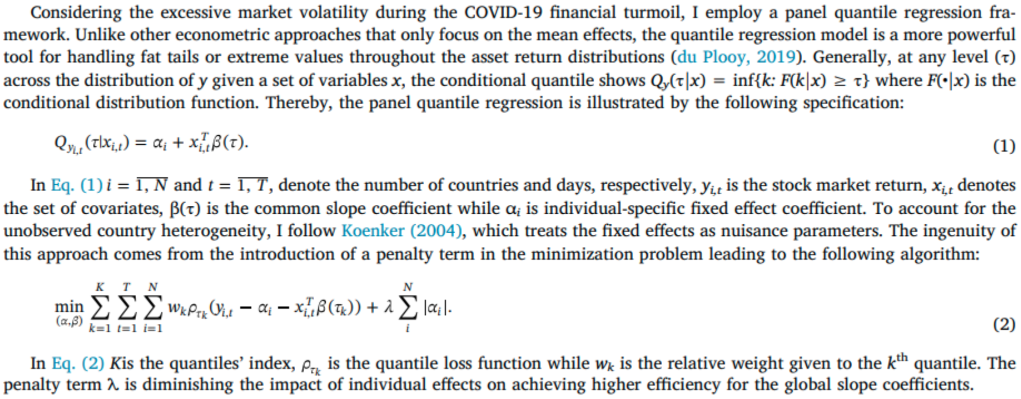
Fig. 1. LSTM based workflow of stock prediction incorporating prices and news sentiments.

8. **Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil**

    1. **Code:** No
    2. **Year:** 2020
    3. **Journal/Conference:** Finance Research Letters
    4. **Abstract:** The impact of public news sentiment on stock returns has received increasing attention in recent years. A growing body of empirical and theoretical studies has focused on understanding whether price movements in financial markets are driven by economic or political news (Smales, 2014; Broadstock and Zhang, 2019; Shi and Ho, 2020). The consensus is that the information arriving from social media channels exerts a significant influence on the stock market dynamic, especially in times of economic or political uncertainty. Given the COVID-19 pandemic and the considerable amount of related news, stock markets around the world have suffered enormous losses in the first three months of 2020. According to Bloomberg, “through 1 p.m. on March 18, the S&P 500 index was off 27% for the year to date, Germany's DAX was down 38% and Japan's Nikkei was off 29%.” Consequently, the governments around the world have undertaken a series of stimulus packages to offset the damages produced by the pandemic and to regain investor's confidence. Although the major stock market indexes have partially recovered in the middle of April 2020, a great deal of financial uncertainty

remains. While the current literature relating the COVID-19 pandemic to financial markets is limited, the existing studies have provided some very interesting results. For example, Corbet et al. (2020a) reveal a negative knock-on impact from the coronavirus on some companies with similar names. In addition, Akhtaruzzaman et al. (2020), show that listed firms across China and G7 countries have experienced significant increases in the conditional correlations for the market returns. This fact is confirmed by Okorie and Lin (2020) which found considerable fractal contagion on the market return and market volatility. Moreover, Conlon and McGee (2020) and Goodell and Goutte (2020) suggest that cryptocurrencies do not act like safe havens during COVID-19 turmoil. In this paper, I contribute to the literature by investigating the stock market's reaction to coronavirus news in the top six most affected countries by the pandemic1 . By employing a panel quantile regression model, I show that the stock markets present asymmetric dependencies with COVID-19 related information. Specifically, the fake news exerts a negative influence on the lower and the middle quantiles throughout the distribution of returns; however, their impact is not statistically significant for the extreme values. Moreover, the media coverage leads to a decrease in returns across middle and upper quantiles and has no effects on the lower ones. Similarly, the financial contagion across companies is detrimental to returns from 50th to 75th quantiles. Furthermore, the estimates show that the gold price dynamic has a nonlinear impact on equity markets, especially during extreme bearish and bullish markets. The rest of the paper has the following structure: Section 2 presents the data, Section 3 discusses the econometric approach, and the results are in Section 4. Section 5 concludes the paper.

5. **Link:** https://www.sciencedirect.com/science/article/pii/S1544612320305912



Considering the excessive market volatility during the COVID-19 financial turmoil, I employ a panel quantile regression framework. Unlike other econometric approaches that only focus on the mean effects, the quantile regression model is a more powerful tool for handling fat tails or extreme values throughout the asset return distributions (du Plooy, 2019). Generally, at any level (τ) across the distribution of y given a set of variables x, the conditional quantile shows $Q_y(\tau|x) = \inf\{k\colon F(k|x) \geq \tau\}$ where $F(\bullet|x)$ is the conditional distribution function. Thereby, the panel quantile regression is illustrated by the following specification:

$$Q_{y_{i,t}}(\tau|x_{i,t}) = \alpha_i + x_{i,t}^T\beta(\tau). \tag{1}$$

In Eq. (1) $i = \overline{1, N}$ and $t = \overline{1, T}$, denote the number of countries and days, respectively, $y_{i,t}$ is the stock market return, $x_{i,t}$ denotes the set of covariates, $\beta(\tau)$ is the common slope coefficient while α_i is individual-specific fixed effect coefficient. To account for the unobserved country heterogeneity, I follow Koenker (2004), which treats the fixed effects as nuisance parameters. The ingenuity of this approach comes from the introduction of a penalty term in the minimization problem leading to the following algorithm:

$$\min_{(\alpha,\beta)} \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N} w_k\rho_{\tau_k}(y_{i,t} - \alpha_i - x_{i,t}^T\beta(\tau_k)) + \lambda\sum_{i}^{N}|\alpha_i|. \tag{2}$$

In Eq. (2) $K$is the quantiles' index, ρ_{τ_k} is the quantile loss function while w_k is the relative weight given to the k^{th} quantile. The penalty term λ is diminishing the impact of individual effects on achieving higher efficiency for the global slope coefficients.

9. **Stock market prediction using machine learning classifiers and social media, news**

    1. Code: No
    2. Year: 2022
    3. **Journal/Conference:** Journal of Ambient Intelligence and Humanized Computing
    4. **Abstract:** Accurate stock market prediction is of great interest to investors; however, stock markets are driven by volatile factors such as microblogs and news that make it hard to predict stock market index based on merely the historical data. The enormous stock market volatility emphasizes the need to effectively assess the role of external factors in stock prediction. Stock markets can be predicted using machine learning algorithms on information contained in social media and financial news, as this data can change investors' behavior. In this paper, we use algorithms on social media and financial news data to discover the impact of this data on stock market prediction accuracy for ten subsequent days. For improving performance and quality of predictions, feature selection and spam tweets reduction are performed on the data sets. Moreover, we perform experiments to find such stock markets that are difficult to predict and those that are more influenced by social media and financial news. We

compare results of different algorithms to find a consistent classifier. Finally, for achieving maximum prediction accuracy, deep learning is used and some classifiers are ensembled. Our experimental results show that highest prediction accuracies of 80.53% and 75.16% are achieved using social media and financial news, respectively. We also show that New York and Red Hat stock markets are hard to predict, New York and IBM stocks are more influenced by social media, while London and Microsoft stocks by financial news. Random forest classifier is found to be consistent and highest accuracy of 83.22% is achieved by its ensemble.

5. **Link:** https://link.springer.com/article/10.1007/s12652-020-01839-w

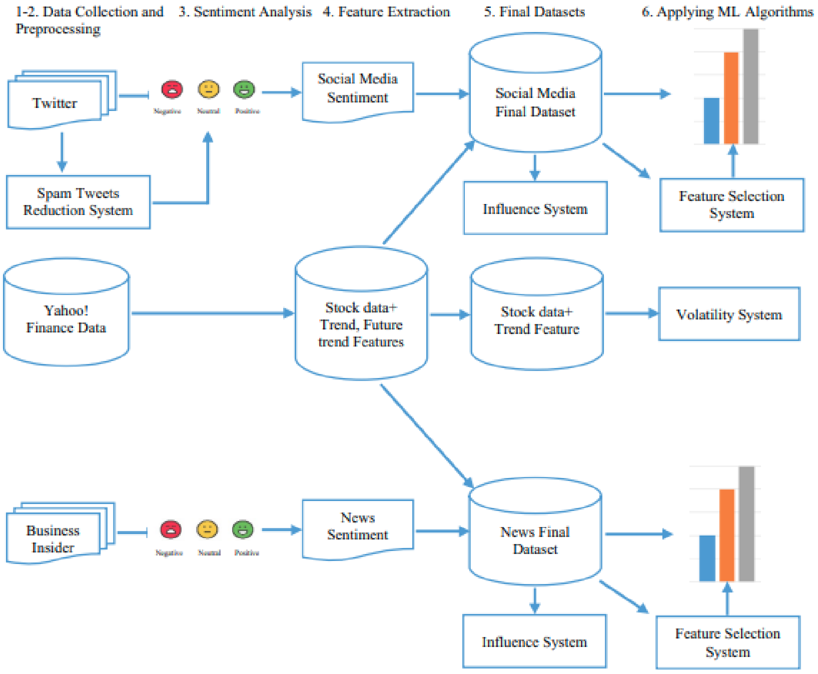


Fig. 2 Flow chart of the steps in our proposed framework for stock market forecasting using financial news and social media

10. **Astock: A New Dataset and Automated Stock Trading based on Stock-specific News Analyzing Model**
    1. Code: GitHub - JinanZou/Astock: Astock
    2. Year: 2022
    3. **Journal/Conference:** IJCAI 2022 (FinNLP)
    4. **Abstract:** Natural Language Processing(NLP) demonstrates a great potential to support financial decision-making by analyzing the text from social media or news outlets. In this work, we build a platform to study the NLP-aided stock auto-trading algorithms systematically. In contrast to the previous work, our platform is characterized by three features: (1) We provide financial news for each specific stock. (2) We provide various stock factors for each stock. (3) We evaluate performance from more financial-relevant metrics. Such a design allows us to develop and evaluate NLP-aided stock auto-trading algorithms in a more realistic setting. In addition to designing an evaluation platform and dataset collection, we also made a technical contribution by proposing a system to automatically learn a good feature representation from various input information. The key to our algorithm is a method called semantic role labeling Pooling (SRLP), which leverages Semantic Role Labeling (SRL) to create a compact representation of each news paragraph. Based on SRLP, we further incorporate other stock factors to make the final prediction. In addition, we propose a self-supervised learning strategy based on SRLP to enhance the out-of-distribution generalization performance of our system. Through our experimental study, we show that the proposed method achieves better performance and outperforms all the baselines' annualized rate of return as well as the

maximum drawdown of the CSI300 index and XIN9 index on real trading. Our Astock dataset and code are available at this https URL.

5. **Link:** https://arxiv.org/abs/2206.06606

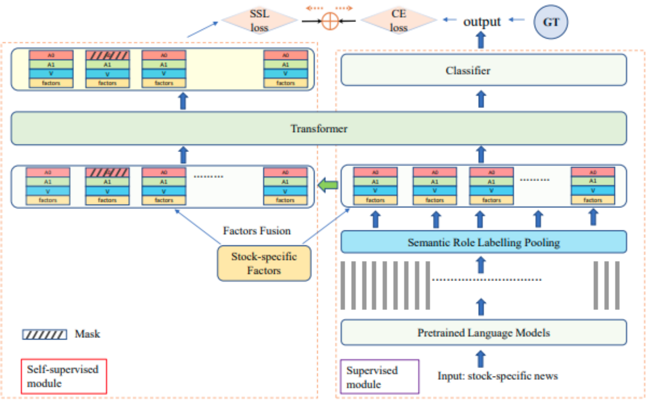


Figure 2: Overall framework of our approach, including a domain adapted pre-trained model (RoBERTa WWM Ext), Semantic Roles Pooling, transformer layer, self-supervised module (left part), and the supervised module (right part). The green arrow represents a duplicate for the SRLP. The final result is generated from the stock movement classifier, and the total loss is obtained from the self-supervised SRLP part and supervised stock movement classification part.