

International Natural Gas Price Trends Prediction with Historical Prices and Related News

Renchu Guan ¹, Aoqing Wang ¹, Yanchun Liang ^{1,2}, Jiasheng Fu ³ and Xiaosong Han ^{1,*}

- ¹ Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of National Education, College of Computer Science and Technology, Jilin University, Changchun 130012, China; guanrenchu@jlu.edu.cn (R.G.); wangaq20@mails.jlu.edu.cn (A.W.); ycliang@jlu.edu.cn (Y.L.)
- ² Zhuhai Laboratory of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Science and Technology, Zhuhai 519041, China
- ³ CNPC Engineering Technology R&D Company Limited, Beijing 102206, China; fujsdr@cnpc.com.cn
- * Correspondence: hanxiaosong@jlu.edu.cn

Abstract: Under the idea of low carbon economy, natural gas has drawn widely attention all over the world and becomes one of the fastest growing energies because of its clean, high calorific value, and environmental protection properties. However, policy and political factors, supply-demand relationship and hurricanes can cause the jump in natural gas prices volatility. To address this issue, a deep learning model based on oil and gas news is proposed to predict natural gas price trends in this paper. In this model, news text embedding is conducted by BERT-Base, Uncased on natural gas-related news. Attention model is adopted to balance the weight of the news vector. Meanwhile, corresponding natural gas price embedding is conducted by a BiLSTM module. The Attention-weighted news vectors and price embedding are the inputs of the fused network with transformer is built. BiLSTM is used to extract used price information related with news features. Transformer is employed to capture time series trend of mixed features. Finally, the network achieves an accuracy as 79%, and the performance is better than most traditional machine learning algorithms.

Keywords: natural gas; machine learning; price trend prediction



Citation: Guan, R.; Wang, A.; Liang, Y.; Fu, J.; Han, X. International Natural Gas Price Trends Prediction with Historical Prices and Related News. *Energies* **2022**, *15*, 3573. <https://doi.org/10.3390/en15103573>

Academic Editor: Fausto Cavallaro

Received: 4 April 2022

Accepted: 11 May 2022

Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, with the gradual consumption of fossil fuel reserves, the status of natural gas in energy supply continues to be improved. According to the report of the U.S. Energy Information Administration, as an important fuel, natural gas plays a really important role in almost all aspects of production process and living environment. In 2020, there were more than 800 billion cubic meters natural gas used in America, with 38% used for power generation, 33% used for industry and 15% used for civil [1]. Natural gas as a kind of main fuel is commonly used for winter heating in Europe and the United States. Compared with other fossil fuels, natural gas is a cleaner, more environment-friendly and more high-quality energy, almost free of sulfur, dust and other harmful substances, thus can fundamentally improve the quality of the environment [2]. In the past few years, in response to the policy of energy conservation and emission reduction, the rural areas of China are gradually using natural gas for large-scale supply in winter [3,4]. Meanwhile, natural gas is a raw material of many commercially organic chemicals.

There are many factors that affect the price of natural gas such as investor sentiment [5], exchange rates for precious metals [6], food price [7], even weather indices [8] and many other factors [9]. However, the influence between factors and commodity prices is uncertain. For example, there is a strong causal relationship between oil prices and two currency pairs, EUR/USD and GBP/USD [10]. Szturo and Candila found there is a lack of stable price relationship between crude oil and currency exchange rates in different periods [11,12].

Therefore, it is a very difficult task to predict the natural gas price accurately. Apparently, fluctuations in commodity prices are mainly due to changes in supplement and

demand. While natural gas is a stable commodity highly regulated by nations and governments, its changes in supply and demand are cyclical in a stable international environment. And the large fluctuations in futures prices are usually due to major international events. It is also proved by [13] that the event shock has a strong impact on the price of crude oil using dummy variable events. However, most existing analysis methods in the economic field only consider few relative factors that influence the prices, such as the stock market index, prices in related industries and government economic reports. Although these indices can clearly indicate some current states of the economy, they are also influenced by the events. Thus, the index of them appears more sluggish. Additionally, the economic index and policy focus more on relevant events in their field. Therefore, the implicit factor in relative news is adopted to measure the multiple and quick-responded impact on the price of natural gas. Natural gas news on natural gas analysis websites contains professional comments or certain events influencing natural gas prices, which can be reflected in economic, political, natural and even human domains.

2. Related Work

When facing the analyses of the price trends of financial products, it is natural to draw an inductive analogy using previous data. In addition, this process of inductive analogies using past periods can be modeled as a time series problem. In the machine learning domain, there are two main methods for the time series problems, statistical-based time series methods and machine-learning methods. Statistical-based methods use some mathematical methods to analyze and predict time-series data such as Difference Methods, Moving Averages and Fourier Decomposition [14,15]. However, when facing unbalanced or biased data, machine learning methods can exploit more complex relationships in the temporal dimensions, such as recurrent neural network (RNN), convolutional neural network (CNN) and attention-based models. RNN uses hidden recurrent layers to transfer past memory. To get more effective past information, some advanced RNN models such as LSTM [16] and GRU [17] use several special methods such as gates to control memory reserve, reset, forget or output. Bidirectional RNN models such as Bidirectional LSTM [18] capture more information from both sides of time series with a two-layer RNN to obtain more stable memory. However, to get recurrent units, the RNN-based models must iterate parameters serially. In order to make full use of computing resources, some scholars proposed temporal convolutional network (TCN) [19] and attention-based models. TCN adopts dilated causal convolution with different convolution kernels. Attention-based models use various attention to integrate the correlations among data in different time series.

Natural language processing is always the focus of machine learning. The text classification methods can be roughly divided into two categories, traditional machine learning methods and feature engineering methods. Traditional machine learning decomposes the text classification task into the two-phase mission, feature engineering and classification. Feature engineering methods mainly contain term frequency-inverse document frequency (TF-IDF), bag-of-words (BOW) and topic models, which can extract the features of the words in the article. After the presentation of content, most traditional machine learning classification models can be used for classification, such as Support Vector Machines (SVM) [20–23], Multi-Layer Perceptron (MLP) [24] and Logistic Regression (LR) [25,26].

With the rapid development of deep learning, text classification is becoming a hotspot. There are some ways extended to solve text classification problems. FastText [27] is a text classification tool that is obviously faster than deep networks in training time. Bidirectional Encoder Representation from Transformers (BERT) [28] is a pre-trained language representation model, which generates representation by a new masked language model (MLM) [29]. It obtained new state-of-the-art results in 11 NLP tasks staggeringly.

The work on economic prediction is more and more in focus. Puka et al. reformulated crude oil prices from a regression problem to a classification problem using neural networks to effectively hedge the price rise risk of West Texas Intermediate (WTI) crude oil [30]. Mouchtaris et al. used a bagging ensemble decision tree to predict gas prices

using SVM with different kernel functions as subtrees [31]. Manowska et al. analyzed Poland's natural gas reserves and proposed a model combining ARIMA with an LSTM artificial neural network for forecasting its consumption, which takes into account historical consumption, energy prices, and Poland's energy policy and proves the effectiveness of the built model [32]. Hu and Zhenda et al. combined Ensemble Empirical Mode Decomposition with Adaptive Noise and LSTM method to analyze the relationship between news sentiment index and West Texas Intermediate (WTI), which surpassed other prediction models in multiple statistical indicators [33]. Pinto et al. integrated the KEA [34] algorithm to extract key phrases from news articles for predicting the closing price of a given trading day with a neural network trained on the extracted key phrases and stock quotes [35]. Wu, Binrong et al. proposed a method to use CNN to extract the information from news headlines and combined it with Google Trends to predict crude oil and obtain a less error rate [36]. A WT-FNN model was built in [37] which tracked and predicted the crude oil prices with a dynamic weight to different-period information. Li et al. implemented a merged model to predict natural gas price, which uses natural gas price and everyday document-level news presentations by sum-pooling Word2Vec words' matrix. The model derives a CNN-LSTM module and an LSTM module which can extract temporal content and price trend, respectively [38]. There is also a method to use sentiment analysis of news to predict gas prices. Guifeng Wang et al. proposed a copula-based contrastive coding method, which can use the dependence between the stock and corresponding economic factors [39].

As a matter of fact, it is a difficult problem to consider news and price trends. When processing news data, there is obviously more than one piece of daily news. The traditional methods for text classification are not feasible. They only consider the semantic relationship of the context in the text, while ignoring the semantic relationship between different texts at the same level. At the same time, the way to embed news text is an essential problem that is the decisive factor extracting vital information related to natural gas prices. Meanwhile, the time lag selection is also important which is conducted and analyzed in Section 4.3.

To address the problem above, we conduct a novel model which can combine daily news with a natural gas price. The model is able to predict future gas prices and achieve an accuracy over 70%. The general overflow of the model is shown in Figure 1. The main idea of the article is as follows.

The main work is divided into three parts, the first part is the data acquisition and normal preprocessing of news text and natural gas prices. The second part is processing text data as the input format of BERT. In addition, the third part is using neural networks for trend prediction of natural gas.

The rest of the paper is organized as follows. Section 3 introduces materials and methods. In Section 3.1, we show the news data and the natural gas prices. In Section 3.3, the main ideas of the paper and the individual components of the model are illustrated. Section 4 shows the results. We conclude our work in the Section 5.

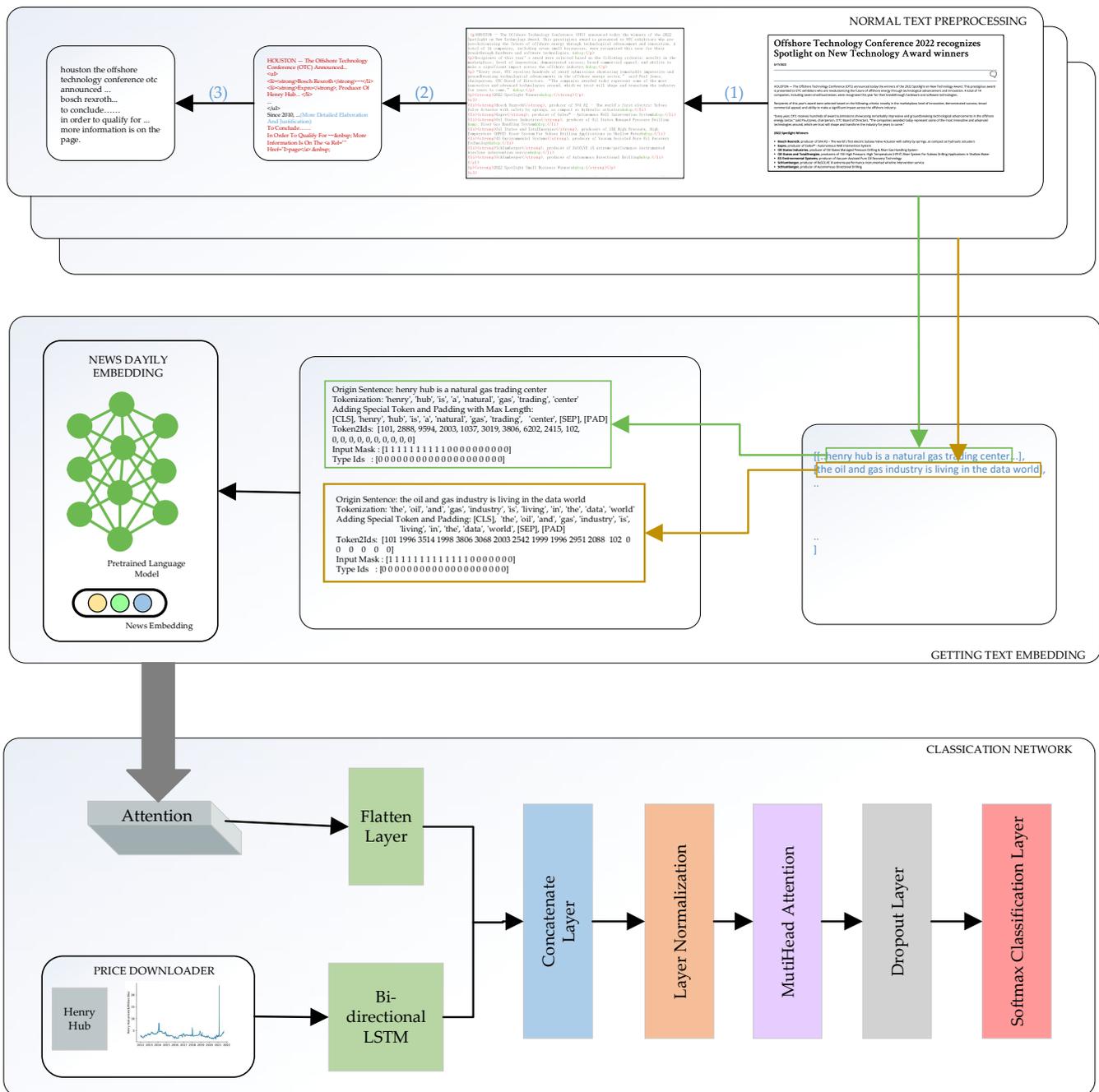


Figure 1. An overall pipeline of our work. There are three steps to standardize the text. Step (1) is to display the codes behind the web pages. Step (2) is to parse information through a spider. Useless information is removed in Step (3), including releasing redundant HTML tags, lowercasing all letters, removing the punctuation marks, intercepting part of text, etc. The details are described in Section 3.2.

3. Materials and Methods

3.1. Dataset

The dataset mainly includes two parts, news data and natural gas price data. We get these data from two representative websites, WorldOil and Henry Hub. The dataset statistics are shown as Table 1 from 1 January 2012 to 3 September 2021.

Table 1. Statistical information from Henry Hub natural gas price and WorilOil news text.

Price Statistics				News Statistics			Word Statistics		
Raise.	Decline.	Unch.	Off.	Avg.	Min.	Max.	Avg.	Min.	Max.
1046	1022	388	778	8.885	1	33	352.2	5	5495

The Price Statistics means the day counts when natural gas price raised, declined, unchanged or was off. Columns of news statics mean the average, minimum and maximum news counts in a day. Columns of word statics enumerate the average, minimum and maximum word counts in a news.

The news data are from World Oil Daily. Then, we introduce the formulation of the news data. World Oil Magazine (WorldOil) is a professional crude oil portal website, with in-depth event analysis and comments, and industry chain news recognized by professional investors. World Oil Trading Company specialized in petroleum service, introducing any service to support drilling, exploring and production. As an upstream trusted source of forecast data, industry trends, and insights, World Oil produces readers with reliable news data that can reflect the price of natural gas. Scrapy is used to crawl the WorldOil website information. There is only some recent news data on the WorldOil website that provides for downloading. Therefore, we search several keywords related to natural gas on the search page of the website and download them. Then, we merge two-part and remove the duplicate articles.

The gas price is downloaded from Henry Hub. As the natural gas pricing center of the New York Commodity Exchange Incorporation (NYMEX), Henry Hub can manage the whole process of natural gas. Due to the large transaction volume and transparent price of Henry Hub, some traditional natural gas producers such as Qatar, Australia and Mozambique are decoupled from crude oil prices. The exchange incorporation in North America established a natural gas delivery pricing mechanism based on Henry Hub prices.

3.2. Materials Preprocessing

This section mainly focuses on the preprocessing of gas natural news. First of all, we need to select the crawled webpages. During the crawling process, part of the text that is not news is also crawled as part of the news data. Regular expressions are used to exclude additional pages that are not news. Meanwhile, non-text information such as (HTML labels and advertisements is removed.

We need to process the news data to the input shape of the news model. Firstly, the redundant and irrelevant text needs to be removed to release useless information. For example, HMTL tags on the Internet indicates a formatting or emphasizing information, such as "a", "strong", "ul", "li", etc. Mixed advertisements and link information are killed by regular expressions.

Since the BERT-Base, Uncased model is used, which only recognizes lowercase characters. All alphabetic characters are transformed into lowercase. The word input size in the original BERT is no more than 512 tokens. The punctuations such as the comma are removed which also occupies a position in the representation vector. Meanwhile, origin BERT requires the input tokens no more than 512. Usually, people put the key ideas that they really want to express at the beginning of the article or the end of the article. Therefore, in order to obtain as much semantic information as possible, we use a normal truncation method—head-tail [40]. We intercept 200 words at the beginning and 200 words at end of the article as the real corpus of the text, which make sure the total word counts less than 512. BERT model takes every word as a token. As for stop words and frequency words, these are only indispensable contextual information [41]. They are necessary to be reserved. After the word segmentation and deduplication, we perform to obtain the longest text length. Then the mentioned news data are tokenized, with zero extended to the longest length. Finally, we add <CLS>, <SEP> as the beginning and end of the articles. BERT is a more complicated multi-layer neural network. Sun, Chi et al. compared the output from

every layer and the pooling results from the last layer output. In addition, the pooling results got better scores. In this way, we use the pooling whole text embedding as the news vector instead of single word embedding.

All the news vectors are queued into the 3D matrix (day_length, max_news_length, max_word_length), where max_news_length is the maximum number of news articles per day, padded with zero matrices when needed.

As for natural gas, the price is downloaded from Henry Hub. In addition, we categorize the price trend of natural gas into three situations by calculating the difference between the open prices of two consecutive days: raising and declining and unchanged. The frequency when unchanged occupies is very low. Therefore, we simply classify the unchanging situation as declining. In this way, predicting the price trend of natural gas is simplified to a binary classification problem.

3.3. Summary of Methods

This section follows the order of our model’s construction: price module, news module, and fusion prediction module of price feature and news presentation.

3.3.1. Price Module

Price Module contains an LSTM model to make use of temporal features of price time series natural gas price. We use BiLSTM to memory the price trend, as Figure 2.

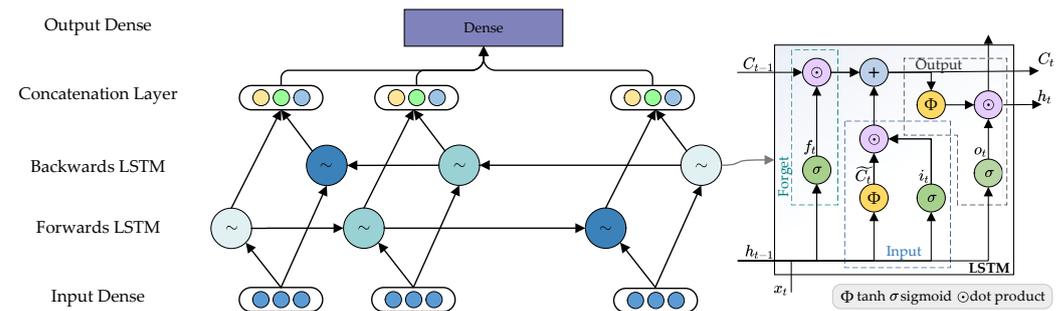


Figure 2. BiLSTM module for natural gas price.

According to the following equations, LSTM [16] transfers two states: h_t and C_t . They can capture more information from the dataset. When data x_t comes, o_t learns what to forget. f_t decides which memory in past is continually flowing in the recurrent hidden parameters.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{2}$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_c), \tag{3}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{4}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{5}$$

$$h_t = o_t \odot \tanh(C_t). \tag{6}$$

In order to prevent forgetting important information, we use bi-directional Long Short-Term Memory layers to enhance the memory. BiLSTM [18] contains two single LSTM networks, which concatenates the hidden embedding of time series data in two directions.

$$LSTM_{output} = \sigma(W'h_t), \tag{7}$$

$$BiLSTM_{output} = \sigma(W_{BiLSTM}[LSTM_{Forwards}, LSTM_{Backwards}] + b_{BiLSTM}). \tag{8}$$

3.3.2. News Module

After news from the website is processed to a clean token, the token is sent to a BERT-Base, Uncased model. BERT-Base model implements a 12-layer bidirectional transformer encoder. It masks parts of words (15%) and predicts them by all the rest of words, omnidirectionally. Secondly, it uses transformers to classify consecutive sentence pairs. Through these two steps of training, the general language representation vectors of every word can be output. We take the matrix that word <CLS> outputs as the representation of the whole news.

However, as a matter of fact, there is more than one piece of daily news. The traditional method of using an article for text classification is not feasible. In order to reflect the importance of different articles, we adopt a self-attention to fuse the multi-news matrix in a day, as Figure 3.

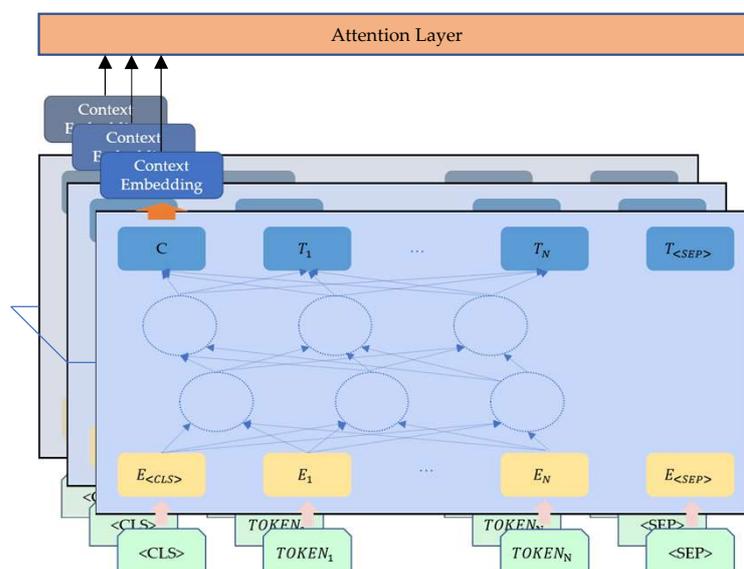


Figure 3. Attentive news module.

Firstly, we multiply the daily news BERT matrices by different weight matrices to obtain the *Query* vectors, *Key* vector and *Value* vectors. Secondly, we calculate a *similarity* between a news *Query* vector and other news *Key* vectors. The *similarity* matrices are multiplied by the corresponding news *Value* vectors. Then, we get the final attentive representation vector of news. We take the scaled dot product normalized by softmax.

$$Attention = \sum_{i=1}^{L_{news}} similarity(Query, Key_i) * Value_i \quad (9)$$

3.3.3. Merged Module

For simplicity, we merge the represented matrices of everyday news and price as the final matrix of features.

We apply a transformer to catch further dependence of the merged model. A multi-Head Attention is adopted as Figure 4. The detailed model parameters are shown in the Table A1.

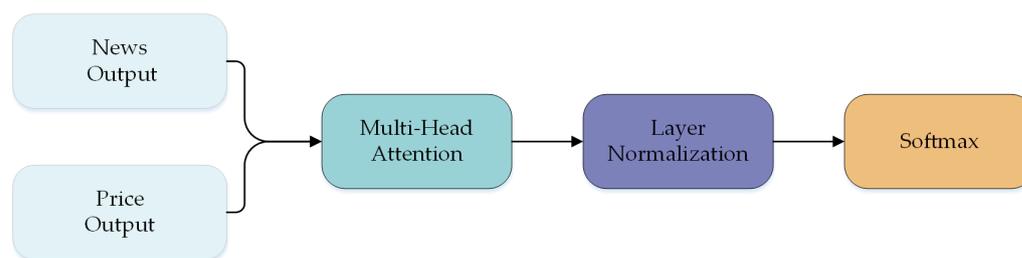


Figure 4. Merged module.

3.4. Evaluation of Results

We use the accuracy, precision, recall and f1-score evaluation to evaluate the effect on the model. *Accuracy*, *Recall* and *F1-Score* can be calculated by following formula,

$$Accuracy = (TP + FN) / (TP + FN + FP + TN) \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1-Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (13)$$

where, *TP*, *TN*, *FP*, *FN* present the number of true positives, true negatives, false negatives and false negatives.

4. Results

4.1. Experimental Setup

All the experiments in this article are conducted under the Linux system environment on one single Tesla 100 GPU with 32 GB memory. Tensorflow2.0's [42] KERAS is used to construct the whole model. The model is compiled with Adam optimizer [43] with a dropout rate of 0.5. Detailed parameter analysis is introduced in Section 4.5. In addition, we use binary cross-entropy as the loss function. Almost all the weighted reLu [44] hyper-parameters are initialized by initializers with Lecun Normal [45] and adjusted on the training stage.

4.2. Comparison Benchmarks

We choose the following models for comparison. The comparative experiments consist of three parts: prediction only with price, prediction only with text, and aggregate textual information and price information. All experiments are conducted with a price lag of 60 days. The News text is presented by adding all word embedding from BERT-Base mentioned in Section 3.3.2 as an additional Document-Level context presentation. The output is to predict the next day's price trends. We construct the following baselines with some tricks.

1. Price Prediction Measure

- **ARIMA:** Autoregressive integrated moving average model. It combines three methods: auto-regression, I-for integrated, moving average. Non-stationary data are made stable by subtracting y to achieve the purpose of accurate prediction. ARIMA is often used for regression. We get the classification result by subtracting ARIMA regression result and real value.

2. Text Prediction Measure

- **FASTTEXT:** It superimposes and averages the word embedding presented by the word-level n-gram algorithm to get the document's vector. Then, a hierarchical SoftMax is used for multi-classification.

3. News-Price Prediction Measure

- LSTM: Long Short-Term Memory is an advanced RNN that can store more information.
- TCN: Temporal Convolutional Network uses dilated causal convolution and padding layers to catch temporal series.
- k-NN: k-Nearest Neighbor. The method uses Euclidean Distance to measure similarity in different time series by mapping each point in the time series as a feature. In order to achieve better scores, we range k from 1 to 31 and take the best accuracy as the final result.
- RANDFOREST: Random Forest classifier. The random forest uses 100 decision tree classifiers to estimate the accuracy of dataset, while the whole dataset is used to build each tree.

4.3. Results

As mentioned above, natural gas price classification is a challenging task influenced by large amounts of potential factors. Generally speaking, it is efficient to guide people in the selection and purchase of financial products when the accuracy comes to 60%. The model gets 79% accuracy after 100 epochs of training. Especially, when using uncertain news data to fetch the factor, the accuracy over 70% of our work is a satisfying result.

A small batch size is used to train the model. Meanwhile, the Lecun Normal initialization and l2 regularization are used to accelerate the convergence of the model and reduce overfitting. So, the effect of the model quickly reaches the best and starts to oscillate. We have experimented with 1000 epochs. However, after 100 epochs, the model begins overfitting with declining accuracy and rising loss on validation data. Actually, after 80 epochs, the validation results trend to be overfitting. In this way, we present the result in 100 epochs. And the accuracy and loss of the training set and the validation set during the training process is shown in Figure 5.

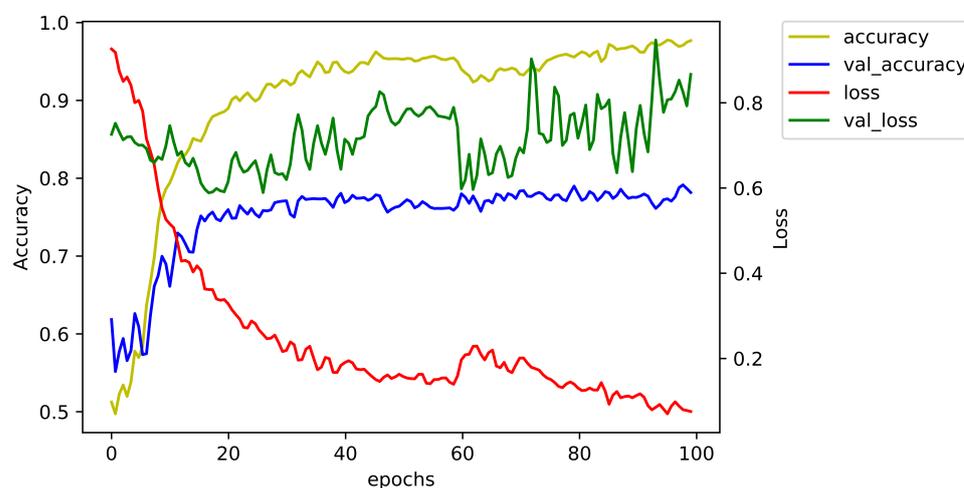


Figure 5. Accuracy and loss in 100 training epochs.

Figure 5 shows that the validation set tends to become large during the training process due to different details of distributions in the training set and the validation set. The training set has 626 positive values and 843 negative values with a positive ratio of 42.6%. In addition, the validation set has 191 positive values and 299 negative values with a positive ratio of 39.0%. However, the overall trend of economical data is often with fewer changes. So, the model gets better accuracy results.

The detailed performance of other baselines is shown in Table 2. LSTM model using combined data achieves the best score in accuracy except for our model. Our model

outperforms over 123.7% with 79.24% in accuracy. TCN’s results are likely to predict all the time series cases to one single class. There are three major reasons that our model is better:

1. BERT model provides an informative embedding of whole news content.
2. Attention in text module catches better hierarchical structure and semantic information.
3. Progressive temporal message is passed to the network through LSTM and transformer in different levels.

Table 2. Results compared with benchmarks. The suffix PRICE in the First column means that input only contains price The suffix TEXT means that the input only contains Text information, and no suffix means that the combination of the two is used.

	Accuracy	Precision	F1-Score	Recall
TCN_PRICE	0.6099	0.5088	0.1593	0.2427
TCN_TEXT	0.6358	0.5333	0.5714	0.5517
TCN	0.5194	0.4255	0.6429	0.5120
KNN_PRICE	0.5151	0.3675	0.3370	0.3516
KNN_TEXT	0.5302	0.3943	0.3812	0.3876
KNN	0.5043	0.3828	0.4420	0.4103
LSTM_PRICE	0.4418	0.3746	0.6319	0.4703
LSTM_TEXT	0.5690	0.4625	0.6099	0.5261
LSTM	0.5453	0.3876	0.2747	0.3215
ARIMA_PRICE	0.5558	0.0	0.0	0.0
Fasttext_TEXT	0.5690	0.5000	0.0005	0.0010
RANDFOREST_PRICE	0.5431	0.3176	0.1492	0.2030
RANDFOREST_TEXT	0.5259	0.3946	0.4033	0.3989
RANDFOREST	0.5259	0.3934	0.3978	0.3956
Our Model	0.7978	0.8483	0.6373	0.7278

4.4. Model Component Ablation Study

We also conduct an ablation study to examine different components’ functions. We remove the transformer structure in the merge module, BiLSTM structure and text attention in News Module. In order to be comparable to our model, prices in past 60-day are used to substitute the BiLSTM structure. Instead of text attention, we flatten the news in a day and transmit it to a single dense. As Table 3, all of our structures improve the accuracy of natural gas prices.

Table 3. Accuracy in component ablation study.

	Validation Data	Test Data
w/o text attention	0.7061	0.7367
w/o text attention & transformer	0.6918	0.7306
w/o text attention & transformer & BiLSTM	0.6469	0.6102
Our Model	0.8126	0.7924

4.5. Parameter Analysis: Probing Sensitivity Evaluation of Results

In this section, we provide a detailed discussion on how the temporal module in different levels affects model performance. As shown in Figure 6, as the time lag in layers increases, the performance of our model’s accuracy goes through a process of rising and decreasing. The finding obviously coincides with our experience. A proper lag fits the network well. If time lag comes too long, detailed information will be neglected. Seasonal dependence cannot be caught with a much detailed time lag. In this way, a more balanced time lag should be chosen to average the impact of the detailed information and the long-term trend with an LSTM lag of two weeks and a transformer lag of one month.

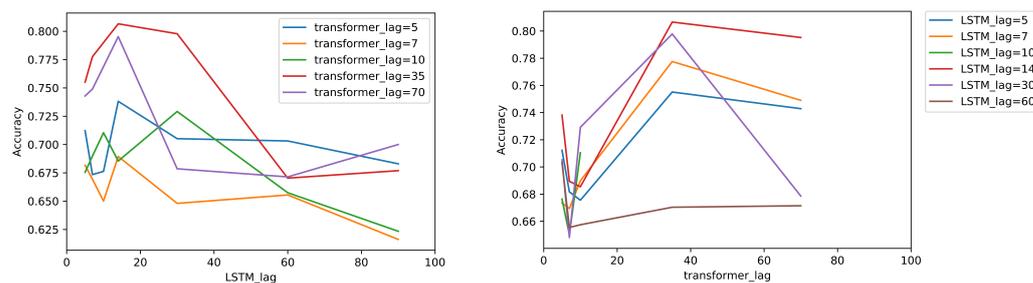


Figure 6. Parameter analysis results in different lags. The figure on the left shows LSTM lag windows' influence on the accuracy. In addition, the right figure shows transformer lags' impact on the accuracy.

5. Conclusions

In this paper, we build a news-price merge model to classify the trend of natural gas prices. Due to the large amounts of factors that influence natural gas prices, we consider the latent events from news. Meanwhile, the historical gas prices are used to capture former information from hidden states. The strengths of our model are as follows.

1. We combine the related news and historical prices of natural gas prices to obtain event and trends information in our model.
2. We use a BERT-attention model for text embedding to retrieve more semantic information. BERT model generates an informative embedding of news content in document level. Text attention offers a weighted matrix given to multiple articles on the same day in multi-document level. In this way, the everyday news embedding is of better hierarchical structure and semantic information.
3. We use two time-related modules. It is important to master the temporal scale. We can abandon useless messages and keep vital information from the past. Meanwhile, we can gain more recent situations and further trend in different time modules.

Due to the reason mentioned above, we adopt a text-price model in different time views. Finally, we present results with several baselines and get an accuracy of over 79%.

However, there are still some deficiencies in the article. In the text representation stage, we roughly use the basic BERT-uncased model to get document embedding. In the future, we will try other methods to view different levels in documents. Meanwhile, the alignment on the news daily information and historical prices is not satisfied. More refined strategies will be used in future works. At present, our model focus on short-term prediction, and we will extend it to a long-term prediction.

Author Contributions: Conceptualization, R.G., Y.L. and X.H.; Data curation, A.W.; Formal analysis, R.G. and Y.L.; Funding acquisition, R.G., Y.L., J.F. and X.H.; Investigation, A.W. and J.F.; Methodology, Y.L. and X.H.; Project administration, X.H.; Resources, X.H.; Software, A.W. and X.H.; Supervision, R.G. and J.F.; Validation, A.W. and X.H.; Visualization, A.W. and X.H.; Writing—original draft, R.G. and Y.L.; Writing—review & editing, Y.L. and X.H. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful for the support of the National Key Research and Development Program of China (2021YFF1201203, 2021YFF1201205), the National Natural Science Foundation of China (61972174 and 62172187), the Science and Technology Planning Project of Guangdong Province (2020A0505100018), Guangdong Universities' Innovation Team Project (2021KCXTD015) and Guangdong Key Disciplines Project (2021ZDJS138).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are from the henry hub and worldoil websites. Gas price is downloaded at <https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm> (accessed on 4 April 2022). In addition, news is crawled at <https://www.worldoil.com> (accessed on 4 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RNN	Recurrent Neural Network
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
TCN	Temporal Convolutional Network
TF-IDF	term frequency-inverse document frequency
BOW	Bag-of-Words
SVM	Support Vector Machines
MLP	Multi-Layer Perceptron
LR	Logistic Regression
WTI	West Texas Intermediate
ARIMA	Autoregressive Integrated Moving Average model
FNN	Fuzzy Neural Network
NYMEX	New York Commodity Exchange Incorporation
HMTL	HyperText Markup Language
BERT	Bidirectional Encoder Representation from Transformers
GPU	Graphics Processing Unit

Appendix A

With different lags the network output dim may be different. With the price lag of 30 and transformer lag of 35, the batch_size is set as 1. The network parameters are shown as Table A1. Total number of parameters in this network is 632,290. All of the parameters are trained.

Table A1. Detailed network parameters.

Network Component Name	Output Dim
Price Input	(1, 35, 30)
Price Reshape	(1, 35, 30, 1)
LSTM1	(35, 30, 32)
LSTM2	(35, 30, 32)
LSTM Bidirectional	(35, 30, 64)
LSTM Concatenation	(35, 1920)
Text Price	(1, 35, 33, 512)
Head _{Num} × Text Attention	(35, 33, 512)
Head _{Num} × Text Attention Flatten	(35, 16,896)
Text Dropout	(35, 16,896)
Text Expand Dim	(1, 35, 16,896)
Text Concatenation	(1, 35, 16,896)
Text Price Concatenation	(1, 35, 18,816)
TransformerHead _{Num} × Text Price Dense	(1, 35, 32)
TransformerHead _{Num} × Transformer Attention	(1, 35, 32)
TransformerHead _{Num} × Layer Normalization1	(1, 35, 32)
TransformerHead _{Num} × Transformer Dense	(1, 35, 32)
TransformerHead _{Num} × Transformer Dropout	(1, 35, 32)
TransformerHead _{Num} × Layer Normalization2	(1, 35, 32)
Overall Dense	(1, 35, 128)
Overall Dropout	(1, 35, 128)
SoftMax Layer	(1, 35, 2)

References

1. U.S. Energy Information Administration (EIA). Total Energy Monthly Data. Available online: <https://www.eia.gov/totalenergy/data/monthly/index.php> (accessed on 4 April 2022).
2. U.S. Energy Information Administration (EIA). Natural Gas and the Environment. Available online: <https://www.eia.gov/energyexplained/natural-gas/natural-gas-and-the-environment.php> (accessed on 4 April 2022).
3. Zhang, D.; Paltsev, S. The future of natural gas in China: Effects of pricing reform and climate policy. *Clim. Chang. Econ.* **2016**, *7*, 1650012. [[CrossRef](#)]
4. Herberg, M.E. *Asia's Uncertain LNG Future*; The National Bureau of Asian Research: Seattle, WA, USA; Washington, DC, USA, 2013.
5. Li, Z.; Huang, Z.; Failler, P. Dynamic Correlation between Crude Oil Price and Investor Sentiment in China: Heterogeneous and Asymmetric Effect. *Energies* **2022**, *15*, 687. [[CrossRef](#)]
6. Gupta, R.; Pierdzioch, C.; Wong, W.K. A Note on Forecasting the Historical Realized Variance of Oil-Price Movements: The Role of Gold-to-Silver and Gold-to-Platinum Price Ratios. *Energies* **2021**, *14*, 6775. [[CrossRef](#)]
7. Kirikkaleli, D.; Darbaz, I. The Causal Linkage between Energy Price and Food Price. *Energies* **2021**, *14*, 4182. [[CrossRef](#)]
8. Tarczyński, W.; Mentel, U.; Mentel, G.; Shahzad, U. The Influence of Investors' Mood on the Stock Prices: Evidence from Energy Firms in Warsaw Stock Exchange, Poland. *Energies* **2021**, *14*, 7396. [[CrossRef](#)]
9. Nuryyev, G.; Korol, T.; Tetin, I. Hold-Up Problems in International Gas Trade: A Case Study. *Energies* **2021**, *14*, 4984. [[CrossRef](#)]
10. Orzeszko, W. Nonlinear Causality between Crude Oil Prices and Exchange Rates: Evidence and Forecasting. *Energies* **2021**, *14*, 6043. [[CrossRef](#)]
11. Szturo, M.; Włodarczyk, B.; Miciuła, I.; Szturo, K. The Essence of Relationships between the Crude Oil Market and Foreign Currencies Market Based on a Study of Key Currencies. *Energies* **2021**, *14*, 7978. [[CrossRef](#)]
12. Candila, V.; Maximov, D.; Mikhaylov, A.; Moiseev, N.; Senjyu, T.; Tryndina, N. On the Relationship between Oil and Exchange Rates of Oil-Exporting and Oil-Importing Countries: From the Great Recession Period to the COVID-19 Era. *Energies* **2021**, *14*, 8046. [[CrossRef](#)]
13. Peng, J.; Li, Z.; Drakeford, B.M. Dynamic Characteristics of Crude Oil Price Fluctuation—From the Perspective of Crude Oil Price Influence Mechanism. *Energies* **2020**, *13*, 4465. [[CrossRef](#)]
14. Chatfield, C. The Holt-winters forecasting procedure. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1978**, *27*, 264–279. [[CrossRef](#)]
15. Heschel, A.J. *The Prophets*; Harper Torchbooks: New York, NY, USA, 1962; Volume 2.
16. Graves, A. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45. [[CrossRef](#)]
17. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS 2014 Workshop on Deep Learning, Montreal, QC, Canada, 12–13 December 2014.
18. Graves, A.; Mohamed, A.; Hinton, G.E. Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Hoboken, NJ, USA, 2013; pp. 6645–6649. [[CrossRef](#)]
19. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
20. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
21. Smola, A.J.; Schölkopf, B. On a Kernel-Based Method for Pattern Recognition, Regression, Approximation, and Operator Inversion. *Algorithmica* **1998**, *22*, 211–231. [[CrossRef](#)]
22. Chen, Z.; Cao, S.; Mao, Z. Remaining useful life estimation of aircraft engines using a modified similarity and supporting vector machine (SVM) approach. *Energies* **2018**, *11*, 28. [[CrossRef](#)]
23. Zhou, Y.; Wu, J.; Yu, Z.; Ji, L.; Hao, L. A hierarchical method for transient stability prediction of power systems using the confidence of a SVM-based ensemble classifier. *Energies* **2016**, *9*, 778. [[CrossRef](#)]
24. Mendonça de Paiva, G.; Pires Pimentel, S.; Pinheiro Alvarenga, B.; Gonçalves Marra, E.; Mussetta, M.; Leva, S. Multiple site intraday solar irradiance forecasting by machine learning algorithms: MGGP and MLP neural networks. *Energies* **2020**, *13*, 3005. [[CrossRef](#)]
25. Wang, F.; Yu, Y.; Wang, X.; Ren, H.; Shafie-Khah, M.; Catal ao, J.P. Residential electricity consumption level impact factor analysis based on wrapper feature selection and multinomial logistic regression. *Energies* **2018**, *11*, 1180. [[CrossRef](#)]
26. Manoharan, H.; Teekaraman, Y.; Kirpichnikova, I.; Kuppusamy, R.; Nikolovski, S.; Baghaee, H.R. Smart Grid Monitoring by Wireless Sensors Using Binary Logistic Regression. *Energies* **2020**, *13*, 3974. [[CrossRef](#)]
27. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, 3–7 April 2017; Volume 2: Short Papers; Lapata, M., Blunsom, P., Koller, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 427–431. [[CrossRef](#)]

28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 5998–6008.
29. Salazar, J.; Liang, D.; Nguyen, T.Q.; Kirchoff, K. Masked Language Model Scoring. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 5–10 July 2020*; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 2699–2712. [[CrossRef](#)]
30. Puka, R.; Łamasz, B.; Michalski, M. Effectiveness of Artificial Neural Networks in Hedging against WTI Crude Oil Price Risk. *Energies* **2021**, *14*, 3308. [[CrossRef](#)]
31. Mouchtaris, D.; Sofianos, E.; Gogas, P.; Papadimitriou, T. Forecasting Natural Gas Spot Prices with Machine Learning. *Energies* **2021**, *14*, 5782. [[CrossRef](#)]
32. Manowska, A.; Rybak, A.; Dylong, A.; Pielot, J. Forecasting of Natural Gas Consumption in Poland Based on ARIMA-LSTM Hybrid Model. *Energies* **2021**, *14*, 8597. [[CrossRef](#)]
33. Hu, Z. Crude oil price prediction using CEEMDAN and LSTM-attention with news sentiment index. *Oil Gas Sci. Technol.-D'IFP Energies Nouv.* **2021**, *76*, 28. [[CrossRef](#)]
34. Witten, I.H.; Paynter, G.W.; Frank, E.; Gutwin, C.; Nevill-Manning, C.G. Kea: Practical automated keyphrase extraction. In *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*; IGI Global: Hershey, PA, USA, 2005; pp. 129–152.
35. Pinto, M.V.; Asnani, K. Stock price prediction using quotes and financial news. *Int. J. Soft Comput. Eng.* **2011**, *1*, 266–269.
36. Wu, B.; Wang, L.; Lv, S.X.; Zeng, Y.R. Effective crude oil price forecasting using new text-based and big-data-driven model. *Measurement* **2021**, *168*, 108468. [[CrossRef](#)]
37. Wang, D.; Fang, T. Forecasting Crude Oil Prices with a WT-FNN Model. *Energies* **2022**, *15*, 1955. [[CrossRef](#)]
38. Li, T.; Han, X.; Wang, A.; Li, H.; Liu, G.; Pei, Y. News-Based Research on Forecast of International Natural Gas Price Trend. In *Fuzzy Systems and Data Mining VI—Proceedings of FSDM 2020, Virtual Event, 13–16 November 2020*; Tallón-Ballesteros, A.J., Ed.; *Frontiers in Artificial Intelligence and Applications/IOS Press: Amsterdam, The Netherlands, 2020*; Volume 331, pp. 194–200. [[CrossRef](#)]
39. Wang, G.; Cao, L.; Zhao, H.; Liu, Q.; Chen, E. Coupling Macro-Sector-Micro Financial Indicators for Learning Stock Representations with Less Uncertainty. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, 2–9 February 2021*; AAAI Press: Palo Alto, CA, USA, 2021; pp. 4418–4426.
40. Sun, C.; Qiu, X.; Xu, Y.; Huang, X. How to fine-tune bert for text classification? In *Proceedings of the China National Conference on Chinese Computational Linguistics, Kunming, China, 18–20 October 2019*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 194–206.
41. Khattab, O.; Zaharia, M. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, 25–30 July 2020*; pp. 39–48.
42. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR* **2016**, arXiv:1603.04467.
43. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015*; Conference Track Proceedings; Bengio, Y., LeCun, Y., Eds.; OpenReview.net; 2015.
44. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, FL, USA, 11–13 April 2011*; Gordon, G.J., Dunson, D.B., Dudík, M., Eds.; JMLR.org: Cambridge, MA, USA, 2011; Volume 15, pp. 315–323.
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proceedings of the 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, 7–13 December 2015*; IEEE Computer Society: Washington, DC, USA, 2015; pp. 1026–1034. [[CrossRef](#)]