# BERT-BiLSTM-Attention model for sentiment analysis on Chinese stock reviews

**Xiaoyan Li**
Huainan Normal University
**Lei Chen**
leichen@hnnu.edu.cn
Huainan Normal University
**Baoguo Chen**
Huainan Normal University
**Xianlei Ge**
Huainan Normal University

**Research Article**

**Keywords:** Stock, Sentiment analysis, BERT, BiLSTM, Attention

**Posted Date:** March 12th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-4023113/v1



**Additional Declarations:** No competing interests reported.

# BERT-BiLSTM-Attention model for sentiment analysis on Chinese stock reviews

Xiaoyan Li[1], Lei Chen[1,*], Baoguo Chen[1] and Xianlei Ge[2]

[1] Department of Computer, Huainan Normal University, Huainan,232000, China; lxy0722_321@163.com; bgchen0706@163.com

[2] Department of Electronic Engineering, Huainan Normal University, Huainan,232000, China; gexianlei@hnnu.edu.cn

[*] Correspondence: leichen@hnnu.edu.cn

## Abstract

COVID-19 has produced significant fluctuations and impacts on the Chinese stock market, and the sentiment analysis of stock reviews is important for the study of economic recovery. Due to the lack of a large amount of labeled data in the existing Chinese stock review data, and the currently popular Bert model mostly failed to consider contextual information according to different contextual backgrounds when extracting features, resulting in the lack of contextual information in the modeled features. To address the above problems, this paper proposes an innovative Chinese stock review sentiment analysis model BERT-BiLSTM-Attention, which encodes the stock review text by BERT to enhance the semantic feature representation of the text, BiLSTM is then utilized to enhance the contextual information of the overall context of the review as well as the model's comprehension of the text sequences, and then Attention mechanism is utilized to obtain important textual information and get the most effective information quickly. Experiments show that the model is effective in sentiment analysis of Chinese stock reviews, with an accuracy of 93.98%, It can be proved that the proposed model well enhances the performance of stock review text classification, and has a strong generalization ability, which can be used for sentiment analysis in many fields.

**Keywords:** Stock; Sentiment analysis; BERT; BiLSTM; Attention

## 1 Introduction

Sentiment analysis is an important application in the field of Artificial Intelligence, it involves identifying and categorizing the sentiments and opinions implicit in a text, using Natural Language Processing methods to identify, extract, quantify, and study a given text, with the aim of finding out the polarity of the text and categorizing it as positive, neutral, or negative, a process also known as textual context mining [1]. With the development of social media, it has become increasingly popular for people to express their sentimental opinions online, from marketing to medicine to governmental decision-making, sentiment analysis can effectively analyze the underlying sentiments of different opinions expressed by people and gain valuable insights [2].

The stock market plays an important role in economic development, and the vast majority of investors are using the Internet to trade stocks. It has become a common phenomenon to express stock opinions on the stock market message board, which has generated a large number of stock reviews. Sentiment analysis of stock market commentary plays a positive role in stock market investment [3]. Bhardwaj et al. [4] conducted pre-processing and feature selection of stock data, and then sentiment analysis was conducted to obtain stock market conditions. Rao and Srivastava [5] used a Naive Bayesian classifier for sentiment classification of stock reviews. The results show that the polarity of sentiment has a strong influence on stock price movements and prove that the sentiment of the previous week has a strong influence on the opening and closing prices of stocks in the next week. Xu and Keselj [6] used the dataset of tweets from the StockTwits website to propose a two-stage sentiment classification method using SVM, and the experimental results show that the overnight activity of users on StockTwits was positively correlated with the stock trading volume of the next day. Xiaodong Li et al. [7] show that the research on sentiment analysis of stock reviews can greatly improve the accuracy of stock market prediction.

Although much research has made contributions to sentiment analysis of stock reviews, as the labeled data sets are usually small in the field of Chinese stock reviews, which makes it difficult to train complex models to improve the accuracy of classification. Therefore, an innovative BERT-BiLSTM-Attention model is proposed in this paper. The specific contributions of this work are as follows:

1. Aiming at the problem of less labeled data in Chinese stock reviews data, this paper proposes an innovative BERT-BiLSTM-Attention model to improve the accuracy of sentiment analysis.
2. Since most BERT models do not consider context information according to different context backgrounds, to solve this problem, this paper proposes to combine BERT model with BiLSTM model and Attention mechanism. BiLSTM can solve the problem of insufficient dependence on long text, enhance the model's ability to capture context information, and improve the model's training and reasoning efficiency through parallel computation. Further, the Attention mechanism is used to extract important text information, which makes the sentiment classification more accurate.
3. In this work, the proposed model is compared with other state-of-the-art deep learning models, and evaluation indicators are introduced to prove that the model proposed in this work is of great significance to the development of the stock market. In addition, the model proposed in this paper has good generalization ability and can be applied to other fields.

The remainder of this paper is organized as follows. Section 2 provides related works. Section 3 provides details of the sentiment analysis method proposed in this paper. Section 4 presents the experimental results and discusses the results. Section 5 summarizes the proposed method and outline future work.

## 2 Related works

With the continuous integration of sentiment analysis methods and application fields, sentiment analysis has now been widely used in a variety of tasks, such as public opinion on political events, customer satisfaction evaluation, stock market research, and so on. Reviewing the development process of sentiment analysis, it can be divided into three stages. The first stage is the method based on sentiment dictionary, which uses the existing dictionary to judge the sentiment tendency of the text according to the occurrence times of positive and negative sentiment words. Guixian Xu et al. [8] constructed an extended sentiment dictionary and utilized the extended sentiment dictionary and designed sentiment scoring rules to realize text sentiment. Murtadha Ahmed [9] proposed SentiDomain, a new method for constructing dome-related sentiment dictionaries, which is trained on weakly supervised unlabeled data by reconstructing input sentence representations from result representations. Although many scholars have done a lot of research based on sentiment dictionaries, the disadvantage is that the sentiment dictionary depends on a manual design, which leads to insufficient coverage of sentiment word. on your various Sections.

The second stage is based on machine learning [10]. Rajkumar et al. [11] used Naive Bayes algorithm and support vector machine algorithm to classify the sentiments of positive or negative reviews of electronic products on Amazon. Suhasini et al. [12] were able to identify sentiments on Twitter using supervised learning. K-nearest neighbor (KNN) and naive Bayes (NB) were the two algorithms compared and the study shows that naive Bayes outperformed the K-nearest neighbor. Jayakody et al. [13] collected data from twitter posts based on product review, then analyzed using the support vector machine (SVM), logical regression, and K-nearest neighbor machine learning algorithm and count vectorizer and term frequency-inverse document frequency mechanisms for converting text into vectors for the data to be inputted into the machine learning model. The highest accuracy score was achieved by logistic regression with a count vectorizer with an accuracy rate of 88.26%. Bhagat et al. [14] used a hybrid approach of naive Bayes and K-nearest neighbor to divide tweets into three classes: positive, negative, and neutral, and they achieved a better accuracy than the random forest. Although machine learning has many advantages over dictionary-based sentiment analysis, it requires manual feature extraction, which has limited generalization ability and poor robustness.

The third stage is a deep learning-based approach. The application of deep learning algorithms in the field of sentiment analysis enables sentiment analysis models to better capture and understand sentiment information in more complex situations ranging from single words to context and semantics, which helps to improve the accuracy and adaptability of sentiment analysis. Jiyao Wei et al. [15] proposed BiLSTM model with multipole orthogonal attention for implicit sentiment analysis. Experimental results on SMP2019 implicit sentiment analysis dataset and two explicit sentiment analysis datasets show that the proposed model can accurately capture the feature differences between sentiment polarities. Guixian Xu et al. [16] integrated the contribution of sentiment information into the traditional TF-IDF algorithm and generates weighted word vectors. The weighted word vectors are input into bidirectional long short term memory (BiLSTM) to capture the context information effectively, experiments show that this method has a high evaluation accuracy. Yue, Wang et al. [17] used CNN-BiLSTM Deep Model to study sentiment analysis, and combined the word vector model (Word2vec), two-way long short-term memory network (BiLSTM) and convolutional neural network

(CNN) in Quora dataset. Experiments show that the accuracy of CNN-BiLSTM model associated with Word2vec word embedding reaches 91.48%. This proves that the performance of hybrid network model is better than that of single neural network in short text. Huang et al. [18] designed a sentiment-enhancing ELSTM model, which was further integrated with CNN, and then introduced the topic-level attention mechanism to adaptively adjust the weight of text hidden representation. Experiments show that this method can effectively improve the performance of sentiment classification. By exploring attention-based and inattention-based GRU networks, Zhang et al. [19] avoided the weight bias caused by using raw soft attention to extract memory information at each attentional step.

Although the above deep learning models have brought great breakthroughs in sentiment analysis research, but most language models are based on unidirectional architectures, i.e., outputs are conditioned only on previous words (left context). When applying such models on downstream tasks, fine tuned models are also limited to be left conditioned. This is a limitation for tasks in which the whole text is available during prediction. BERT introduces a bidirectional language model architecture in order to explore such knowledge. In 2018, Devlin et al. [20] introduces a new language representation model, namely BERT (Bidirectional Encoder Representations from Transformers). This model has successfully improved recent works in finding representations of words in a digital space from their context. MG Sousa et al. [21] used the bidirectional encoder representation from transformer BERT to perform sentiment analysis of stock market news articles. By fine-adjusting the BERT model, they conducted experiments on artificially labeled data sets, and the results showed that the output of the proposed model could provide valuable information for predicting the subsequent trend of Dow Jones Industrial Index (DJI). Syaiful Imron et al. [22] Extract Reviews on Bukalapak based on aspect-based sentiment analysis, using BERT as word embedding and LSTM for aspect extraction. CNN was used for emotion extraction, and the experiment showed that Classification accuracy increased by 2.04% compared to classification without using stemming on datasets. Quoc Thai Nguyen et al. [23] Based on the Vietnamese review dataset, uses only the [CLS] token as the input for an attached feed-forward neural network, then BERT output vectors are used as the input for classification. Experimental results on two datasets show that models using BERT slightly outperform other models using GloVe and FastTex. Mingzheng Li et al. [24] proposed an analysis model of Chinese stock reviews based on BERT, use a BERT pre-training language model to perform representation of stock reviews on the sentence level, and subsequently feed the obtained feature vector into the classifier layer for classification. The proposed model can obtain the best results which are indicated to be effective in Chinese stock review sentiment analysis. Abayomi Bello et al. [25] used BERT for natural language processing with other variants. The experimental results show that the combination of BERT and CNN, BERT and RNN performs well compared with Word2vec and Word2VEC without variants in terms of accuracy and f1 score.

Inspired by the success of BERT model, this work proposes the stock market sentiment analysis model combining BERT with BiLSTM and Attention mechanism, aimed to improve the accuracy of sentiment classification for stock market reviews. It can be proved that the model proposed in this work has good generalization ability and robustness, it can be used in different fields.

## 3 Methodology

This work proposes an innovative model, the BERT-BiLSTM-Attention model. The model consists of five parts: input layer, BERT layer, BiLSTM layer, Attention layer and output layer. The model structure diagram is shown in Figure 1. The specific feature extraction process is as follows: Firstly, the pre-processed stock comment text was input into the pre-trained BERT model to obtain context-dependent word vector representation of each word. Then, the word vector extracted by BERT was input into BiLSTM to capture contextual information and contextual features in the text. Thirdly, to further improve the performance of key feature extraction in the text, the Attention layer is connected after BiLSTM, and finally the comprehensive feature representation of the text is formed.

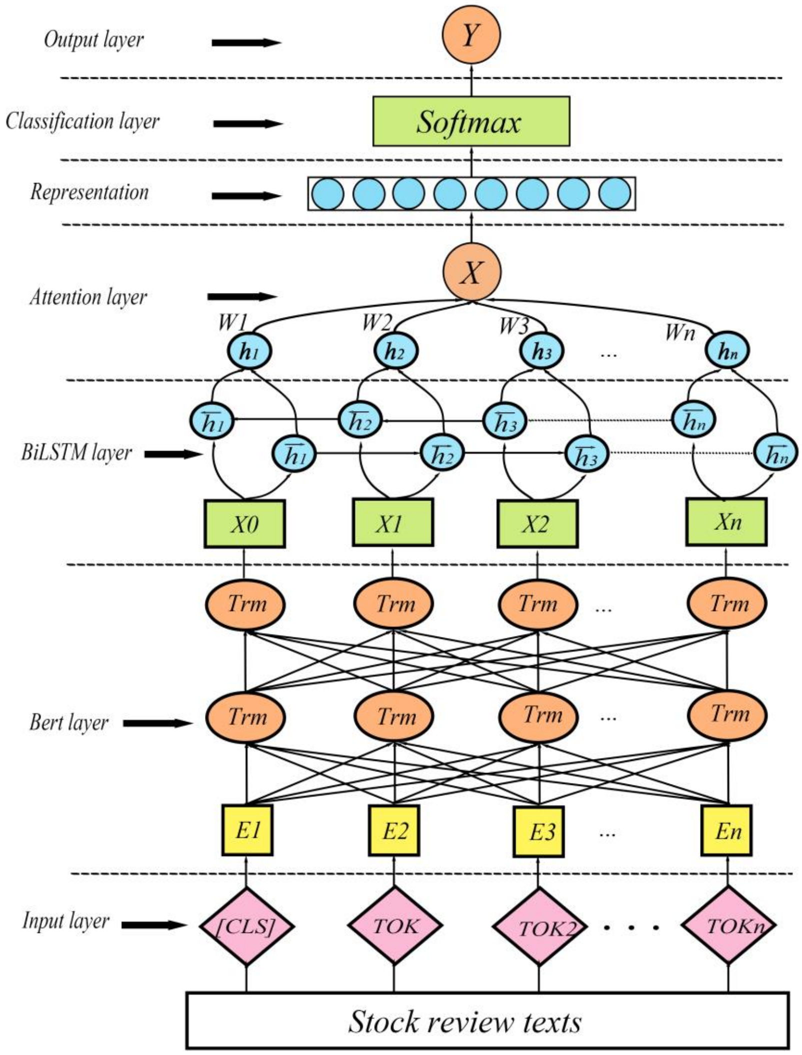


**Figure 1.** BERT-BiLSTM-Attention model

### 3.1 BERT

BERT is a bidirectional representation pre-trained language model based on the Transformer model, which can consider context information in both left and right directions to capture richer language features. For the input to the BERT model, the input sequence is expressed as: $[CLS], W_1, W_2, \dots, W_n, [SEP]$, where $[CLS]$ is a special sentence start marker, $[SEP]$ is the sentence separation marker. The stock review text is subjected to preprocessing operations such as syncopation and disambiguation to obtain a serialized representation $X = [X_1, X_2, \dots, X_n]$, where $X_i$ is the $i$th word or word segmentation in the text, each of which can be converted to a word embedding vector by the word embedding layer, expressed as $E = [E_1, E_2, \dots, E_n]$, where $E_i$ represents the $i$th given embedding vector. The input vector is passed through a multilayer Transformer Encoder to produce word-level output, expressed as $T = [T_1, T_2, \dots, T_n]$, where $T_i$ represents the representation vector of the $i$th word.

The BERT model uses training followed by fine-tuning for sentiment classification. The pre-training task of BERT consists of two phases MLM (Masked Language Model) and NSP (Next Sentence Prediction), which are used to learn lexical and sentence-level representations, respectively. MLM is a completion task that randomly replaces some words in the input text by randomly replacing some words with special symbols [MASK] and then letting the model predict the masked words. NSP is a binary categorization task that, given two sentences, lets the model determine whether they are consecutive or not.

There are two possibilities for fine-tuning BERT, the first is to fine-tune only the categorized output header, keeping a large number of parameters of the Pre-training unchanged, and only focusing on the categorized output header. The other is to fine-tune the parameters of BERT as a whole during the fine-tuning process. The first one is used in this work, add BiLSTM-CNN layers to the twelfth layer, these classification layers are jointly fine-tuned with BERT.

The working model diagram of BERT is shown in Figure 2.

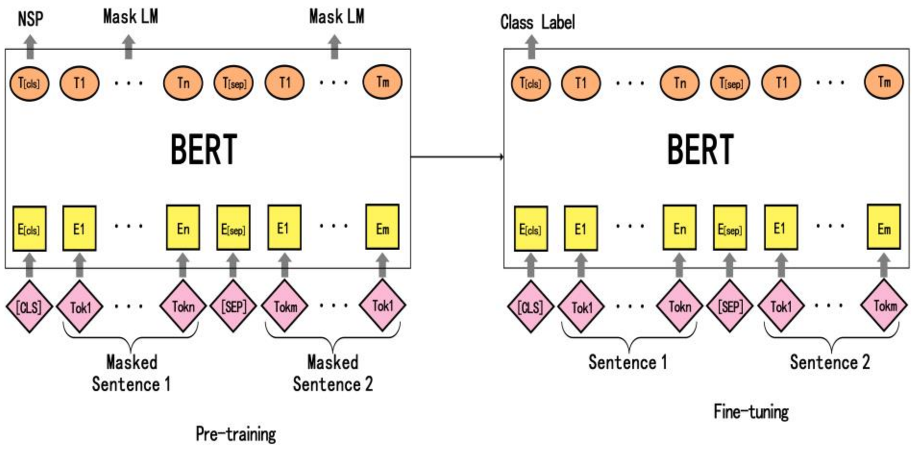


**Figure 2.** BERT model

### 3.2 BiLSTM

BiLSTM is a bidirectional long and short term memory model (see Figure 3), which is a combination of forward LSTM and backward LSTM. As a variant of RNN, LSTM can solve problems such as gradient disappearance or gradient explosion of RNN. Based on RNN, LSTM introduces storage and memory functions and controls information input and output through a threshold mechanism. To control the forgetting and memory of information, an LSTM unit includes an input gate $i$, a forgetting gate $f$, an output gate $o$, and a memory unit $c$. The formula is shown as follows:

$$i_t = \sigma(W^i x_t + U^i h_{t-1} + b^i) \quad (1)$$

$$f_t = \sigma(W^f x_t + U^f h_{t-1} + b^f) \quad (2)$$

$$o_t = \sigma(W^o x_t + U^o h_{t-1} + b^o) \quad (3)$$

$$\mu_t = tanh\ (W^\mu x_t + U^\mu h_{t-1} + b^\mu) \quad (4)$$

$$c_t = i_t * \mu_t + f_t * c_{t-1} \quad (5)$$

$$h_t = o_t * tanh\ (c_t) \quad (6)$$

$i_t$ is the information obtained by the input gate, $x_t$ is the feature of the input, $W^i$ denotes the input gate weight, $h_{t-1}$ is the hidden layer of the previous state, $U^i$ is the transformation matrix of the hidden layer of the input gate, and $b^i$ denotes the input gate bias. $f_t$ is the information obtained by the forgetting gate, $W^f$ denotes the forgetting gate weight, $U^f$ is the transformation matrix of the hidden layer of the forgetting gate, and $b^f$ denotes the forgetting gate bias. $o_t$ is the information obtained by the output gate, $W^o$ denotes the output gate weight, $U^o$ is the transformation matrix of the hidden layer of the output gate, and $b^o$ denotes the output gate bias. $\mu_t$ is the updated input information, $c_t$ is the current state of the memory cell, and $h_t$ is the updated hidden layer, where $\sigma$ and $tanh$ are activation functions.

From the above formulas, it can be seen that LSTM also enhances the ability to capture remote semantic dependencies and memorize historical information.

The model structure of LiSTM can be seen in Figure 3.

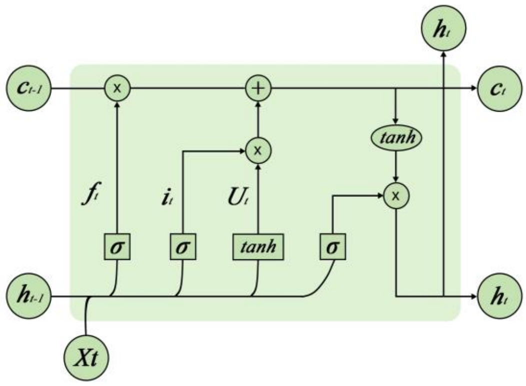


**Figure 3.** LSTM model

The model design concept of BiLSTM is to make the feature data obtained at time $t$ have information between the past and the future at the same time. The formula is as follows:

$$\overrightarrow{h}_t = LSTM(X_t, \overrightarrow{h}_{t-1}) \tag{7}$$

$$\overleftarrow{h}_t = LSTM(X_t, \overleftarrow{h}_{t-1}) \tag{8}$$

$$h_t = w_t\overrightarrow{h}_t + v_t\overleftarrow{h}_{t+}b_t \tag{9}$$

Where $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ represent the forward output and backward output of the LSTM at time $t$ respectively, $h_t$ indicates the output of the BiLSTM at time $t$, $w_t$ denotes the weight matrix of forward output when $v_t$ denotes the weight matrix of backward output. $b_t$ indicates the offset at time $t$.

Existing research proves that BiLSTM model has better text feature extraction efficiency and performance than a single LSTM structure model. The model figure of BiLSTM can be seen in Figure 4.

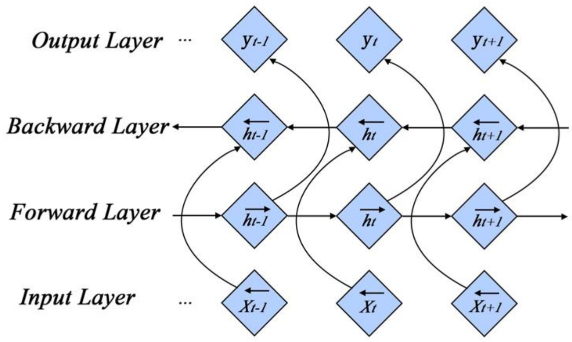


**Figure 4.** BiLSTM model

### 3.3 Attention mechanism

Distinct parts of the phrase have distinct meanings in the actual text, which influences the overall sentimental statement. Consequently, the attention mechanism is added to the BiLSTM model's output layer in this work, giving the key and important words more weight, to reflect the varying relevance of each sentence's component to the feelings that the entire phrase aims to communicate.

This study introduces the attention mechanism to compute the semantic connection coefficient, or how each word in a text sequence relates to other terms, determine the word vector weights based on correlation coefficient. Word context vector is the weighted linear combination of word vectors and is combined as a new word vector feature representation with the original word vector. Larger weighted words can be made to receive more attention in this way using the word context vector. As the words that are furthest apart in the sequence have less semantic links, to highlight the information contained in the sentence keywords, the text is split into sentence sets, and the context vector of those sentence words is computed.

Three steps make up the attention implementation process:

Given a text sequence of $k$ sentences and a sentence length of $m$, the word vector representation of the $j$th word in the $i$ th sentence is denoted as $x_{ij}$. The correlation coefficient between the $i$ th sentences' $j$ th and $k$ th words is denoted by the symbol $\alpha_{ij,k}$.

The context variables of the words are as follows:

$$g_{ij} = \sum_{k=1,k\neq j}^{m} \alpha_{ij,k}\, x_{i,k} \tag{10}$$

Where $\alpha_{ij,k}$ is attention weights, which was obtained by softmax regression:

$$\alpha_{ij,k} = \frac{\exp\left(score(x_{i,j}, x_{i,k})\right)}{\sum_{h=1}^{m} \exp\left(score(x_{i,j}, x_{i,h})\right)} \tag{11}$$

Where

$$score\left(x_{i,j}, x_{i,k}\right) = v_t^T \tanh\left(w_t x_{i,j} + w_u x_{i,k}\right) \tag{12}$$

The correlation coefficient between the current word and other words is calculated using the score function to determine how closely connected the current word is to other words. The higher score value indicates that the word is given more weight in the current word's context vector $v_t$, $w_t$ and $w_u$ are training parameters, tanh is activation function for feature extraction. The word's context vector is split with the original word vector to create a new word vector representation, which can be shown as follows:

$$x_{i,j}^{l} = \left[x_{i,j}, g_{i,j}\right], l = 1 \ldots, n \tag{13}$$

The output y can be expressed as follows:

$$y = (x_{i,j}^{1}, x_{i,j}^{2}, \ldots, x_{i,j}^{n}) \tag{14}$$

## 4 Experiments

Experiments were carried out to evaluate the efficacy of the proposed method in sentiment analysis of stock reviews. This part contains following 3 sections, experimental setup, experimental evaluation index, experimental results and discussion.

### 4.1 Experimental setup

#### 4.1.1 Dataset

The experimental dataset in this study is the Chinese stock review dataset. Few datasets exist with labels for reviews of Chinese stocks, this is one of the problems that we need to solve in this work. The datasets comprise 9204 reviews, including 4602 positive reviews and 4602 negative reviews. They are labeled as"positive" and "negative"respectively. The dataset is divided into the training and test sets in a ratio of 8:2. The instance of the datasets are demonstrated in Table 1.

**Table 1.** Instance of dataset

| Text | Label |
|---|---|
| 除了买不到的，就是不敢买的，奇葩股市 | Negative |
| 个股全部拉稀，大盘居然还是红的，真他妈假 | Negative |
| 中国股市的春天正在来到 | Positive |
| 二十年的老股民了，从没想现在这么无奈和恐惧 | Negative |

| | |
|---|---|
| 几次股灾，都没失去信心。然而这次... | |
| 周末 利好 势在必行,朋友圈已放出,今日抄底钱 | Positive |
| 杭州西湖夜色茫茫， 杨柳梢风动 ，明日 A 股一山更比一山高 | Positive |
| 新年新气象、新的行情、大展鸿图、大家收获满满！ | Positive |
| 老子就想问一下这种指数涨个股跌的状态还要持续多久，垃圾股市，尼玛 | Negative |
| 有点怕，仓位降降先 | Negative |
| 在这么走下去，你们就不怕最后的崩盘吗吗？ | Negative |

### 4.1.2 Experimental parameter setting

In this paper, the Chinese pre-training model "BERT-Base-Chinese" released by Google is used as the model input. The vector dimension of the pre-training words for this model is 768. The 12-layer Transformer is used, and the activation function is Relu. The maximum training length is 128. To avoid overfitting, dropout is used. Optimization is used to reduce losses during training. Adam is used for optimizer. The BERT-BiLSTM-Attention model is constructed by the Python language. The operating system of our experiments is Windows 11, the processor is Intel(R) Core (TM) i9-10900K CPU @3.70GHZ 3.70GHZ and GPU is GeForce RTX 4070. The optimal parameter settings are shown in Table 2.

**Table 2. Parameters' setting of BERT-BiLSTM-Attention**

| Parameters | Value |
|---|---|
| Bert_model | bert-base-Chinese |
| lstm_hidden_size | 64 |
| dropout_rate | 0.3 |
| max_length | 128 |
| batch_size | 32 |
| learning_rate | 1e-4 |
| word vector dimension | 768 |
| L2 regularization coefficient | 0.7 |
| optimizer | Adam |

### 4.2 Experimental evaluation index

In this work, accuracy rate, recall, F1 score, and confusion matrix are used as evaluation criteria, and the calculation formula is as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN})/\text{N} \tag{15}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{16}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{17}$$

$$\text{F1score} = 2\text{PR}/(\text{P} + \text{R}) \tag{18}$$

Where TP is the number of samples correctly predicted by the sentiment classifier to be in a certain sentiment category, TN is the number of samples correctly predicted by the sentiment classifier to be not in a certain sentiment category, FP is the number of samples not in a certain sentiment category incorrectly predicted by the sentiment classifier to be in that category, and FN is the number of samples incorrectly predicted by the sentiment classifier to be in a certain sentiment category to be not in that category.

The confusion matrix is used to summarize the predictions of the classification model. The true and expected categories determined by the classification model summarize the records of the data set in the form of a matrix. The columns of the matrix correspond to the expected values and the rows correspond to the true values.

### 4.3 Experimental results and discussion

The experimental results of sentiment analysis of the proposed model are showed by Table 3 and Figure 5.

**Table 3.** Classification performance of BERT-BiLSTM-Attention model

| | precision | recall | F1score | Support |
|---|---|---|---|---|
| Negative | 0.9279 | 0.9514 | 0.9395 | 906 |
| Positive | 0.9519 | 0.9285 | 0.9400 | 937 |
| accuracy | | | 0.9398 | 1843 |
| macro avg | 0.9519 | 0.9400 | 0.9398 | 1843 |
| Weighted avg | 0.9401 | 0.9398 | 0.9398 | 1843 |

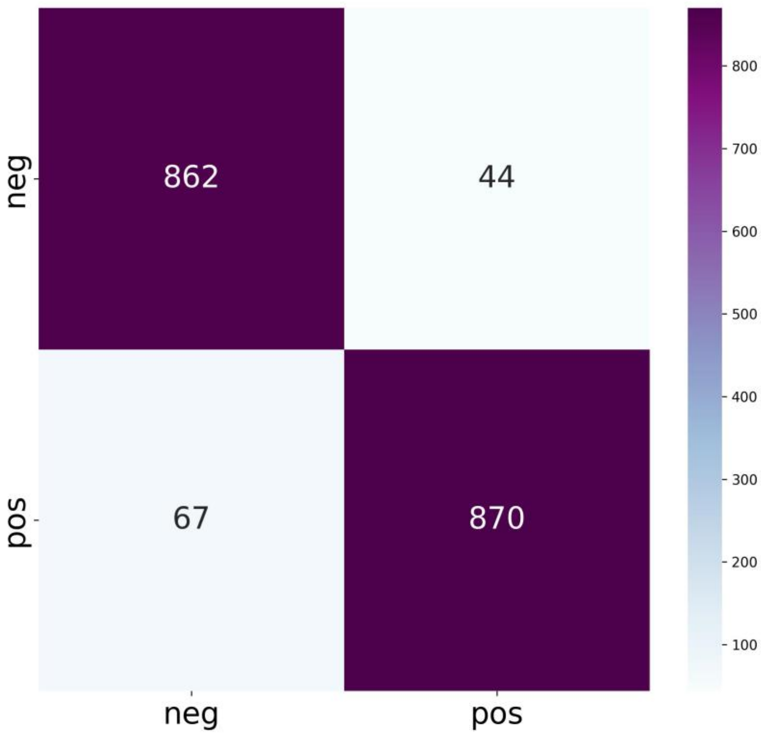


**Figure 5.** Confusion matrix of BERT-BiLSTM-Attention

From the Table 3, it can be seen that the accuracy of BERT-BiLSTM-Attention model for sentiment analysis on test datasets can be up to 0.9398. The confusion matrix of BERT-BiLSTM-Attention model is shown in Figure 5, There has 1843 reviews in test dataset, the positive reviews is 937, the negative reviews is 906. From confusion matrix, it shows that 870 reviews are expected to positive reviews, 862 reviews are expected to negative reviews. These experimental results show that BERT-BiLSTM-Attention model has good performance in sentiment analysis on stock reviews.

To prove the validity of the proposed model, we designed three comparative experiments with BERT-BiLSTM-CNN model, BERT-BiLSTM model, BERT model on the same datasets respectively.

The experimental results of sentiment analysis of BERT-BiLSTM-CNN model are showed by Table 4 and Figure 6.

**Table 4.** Classification performance of BERT-BiLSTM-CNN model

| | precision | recall | F1score | Support |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Negative | 0.9062 | 0.9172 | 0.9117 | 906 |
| Positive | 0.9190 | 0.9082 | 0.9136 | 937 |
| Accuracy | | | 0.9126 | 1843 |
| macro avg | 0.9126 | 0.9127 | 0.9126 | 1843 |
| Weighted avg | 0.9127 | 0.9126 | 0.9126 | 1843 |

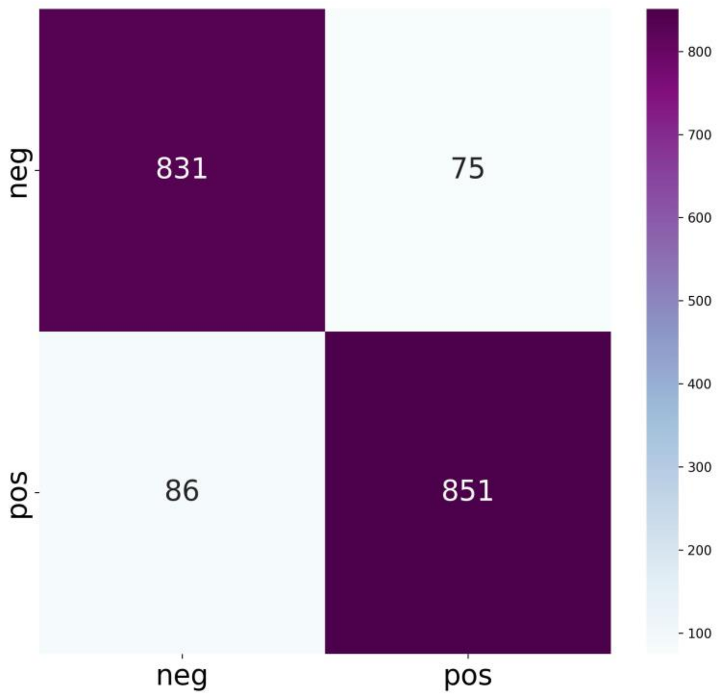


**Figure 6.** Confusion matrix of BERT-BiLSTM-CNN

From the Table 4, it can be seen that the accuracy of BERT-BiLSTM-CNN model for sentiment analysis on test datasets is 0.9126. The confusion matrix of BERT-BiLSTM-CNN model is shown in Figure 6, it shows that 851 reviews are expected to positive reviews, 831 reviews are expected to negative reviews. These experimental results show that BERT-BiLSTM-Attention model has better performance than BERT-BiLSTM-CNN in sentiment analysis on stock reviews.

The experimental results of sentiment analysis of BERT-BiLSTM model are showed by Table 5 and Figure 7.

**Table 5.** Classification performance of BERT-BiLSTM model

| | precision | recall | F1score | Support |
|---|---|---|---|---|
| Negative | 0.9204 | 0.9448 | 0.9325 | 906 |
| Positive | 0.9452 | 0.9210 | 0.9330 | 937 |
| accuracy | | | 0.9327 | 1843 |
| macro avg | 0.9328 | 0.9329 | 0.9327 | 1843 |
| weighted avg | 0.9330 | 0.9327 | 0.9327 | 1843 |

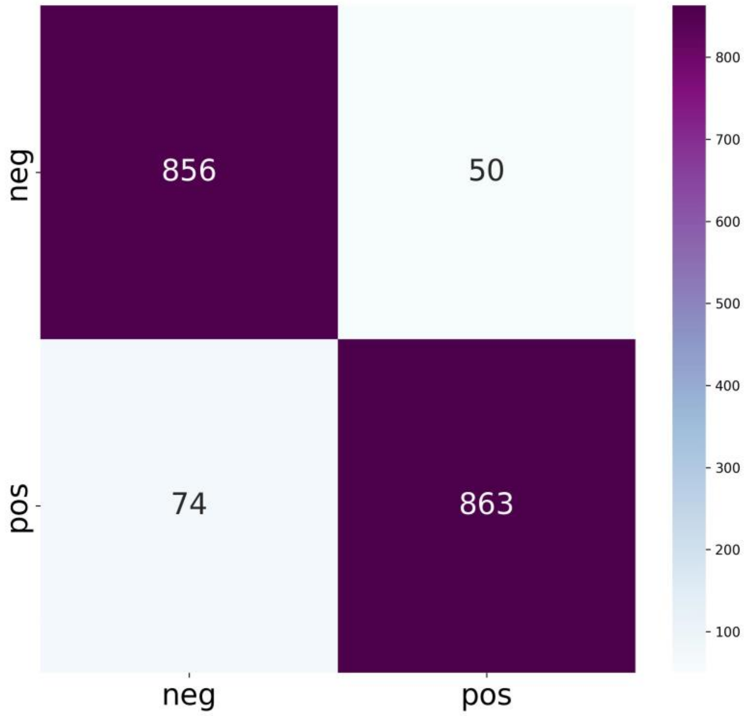


**Figure 7.** Confusion matrix of BERT-BiLSTM

From the Table 5, it can be seen that the accuracy of BERT-BiLSTM model for sentiment analysis on test datasets is 0.9327. The confusion matrix of BERT-BiLSTM model is shown in Figure 7, it shows that 863 reviews are expected to positive reviews, 856 reviews are expected to negative reviews. These experimental results show that BERT-BiLSTM-Attention model has better performance than BERT-BiLSTM in sentiment analysis on stock reviews.

The experimental results of sentiment analysis of BERT model are showed by Table 6 and Figure 8.

**Table 6.** Classification performance of BERT model

| | precision | recall | F1score | support |
|---|---|---|---|---|
| Negative | 0.9388 | 0.9316 | 0.9352 | 906 |
| Positive | 0.9343 | 0.9413 | 0.9378 | 937 |
| accuracy | | | 0.9365 | 1843 |
| macro avg | 0.9366 | 0.9364 | 0.9365 | 1843 |
| weighted avg | 0.9365 | 0.9365 | 0.9365 | 1843 |

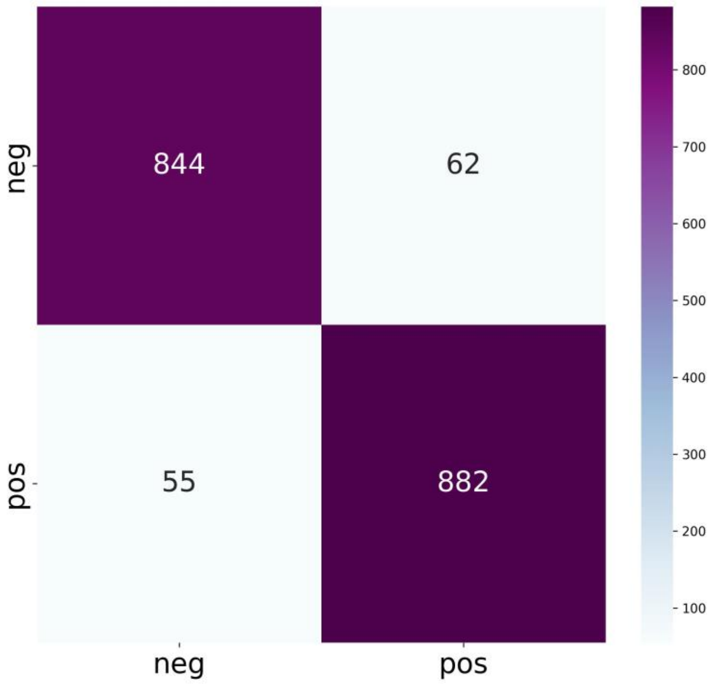


**Figure 8.** Confusion matrix of BERT

From the Table 6, it can be seen that the accuracy of BER model for sentiment analysis on test datasets is 0.9365. The confusion matrix of BERT model is shown in Figure 8, it shows that 844 reviews are expected to positive reviews, 882 reviews are expected to negative reviews. These experimental results show that BERT-BiLSTM-Attention model has better performance than BERT in sentiment analysis on stock reviews.

The comparison results of proposed model and three comaprative models are shown in Table 7.

**Table 7.** Model comparsion results

| Model | Accuracy | Precision | F1score |
|---|---|---|---|
| BERT-BiLSTM-Attention | 0.9398 | 0.9519 | 0.9400 |
| BERT-BiLSTM-CNN | 0.9126 | 0.9190 | 0.9136 |
| BERT-BiLSTM | 0.9327 | 0.9452 | 0.9330 |
| BERT | 0..9365 | 0.9343 | 0.9378 |

Furthermore, considering the performance of the model is usually affected by the number of epochs, this work also executes experiment through epochs which is shown in Figure 9. In general, the model performs better as the number of epochs increases, however, overfitting issues could arise if epoch grows to a certain level. As seen from Figure 9, when the Epoch is 2, the performance of the proposed model is the best, then the accuracy decreases and fluctuates within a small range.

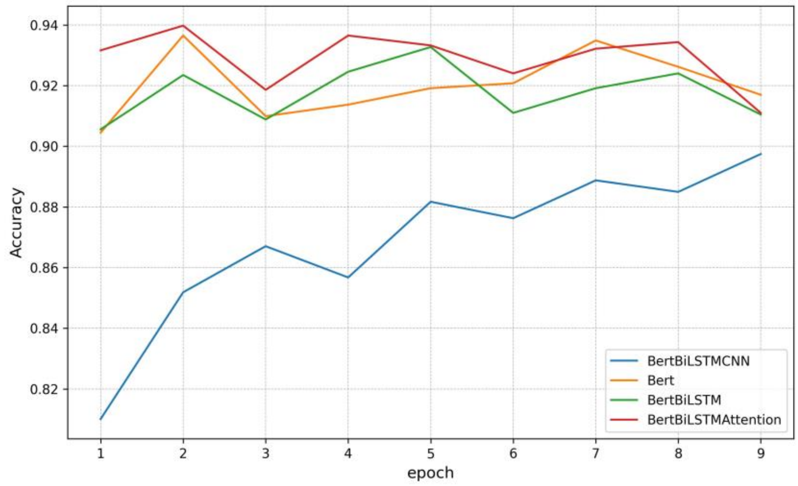


**Figure 9.** Relationship between Epochs and Accuracy

From the above experiment results, it can be seen that the BERT-BiLSTM-Attention model proposed in this work performs better than other competitive models and reflects its superiority.

# 5 Conclusion

In the stock market, sentiment has a significant impact on the volatility of stock prices. Therefore, sentiment analysis of stock reviews can help to understand the emotional tendencies of market participants and provide investors with more accurate information for decision making. As the number of current labeled Chinese stock reviews are very small, this work proposed an innovative model BERT-BiLSTM-Attention to solve the sentiment classification for these limited labeled datasets. The proposed model can not only take advantage of BERT to consider the contextual information on the left and right sides of a word at the same time, but also incorporate the BiLSTM model to enhance the long-distance dependencies as well as the ability to capture bi-directional sequential information. The Attention mechanism is further introduced to enable the model to focus on the most relevant parts of the current context while performing the task, thus improving the model performance. Experiments demonstrate that proposed model can effectively improve the accuracy of sentiment classification and achieve significant performance gains in ssentiment analysis tasks.

To prove the superiority of the proposed model, this work designed competitive models for comparative tests. The results of the experimental evaluation indicators show that the proposed model outperforms other models in terms of accuracy, F1score, and so on. The performance of BERT-BiLSTM-Attention model is better than other competitive models. Future work can further validate the robustness and utility of the model in a wider range of applications in finance.

**Author Contributions:** Conceptualization, Xiaoyan Li.; methodology, Xiaoyan Li.; validation, Xiaoyan Li., Lei Chen.; formal analysis, Xiaoyan Li.; investigation, Baoguo Chen.; resources, Xiaoyan Li.; data curation, Rodolfo C. Raga Jr.; writing—original draft preparation, Xiaoyan Li.; writing—review and editing, Xiaoyan Li., Lei Chen; visualization, Xiaoyan Li.; supervision, Baoguo Chen., Rodolfo C. Raga Jr.; project administration, Lei Chen,; funding acquisition, Xiaoyan Li. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by 2023 Scientific research project of Anhui Provincial Education Department (Grant No. 2023AH051551) and the Opening Foundation of State Key Laboratory of Cognitive Intelligence, iFLYTEK (COGOS-2023HE02).

**Data Availability:** The datasets used to support the results of this study are available from the Github (https://github.com/algosenses/Stock_Market_Sentiment_Analysis/tree/master/data) website.

**Conflicts of Interest:** The authors declared that there is no conflict of interest in publishing this paper, and the authors confirmed that there is no conflict of interest for author and the co-authors.

# References

1. Pham, T.; Vo, D.; Li, F.; Baker, K.; Han, B.; Lindsay, L.; Pashna, M.; Rowley, R. Natural language
processing for analysis of student online sentiment in a postgraduate program. Pac. J. Technol. Enhanc. Learn. 2020, 2, 15–30. [CrossRef]
2. Jayakody, J.P.U.S.D.; Kumara, B.T.G.S. Sentiment analysis on product reviews on twitter using Machine Learning Approaches. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 7–8 December 2021; pp. 1056–1061.
3. Sheu H-J, Lu Y-C, Wei Y-C (2010) Causalities between sentiment indicators and stock market returns under different market scenarios. Int. J Bus Fin Res 4(1):159–171
4. Bhardwaj A, Narayan Y, Vanraj P, Dutta M (2015) Sentiment analysis for indian stock market prediction using sensex and nifty. In: Procedia computer science, vol 70, pp 85–91.
5. Rao T, Srivastava S (2012) Analyzing stock market movements using Twitter sentiment analysis. In: ASONAM'12 Proceedings of the 2012 international conference on advances in social networks analysis and mining (ASONAM 2012), pp 119–123.
6. Xu F, Kešelj V (2014) Collective sentiment mining of microblogs in 24-hour stock price movement prediction. In: 16th IEEE conference on business informatics, CBI 2014, vol 2, pp 60–67
7. Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing & Management, 57(5), 102212.
8. Xu, G., Yu, Z., Yao, H., Li, F., Meng, Y., & Wu, X. (2019). Chinese text sentiment analysis based on extended sentiment dictionary. IEEE access, 7, 43749-43762.
9. Ahmed, M., Chen, Q., & Li, Z. (2020). Constructing domain-dependent sentiment dictionary for sentiment analysis. Neural Computing and Applications, 32, 14719-14732.
10. Malviya, S., Tiwari, A. K., Srivastava, R., & Tiwari, V. (2020). Machine learning techniques for sentiment analysis: A review. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 12(02), 72-78.
11. Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In 2013 fourth international conference on computing, communications and networking technologies (ICCCNT) (pp. 1-5). IEEE.
12. Suhasini, M.; Srinivasu, B. Emotion detection framework for twitter data using supervised classifiers. In Data Engineering and Communication Technology; Springer: Singapore, 2020; pp. 565–576
13. Jayakody, J.P.U.S.D.; Kumara, B.T.G.S. Sentiment analysis on product reviews on twitter using Machine Learning Approaches. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 7–8 December 2021; pp. 1056–1061.
14. Bhagat, C.; Mane, D. Text categorization using sentiment analysis. In Proceedings of the International Conference on Computational Science and Applications, Saint Petersburg, Russia, 1–4 July 2019; Springer: Singapore, 2020; pp. 361–368.
15. Wei, J., Liao, J., Yang, Z., Wang, S., & Zhao, Q. (2020). BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. Neurocomputing, 383, 165-173.
16. Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. Ieee Access, 7, 51522-51532.
17. Yue, W., & Li, L. (2023). Sentiment Analysis using a CNN-BiLSTM Deep Model Based on Attention Classification. International Information Institute (Tokyo). Information, 26(3), 117-162.
18. Huang, F., Li, X., Yuan, C., Zhang, S., Zhang, J., & Qiao, S. (2021). Attention-emotion-enhanced convolutional LSTM for sentiment analysis. IEEE transactions on neural networks and learning systems, 33(9), 4332-4345.
19. Zhang, Z., Wang, L., Zou, Y., & Gan, C. (2018). The optimally designed dynamic memory networks for targeted sentiment classification. Neurocomputing, 309, 36-45.
20. Jacob Devlin, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". 2018.
21. Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019, November). BERT for stock market sentiment analysis. In 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI) (pp. 1597-1601). IEEE.
22. Imron, S., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2023). Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 7(3), 586-591.
23. Nguyen, Q. T., Nguyen, T. L., Luong, N. H., & Ngo, Q. H. (2020, November). Fine-tuning bert for sentiment analysis of vietnamese reviews. In 2020 7th NAFOSTED conference on information and computer science (NICS) (pp. 302-307). IEEE.
24. Li, M., Chen, L., Zhao, J., & Li, Q. (2021). Sentiment analysis of Chinese stock reviews based on BERT model. Applied Intelligence, 51, 5016-5024.

25. Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT framework to sentiment analysis of tweets. Sensors, 23(1), 506.