Article
China’s Public Firms’ Attitudes towards Environmental Protection Based on Sentiment Analysis and Random Forest Models

Cai Li¹, Luyu Li²,*, Jiaqi Zheng³, Jizhi Wang⁴, Yi Yuan⁵, Zezhong Lv⁶, Yinghao Wei⁷, Qihang Han⁸, Jiatong Gao⁹ and Wenhao Liu¹⁰

Abstract: In this article, we investigated changes in public firms’ attitudes towards environmental protection in 2018–2021 in China. We crawled the firm–investor Q&A record on the website of East Money, extracted the carbon- and environment-related corpus, and then applied the sentiment analysis method of NLP (natural language processing) to calculate the sentiment weight of each firm-level record to estimate the attitude before and after towards carbon reduction. We found that there were significant changes in firms’ attitudes towards carbon reduction and environmental protection after the COVID-19 pandemic and the implementation of environment-related policies. We also found a heterogeneous effect of the attitude in different industries. In addition, we built several models to examine the relationship between a firm’s carbon reduction attitude and its financial performance. We found that: A goal with consequent specific policies can raise the positive attitudes of firms toward carbon reduction topics; firms’ attitudes toward ecological topics are different from industry to industry, which means that there are different needs and situations in the trend of carbon reduction from industry to industry. COVID-19 influenced firms’ attitudes toward carbon reduction and environmental protection, calling back the classic dilemma or trilemma of economic growth, carbon reduction, and energy consumption or, perhaps, epidemic control today. The stock situation also influenced the attitude toward environmental protection.

Keywords: CO₂ emission reduction; sentiment analysis of China’s public firms; carbon reduction sentiment weight prediction

1. Introduction
From 2018 to 2021, China experienced big events, such as the COVID-19 pandemic, economic transformation, trade war, and environmental topics. Alongside these events, China has proposed carbon reduction targets of “carbon neutrality” and a “carbon peak”. Under these circumstances, we aim to explore how the attitudes of public firms dynamically change. The attitudes of firms toward energy conservation and emission reduction are affected by many factors. According to past field research, some Chinese firms believe that
emission reduction has restricted the development of enterprises, while some believe that it is beneficial in the long term, and the attitude is influenced by industries, technology for and cost of emission reduction, size of firms, and other attributes of firms [1]. Especially under the COVID-19 pandemic, economic policy uncertainty has risen, and emission reduction behavior is also affected by economic policy uncertainty (EPU) [2]. When policy uncertainty increases, manufacturing companies tend to use cheap and highly polluting fossil energy [3]. At present, research on China’s emission reduction issues mostly focuses on regional research [4], several typical industries [3], the relationship of energy consumption, emissions, and the economy [5], and the trade-off between emissions and economic development [6]. Recent research showed that with these policies, China can achieve the carbon intensity target by 2030, but with a negative impact on economic growth [6]; also in addition, energy consumption and economic growth are mutually important influencing factors [4], leading to a trilemma among energy consumption, carbon emissions, and economic growth [5]. However, although short-term effects exist, in the long term, a positive correlation of economic growth and carbon reduction was observed in BRICS and OECD economies [5].

There is not much research work on the firm level, and the existing research focuses more on emission behavior rather than attitude. As for the literature on the attitudes of firms toward emissions, Xing Lu’s work [1] is important, reflecting 120 firms’ attitudes directly through surveys. Firm-level research also showed that in the long term, firms prefer optimizing energy consumption and investing in green technologies, especially non-state-owned firms and firms with high external financing dependence [7]. In response to policies with uncertainty on carbon emission intensity, manufacturing firms prefer to use cheap and dirty fossil fuels [6].

In this article, we studied the dynamic changes in Chinese public firms’ attitudes toward environmental protection from 2018 to 2021 and explored the factors influencing their attitudes to verify how the environmental protection policies, the COVID-19 pandemic, the industries of companies, and the stock performance of companies touch the nerves of companies.

Our contribution mainly lies in: (1) Starting from a Q&A of the listed firms with investor organizations, we constructed a collection of text data of Chinese firms’ comments about carbon reduction; (2) we applied sentiment analysis (NLP methods) to estimate the firms’ attitudes towards carbon reduction, and with the estimated results, we segmented the time span into three periods with two key time points (when the government’s goal was set and when consequent policies were released), leading to the conclusion that, on their own, goals cannot raise the positive attitudes of firms, but goals with consequent policies can; (3) as applying NLP methods in order to estimate firms’ attitudes towards carbon reduction is a complicated and dirty project that involves collecting text data, text mining and cleanup, and conducting NLP methods, we provided another more elegant access to firms’ attitudes toward carbon reduction through financial and industry data, which were modeled by random forests; (4) we explored the industry factor and found that the attitude score differed from industry to industry; (5) we investigated how COVID-19 influenced firms’ attitudes toward carbon reduction, finding that the attitudes did not float significantly before and after COVID-19, but if we controlled the financial data of firms, a more positive attitude could be observed.

2. Materials and Methods

2.1. Workflow

To estimate firms’ attitudes towards environmental protection topics, we collected over 304,000 records of investor Q&A texts with their timestamps from the website of East Money [8], and then extracted the texts relevant to environmental protection, including those about carbon reduction, to calculate the attitude weight score by using sentiment analysis. Then, we analyzed the attitude weight score by industry, period, and other
financial variables from the Choice dataset from East Money. A detailed explanation is shown in Table 1.

**Table 1. Variables and explanations.**

| Variables | Explanation | Source |
|-----------|-------------|--------|
| company   | Name of the invested public firm | East Money’s website |
| com_code  | The stock code of the company | East Money’s website |
| date      | The date when the Q&A was conducted by investors and the firm | East Money’s website, cleaned by the author; only texts about the environment were preserved |
| text      | The text record of the Q&A | East Money’s website, cleaned by the author; only texts about the environment were preserved |
| weight    | The sentiment score calculated from the variable text using sentiment analysis | Calculated by the author |
| period (p1, p2, p3) | The time category of the Q&A record. P1 refers to those from before the “Double Carbon” goal was set. P3 refers to those from after the goal was incorporated into the government’s work report. P2 is the time between p1 and p3. Then, we split p1 into p0 and p1 according to the time of the COVID-19 outbreak in China. When p0 is included, p1 refers to the period between the Wuhan shutdown and when the “Double Carbon” goal was proposed. | The writer set this according to the variable date |
| current_value | of the stock value | Choice dataset from East Money |
| percentage of increase | of the stock value | |
| amount of float | of the stock value | |
| volume | of the stock value | |
| recent trading volume | of the stock value | |
| speed of increment | of the stock value | |
| turnover | of the stock value | |
| volume of transaction | of the stock value | |
| highest value | of the stock value | |
| lowest value | of the stock value | |
| open | of the stock value | |
| close | of the stock value | |
| stock amplitude | of the stock value | |
| quantity relative ratio | of the stock value | |
| price-earnings ratio.TTM. | of the stock value | |
| price-earnings ratio.LYR. | of the stock value | |
| price/book value ratio | of the stock value | |
| market_value | of the stock value | |
| total value | of the stock value | |
| industry | 96 different industries; the industry to which the company belongs | |
| the percentage of float in 60 days | of the stock value | |
| the percentage of float in this year | of the stock value | |

The main steps are shown in Figure 1. We first cleaned up the text data to preserve only the text that was relevant to carbon reduction. Then, we used sentiment analysis to score the attitude of the sentiment in each text datum, which reflected the attitudes of firms in each investor Q&A session. With the sentiment score, we analyzed how firms’ attitudes varied by different periods (segmented by COVID-19 and carbon policies) and by industry by using the Wilcoxon test to verify the significance of group differences. In the next step, with the estimated sentiment score, combined with other stock data collected from the Choice dataset from East Money, we built several models to explore the relationship
between the sentiment weight and these indicators. Then, we obtained predictive results and the RMSE indicator from random forest models, to estimate the performance of the models.

**Figure 1.** Flowchart of our work.

In the processing of the descriptive statistics, we found that, after the goals were proposed by President Xi and incorporated into government work reports, the frequency of words related to the environment mentioned in investor Q&As increased. Through the sentiment analysis, we obtained the sentiment score of the firms in each Q&A session. Then, according to the results of the scores, we verified that there was a significant increase in positive attitudes toward the environment after the “Double Carbon” goal was incorporated into the government report, but not after the goal was set, and there were significant differences between different industries. According to the linear models, we found significant influences from COVID-19, stock values (and floats of stock values), and the industry. Finally, as the NLP method involves a heavy workload in data collection and cleaning, we built models to predict the attitude scores from numerical financial data, which were much easier to collect. The RMSE (of the predicted result and the real data) of each model was calculated to compare the performance of the models and return the best random forest model. This part is summarized in Table 2.

**Table 2.** Methods and findings.

| Methods                          | Findings                                                                 |
|----------------------------------|---------------------------------------------------------------------------|
| Descriptive statistics           | After the goals were proposed by President Xi and incorporated into government work reports, the frequency of words related to the environment mentioned in investor Q&As increased. |
Table 2. Cont.

| Methods                        | Findings                                                                 |
|--------------------------------|--------------------------------------------------------------------------|
| Sentiment analysis (one of the NLP methods) | We obtained the sentiment score for carbon reduction.                     |
| Group analytics (Wilcoxon test) | (1) There was a significant increase in positive attitudes toward the environment after the “Double Carbon” goal was incorporated into government reports, but not after the goal was set. (2) There were significant differences between different industries. |
| Analytics on the sentiment score | model1: lm1 (1) COVID-19 showed a significant influence on the sentiment score. (2) The stock value, float of the stock value, and industry also influenced the sentiment score. |
|                                | model2: lm2 The sentiment score was significantly influenced by whether a firm was in the technology industry. |
|                                | model3: rf1 A non-NLP way to predict firms’ attitudes was provided.        |
|                                | model4: rf2                                                                 |
|                                | Applied the four models for prediction and estimated the models by using the RMSE (a standard machine learning procedure). Model3 (rf1) had the best RMSE, which means the lowest error in prediction. |

2.2. NLP

For further verification and inspection, we applied the sentiment analysis method of NLP (natural language processing) to calculate the sentiment weight of each record in the Q&A text data to estimate the changes before and after the “Double Carbon” goal was set. In recent research, NLP methods have been extensively used to explore the non-numerical aspects of organizations, such as corporate culture, attitude, CSR, the personality traits of CEOs, etc. In Kai Li’s work [9], they used Word2Vec to build dictionaries for corporate culture. In Shavin Malhotra’s work [10], linguistic techniques were also applied to attain a CEO’s traits from a spoken text. In our work, we used sentiment analysis to estimate a corporation’s attitude towards the carbon emission goals and calculated a numerical result to represent the extent of firms’ negative and positive attitudes.

2.3. Sentiment Analysis

Sentiment analysis is an NLP method that was first contributed by Turney [11] and Pang [12], who estimated binary attitudes in comments toward movies and commodities. Word segmentation methods can mainly be categorized into 4 groups: dictionary-based (keyword) word segmentation, word association, statistic-based word segmentation, and understanding-based word segmentation methods [13]. dictionary-based word segmentation methods match text data with the words in a constructed dictionary to obtain a word segmentation result [14]. Statistical methods, such as the support vector machine (SVM), N-gram grammar model (N-gram), hidden Markov model (HMM), and so on, usually use training data to build models [13]. The most common methods of word segmentation are usually combinations of dictionary-based and statistical models. In addition, with the development of deep learning, we have obtained more complex word segmentation methods that are closer to the human brain’s understanding, such as BERT (a bidirectional neural network model). Such models are usually more accurate, have more complex algorithms, and are slower to implement.
As for our research, we aimed to estimate the emotional polarity of text data with more of a focus on sentimental words, rather than other words. In addition, the terms in the investor Q&As were mostly commonly used, standard, modern words; thus, we chose an agile way to detect the words in sentiment dictionaries. A sentiment dictionary maps words and the human emotions that they stand for, and it stores the emotions as computational values, such as numerical or True/False values. For example, we can use a positive number to represent a positive emotion and a negative one in a similar way. The absolute value can reflect the extent of an emotion. An example of the simplest dictionary is given in Table 3.

Table 3. The simplest sentiment dictionary.

| Word    | Sentiment Weight |
|---------|------------------|
| sad     | -1               |
| very sad| -2               |
| happy   | 1                |
| very happy | 2          |

The most prevailing Chinese sentiment dictionaries include Tsinghua Li Jun’s positive and negative sentiment dictionary [15], the Chinese Academy of Sciences’ Chinese sentiment degree dictionary, Dalian University of Technology’s sentiment dictionary, and Tan Songbo’s positive and negative sentiment dictionary based on a hotel evaluation corpus. We compiled the emotional dictionary of the Information Retrieval Laboratory of the Dalian University of Technology [16] and Li Jun’s positive and negative sentiment dictionary from Tsinghua University and used the combination in the word segmentation method after deduplication in the dictionaries.

For the word segmentation results, we removed the stop words before calculating the sentiment score. Commonly used stop vocabularies include the stop vocabulary of Harbin Institute of Technology, the stop vocabulary of Baidu, and the stop vocabulary of the Machine Intelligence Laboratory of Sichuan University. We integrated the Baidu stop word list and the stop word database of the Machine Intelligence Laboratory of Sichuan University [17] and removed the stop words from the segmentation results based on the integrated stop word list.

We applied the combined dictionaries to our text records. The process can be briefly interpreted as estimating the positive level of the corpus according to the words in the text. For example, if a text record was “Thanks, the company actively pays attention to carbon-emission-related policies and actively participates in it. The current financial report does not have this business”, we obtained “thanks | company | actively | pays attention to | carbon emission | related policies | actively | participate”, which are 8 phrases after the word segmentation and removal of stop words (which we did in the former steps). The sentiment dictionary provides a mapping between words and scores.

2.4. Random Forests

The random forests technique is a machine learning method that is advantageous in terms of the usage efficiency of data because of its ability to use out-of-bag (OOB) samples and to rank variables according to their importance [18]. The recent research work by Heinrich [19] used random forest regression for carbon emission estimation to find the importance ranking of variables. In our work, with random forests, we attained the best-predicted results with the lowest RMSE and derived a ranking of importance.

All statistical analysis work was implemented with R version 4.1.2 (and packages for it).

2.5. Data

Our data sources were records of Q&A sessions between investor organizations and Chinese public firms, which were taken from the datasets of East Money. We crawled for company names, stock codes, and investor survey questions and answers on different...
The crawling results included a total of more than 304,000 records of data from 2018 to 2021, and each piece of data contained multiple questions and answers. This amount of data is meaningful in machine learning [20]. The data included records of more than 304,000 questions and answers from 2690 public firms from 13 November 2018 to 12 November 2021. We published the crawled data and some subsequent collated data [21].

The data contained a total of 304,322 questions and answers between public firms and investors. Questions and answers for the same company on the same day were counted as one record; thus, each question and answer contained multiple questions and answers. The period of the data was from 13 November 2018 to 12 November 2021, including the time frame before and after the “Double Carbon” goal was proposed. A total of 2690 public firms were covered by the data. Now, there are currently more than 4000 companies in China’s stock market, and there were 3584 in 2018 as of the beginning of the data collection [22].

To improve the accuracy of the analysis and facilitate the comparison of the situation before and after the relevant policy, we set two key time points (summarized in Table 4). One is 22 September 2020, when President Xi Jinping first mentioned the terms “carbon neutrality” and “carbon peaking” at the United Nations. The other time point is when the “Double Carbon” policy was written into the State Council’s government work report on 5 March 2021.

Table 4. The time distribution of the data.

| Date                  | Number of data records | The proportion of the total volume |
|-----------------------|------------------------|-----------------------------------|
| 13 November 2018 to 22 September 2020 | 134,731 | 44.27% |
| 22 September 2020 to 5 March 2021 | 53,309 | 17.52% |
| 5 March 2021 to 12 November 2021 | 116,282 | 38.31% |

From the perspective of the time distribution of the data, in the raw data, there were 134,731 units of survey data before 22 September 2020 and 53,309 units of survey data between the two dates. From 5 March 2021 to the present, there were a total of 116,282 units of survey data.

3. Results and Discussion

3.1. Descriptive Statistical Results of the Raw Data

Since our original data included all of the Q&A records of public firms, not all records were related to environmental protection, which meant that we needed to extract the records that were related. However, we could still find how much the importance of environmental protection changed over time from 2018 to the end of 2021 according to the proportion of records mentioning keywords of environmental protection in the whole collection of records. Thus, before selecting the text records related to environmental topics, we performed a descriptive statistical analysis on the text records related to energy conservation and emission reduction.

The following (Table 5) presents the frequency of relevant text records in different periods. From the results of the statistical description, after the “Double Carbon” goal was proposed, especially after being incorporated into the government work report, the data in the investor Q&As showed an increasing interest in reducing carbon dioxide emissions (Figures 2 and 3).
Table 5. Frequency of the appearance of relevant words in the data.

| Key words               | Period  | Frequency | The Proportion of Surveys In Each Period |
|-------------------------|---------|-----------|----------------------------------------|
| carbon                  | period1 | 5379      | 3.9924%                                |
|                         | period2 | 3673      | 6.89%                                  |
|                         | period3 | 20,459    | 17.59%                                 |
|                         | total   | 29,511    |                                        |
| low carbon              | period1 | 335       | 0.248%                                 |
|                         | period2 | 638       | 1.197%                                 |
|                         | period3 | 3678      | 3.16%                                  |
|                         | total   | 4651      |                                        |
| carbon neutralization   | period1 | 8         | 0.005927757%                          |
|                         | period2 | 699       | 1.3%                                   |
|                         | period3 | 8702      | 7.48%                                  |
|                         | total   | 9409      |                                        |
| carbon peak             | period1 | 0         | 0                                      |
|                         | period2 | 182       | 0.34%                                  |
|                         | period3 | 5239      | 4.5%                                   |
|                         | total   | 5421      |                                        |
| emission reduction      | period1 | 2316      | 1.7%                                   |
|                         | period2 | 1287      | 2.4%                                   |
|                         | period3 | 5478      | 4.7%                                   |
|                         | total   | 9081      |                                        |
| energy saving           | period1 | 8191      | 6.0795%                                |
|                         | period2 | 2768      | 5.19%                                  |
|                         | period3 | 11,430    | 9.829%                                 |
|                         | total   | 22,389    |                                        |

Figure 2. The frequency of keywords in the dataset.
As for each keyword, we can see dramatic increases in the proportions of term from period1 to period2, and from period2 to period3. To further analyze what the trend stood for, we used the subsequent processing of the sentiment analysis. Specifically, although the trend could reflect the increasing attention to the keywords of the “Double Carbon” goal, we still need an accurate method to estimate the firms’ attitudes towards the keywords. This is what we do in the next section on sentiment analysis.

3.2. Data Pre-Processing

We grouped the data into two categories. The first category included data that contained keywords about double carbon, energy saving, or emission reduction. The keyword set included the seven keywords of “carbon”, “energy-saving”, “emission reduction”, “environmental protection”, “low carbon”, “carbon neutral”, and “carbon peak”.

The other category comprised the rest of the data. Finally, after classification, there were 75,786 units of data in the first category. Among them, the units of data before 22 September 2020 numbered 30,025, the units of data after 5 March 2021 numbered 34,639, and the number between the two points was 11,122 (Table 4).

However, the text records included all of the questions and answers in one session, which meant that not all of the text was about our topic. Therefore, before the next step, we thoroughly cleaned the text to preserve only the Q&As that contained the seven keywords. For example, the record of Hailiang Shares on 16 September 2021 included the basic information and five Q&As, but only Q&A2 and Q&A4 were related to the carbon reduction topic. Thus, after our pre-processing, only Q&A2 and Q&A4 were preserved in the record, and we applied this function to all of the text records of the Q&As. Compared to the original data, the pre-processed data stuck more tightly to the main topic, which was beneficial in improving the performance of the consequent models. We have also published the cleaned data [23].
3.3. Sentiment Weight and Characteristics

3.3.1. The Result of the Sentiment Analysis and its Distribution

We calculated the sentiment weight of each record and show the results in Table 6.

| Period     | Period1                  | Period2                  | Period3                  |
|------------|--------------------------|--------------------------|--------------------------|
| Min:       | −5.000                   | −3.00                    | −5.00                    |
| 1st Qu:    | 3.000                    | 3.00                     | 3.00                     |
| Median:    | 6.000                    | 7.00                     | 8.00                     |
| Mean:      | 9.155                    | 10.28                    | 12.41                    |
| 3rd Qu:    | 12.000                   | 15.00                    | 16.00                    |
| Max:       | 210.000                  | 113.00                   | 809.00                   |
| Count      | 28,977                   | 10,818                   | 33,290                   |

For each unit of text data, we received a sentiment weight, representing the firm’s attitude in a specific Q&A session. A negative value represented a negative attitude, and a positive one represented the opposite. The larger the absolute value was, the greater the extent of the attitude was. In this table (Table 6), we can observe an increasing trend in the median, mean, and third-quartile values. In addition, we have included an interactive picture (Figure 4) to show how the entire sentiment weight flowed over time. If there were multiple records on the same date, we calculated the mean as the sentiment score on that day.

![Figure 4](https://github.com/luyuyuyu/gov_mkt_carbon_nlp/blob/main/p1.html)

This figure (Figure 4) presents the sentiment scores of firms’ attitudes over the whole timeline. We added three vertical lines to the plot: (1) The time point of the Wuhan shutdown because of COVID-19; (2) the time point of the “Double Carbon” goal proposed...
by President Xi; (3) the goal was incorporated into the government’s work report. We used a 30-day rolling average on the data.

We can observe that: (1) There was a low-level weight around the period of the Wuhan shutdown. A possible reason can be the negative influence of COVID-19 on emotions and expectations, which will be one of our focal points later; (2) a dramatic soar of the sentiment weight in the last period after the goal was incorporated into the government report; (3) according to the figure, we cannot tell whether the change after the goal was proposed on 22 September 2020 is significant, which will be discussed in the next section.

We then calculated the average sentiment weight on the same date in each period group (there were multiple records made on the same day by different companies). As shown in Figure 5, we found that the p2 records showed a slightly higher average sentiment weight in the distribution than that of p1, and p3 had a higher average sentiment weight than those of both p1 and p2.

![Figure 5. Distribution of the daily average sentiment weight in each period.](image)

### 3.3.2. The Group Differences in Sentiment Weight

To verify the significance of the differences in each period, we conducted a Wilcoxon test (results in Table 7). In addition, we have visualized the test results in Figure 6.

| Variable | Group 1 | Group 2 | p          | p.Adj | p.format | p.Signif | Method   |
|----------|---------|---------|------------|-------|----------|----------|----------|
| 1 avg_senti | p1     | p2     | 0.872064   | 0.87  | 0.87     | ns       | Wilcoxon |
| 2 avg_senti | p1     | p3     | 3.02 × 10^{-13} | 9.10 × 10^{-13} | 3.00 × 10^{-13} | ***     | Wilcoxon |
| 3 avg_senti | p2     | p3     | 6.71 × 10^{-9} | 1.30 × 10^{-8}  | 6.70 × 10^{-9}  | ***     | Wilcoxon |

ns: \( p > 0.05 \); ***: \( p \leq 0.001 \).
The test verified that there was a significant increase in the average sentiment weight after the “Double Carbon” goal was incorporated into the government’s work report (p2 vs. p3), but not after the goal was set (p1 vs. p2), which implied that firms would not change their attitudes only because of the government’s goal, but further policies would push them to be significantly more positive in their attitudes toward environmental protection (at least with their attitudes in public).

As shown in Figure 4, we also observed a low level around the Wuhan shutdown, leading us to the influence of COVID-19 on the attitude towards carbon reduction; thus, we split period0 (before the Wuhan shutdown) from period1. However, there was no significant difference between p0 and p1 or p2, but only between it and p3 (Figure 7).

We also explored whether there were significant differences in sentiment weight among all 96 industries. We grouped the firms’ sentiment results by the industries to which they belonged and conducted a Wilcoxon test to verify each combination of industries. Thus, we obtained 4560 pairs for comparison, and 3122 of them were significant (Table 8), which meant that industries had a significant influence on the sentiment weights of the firms.
There were significant differences in the sentiment weights among industries. The complete results are available at: https://github.com/luyuyuyu/gov_mkt_carbon_nlp/blob/main/by_field.csv (accessed on 3 March 2022).

A total of 4554 rows were omitted, and 3122 of the 4560 comparison groups had significant differences. Significant code: *: \( p \leq 0.05 \); **: \( p \leq 0.01 \); ***: \( p \leq 0.001 \).

Table 8.

| Variable | Group1                      | Group2                               | \( p \)       | \( p.\text{Adj} \) | \( p.\text{Format} \) | \( p.\text{Signif} \) | Method |
|----------|-----------------------------|--------------------------------------|---------------|---------------------|------------------------|------------------------|--------|
| weight   | Internet technology         | Internet business                    | 0.005164      | 1                   | 0.00516                | **                     |        |
| weight   | Internet technology         | Chemical fertilizer and pesticide    | \( 1.74 \times 10^{-10} \) | \( 5.50 \times 10^{-7} \) | \( 1.70 \times 10^{-10} \) | ***                   |        |
| weight   | Internet technology         | New materials                        | 1.13 \times 10^{-28} | 4.40 \times 10^{-25} | <2 \times 10^{-16}    | ***                   |        |
| weight   | Internet technology         | Chemical materials                   | 0.001042      | 1                   | 0.00104                | **                     |        |
| weight   | Internet technology         | Chemical products                    | 0.029401      | 1                   | 0.0294                 | *                      |        |
| weight   | Internet technology         | chemical/ pharmaceutical             | \( 8.32 \times 10^{-7} \) | 0.0023              | \( 8.30 \times 10^{-7} \) | ***                   |        |

Figure 6. The group difference test of the daily average sentiment weight in each period. ns: \( p > 0.05 \); ***: \( p \leq 0.001 \).

The test verified that there was a significant increase in the average sentiment weight after the "Double Carbon" goal was incorporated into the government's work report (\( p_2 \) vs. \( p_3 \)), but not after the goal was set (\( p_1 \) vs. \( p_2 \)), which implied that firms would not change their attitudes only because of the government's goal, but further policies would push them to be significantly more positive in their attitudes toward environmental protection (at least with their attitudes in public).

As shown in Figure 4, we also observed a low level around the Wuhan shutdown, leading us to the influence of COVID-19 on the attitude towards carbon reduction; thus, we split period0 (before the Wuhan shutdown) from period1. However, there was no significant difference between \( p_0 \) and \( p_1 \) or \( p_2 \), but only between it and \( p_3 \) (Figure 7).

Figure 7. There are significant differences of firms’ attitudes between \( p_0 \) and \( p_3 \), \( p_1 \) and \( p_3 \), \( p_2 \) and \( p_3 \). ns: \( p > 0.05 \); ***: \( p \leq 0.001 \).

3.4. Predictive Models of Firms’ Sentiment Weights and Stock Data Based on Advanced Tree Models

We further explored the stock data of all of the firms observed in our attitude dataset to find the relationship of the sentiment weight with other values, such as the stock value, industry, and so on. The stock data came from the Choice dataset from East Money.
We used a linear regression model as the baseline model and a random forest models for further improvement. We split the dataset randomly with a proportion of 7:3, with 7 as the training data and 3 as the test data, in order to assess the performance of each model, as is done with supervised machine learning [24]. Then, we used our best model to rank the importance of the variables. The RMSE indicator was used to assess the models, and our best model was the one that satisfied:

\[
\text{Min (RMSE (predict_result_best_model (train_data), test_data))}
\]  

Then, we improvised the random forest models by optimizing the number of trees [25] and other hyper-parameters [26].

3.4.1. Model 1

Result 1:

As we can see in the table in the Appendix A, unlike with the verification of the observation of the significance of group differences, COVID-19 had a significant influence (in this model, industries were compared with “White household appliances”, and periods were compared with “periodp0”).

\[
\text{model 1: weight = a1 × date + a2 × curret_value + a3 × percentage of increase + a4 × amount of float + a5 × volume + a6 × recent trading volume + a7 × open + a8 × price-earnings ratio.TTM. + a9 × total value + a10 × industry + a11 × percentage of float in 60 days. + a12 × percentage of float in this year. + a13 × period}
\]  

3.4.2. Model 2

We added another variable to check whether a company’s belonging to a technology field had a significant influence on the sentiment weight. The companies in technology industries were marked as 1, the traditional industries as –1, and industries that did not directly produce carbon as 0 (mostly service industries, such as banking and finance; see Table A2 in the Appendix A). Since the variable came from the industries, we removed the industry variable to avoid multicollinearity problems.

\[
\text{model 2: weight = a1 × date + a2 × curret_value + a3 × percentage of increase + a4 × amount of float + a5 × volume + a6 × recent trading volume + a7 × open + a8 × price-earnings ratio.TTM. + a9 × total value + a10 × whether_tech + a11 × percentage of float in 60 days. + a12 × percentage of float in this year. + a13 × period}
\]

Result 2 (Table 9):

In this model, we observed whether technology industries significantly influenced the sentiment weight result. However, R-squared was reduced because we removed the more specific variable—industry.

**Table 9. Results of model 2.**

| Term               | Estimate | Std.Error | t Value | p.Value       | Signif Codes |
|--------------------|----------|-----------|---------|---------------|--------------|
| 1 (Intercept)      | 101.4944 | 13.512    | 7.511426| 5.94 × 10^{−14} | ***          |
| 2 date             | −0.00515 | 0.000748  | −6.89541| 5.43 × 10^{−12} | ***          |
| 3 current_value    | −0.29746 | 0.078357  | −3.79628| 0.000147      | ***          |
| 4 percentage of increase | 0.732461 | 0.034771  | 21.0652 | 4.35 × 10^{−98} | ***          |
| 5 amount of float  | 0.212833 | 0.102069  | 2.085193| 0.037057      | *            |
| 6 volume           | −1.53 × 10^{−8} | 5.04 × 10^{−9} | −3.04076 | 0.002361 | **           |
3.4.3. Stepwise Feature Selection

Before we used the random forest models, we used a forward stepwise selection to assist with the choice of the variables in model 1 (Table 10). There were several variables related to stock, leading to a potential interrelationship within the set of variables. We started from the intercept term, adding a variable in each step according to the contribution to the difference in the AIC after adding it, and we ended with all 13 variables in model 1. The result shows that after adding the other 12 variables, the variable “volume” did not contribute to the model, and should be deleted from the set of variables.

Table 10. Steps of the forward selection.

| Term                      | Estimate | Std.Error | t Value | p.Value | Signif Codes |
|---------------------------|----------|-----------|---------|---------|--------------|
| recent trading volume     | 0.000225 | 4.14 × 10⁻⁵ | 5.447573 | 5.13 × 10⁻⁸ | ***          |
| open                      | 0.308644 | 0.078796  | 3.91703 | 8.98 × 10⁻⁵ | ***          |
| price-earnings ratio.TTM. | 0.004332 | 0.000467  | 9.268343 | 1.96 × 10⁻⁴⁰ | ***          |
| total value               | -2.44 × 10⁻¹² | 1.36 × 10⁻¹² | -1.7983 | 0.072136  | .            |
| percentage of float in 60 days | -0.03591 | 0.002974  | -12.0773 | 1.55 × 10⁻³³ | ***          |
| percentage of float in this year | 0.005856 | 0.000733  | 7.986153 | 1.42 × 10⁻¹⁵ | ***          |
| periodp1                  | 1.643733 | 0.329841  | 4.983406 | 6.27 × 10⁻⁷ | ***          |
| periodp2                  | 3.485751 | 0.456308  | 7.639027 | 2.23 × 10⁻¹⁴ | ***          |
| periodp3                  | 6.488219 | 0.576821  | 11.24823 | 2.56 × 10⁻₂⁹ | ***          |
| whether_tech0             | 1.149543 | 0.362972  | 3.16703 | 0.001541  | **           |
| whether_tech1             | 0.365539 | 0.139087  | 2.628136 | 0.008588  | **           |

Significant codes: .: p ≤ 0.1; *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001.
Table 10. Cont.

| +<none> | 12,656,310 | 281,527 |
| +total_value | 1 | 190 | 12,656,120 | 281,529 |

From the lines above, we found that adding the industry variable to the starting model (weight of ~1) would lead to the best AIC. Thus, the stepwise selection started with a weight of 1 + the field in the next step.

Step 2:

AIC = 276,219.6

| Weight ~ 1 + industry |
|-----------------------|
| Df | Sum of Sq | RSS | AIC |
| + period | 3 | 108,204 | 11,256,388 | 275,737 |
| +%increase_this_year | 1 | 64,285 | 11,300,307 | 275,932 |
| + date | 1 | 59,843 | 11,304,748 | 275,952 |
| +amount of float | 1 | 36,591 | 11,328,000 | 276,057 |
| +open | 1 | 5802 | 11,358,790 | 276,196 |
| +current_value | 1 | 5776 | 11,358,816 | 276,196 |
| +amount_of_float | 1 | 2508 | 11,362,083 | 276,210 |
| +total_value | 1 | 1660 | 11,362,931 | 276,214 |
| +volume | 1 | 940 | 11,363,652 | 276,217 |
| +%increase_60days | 1 | 298 | 11,364,294 | 276,220 |
| +recent_trading_volume | 1 | 170 | 11,364,422 | 276,221 |
| +price-earnings ratio.TTM. | 1 | 114 | 11,364,477 | 276,221 |

From the lines above, we found that adding the period would lead to the best AIC. Thus, the stepwise selection started with a weight of ~1 + the field + the period in the next step.

Several steps were omitted, and the AIC continued improving until the model became weight ~ industry + period + %increase_this_year + %increase + %increase_60days + date + amount_of_float + current_value + open + total_value + recent_trading_volume + price-earnings ratio.TTM.

We can see that in this step, adding the variable “volume” was not better than adding nothing (<none>) according to the AIC. Thus, the stepwise variable selection suggested that we delete the volume variable.

| Df | Sum of Sq | Rss | AIC |
| <none> | 11,104,597 | 275,064 |
| +volume | 1 | 37.37 | 11,104,560 | 275,066 |

3.4.4. Model 3

In the importance ranking (Figure 8) in model 3, we found that the date and the percentage of the float of stock were the two most important variables, while the period ranked last. The reason was that the period variable was related to the date variable, and if the date mostly explained the changes in attitudes, the part left to the period would be less, which was verified when we removed the date from model 3 and ranked the importance of the variables again (Figure 9).

model3: random forest 1(weight) = rf (period, %increase_this_year, %increase, %increase_60days, date, amount_of_float, current_value, open, total_value, recent_trading_volume, price-earnings ratio.TTM.)
ranked last. The reason was that the period variable was related to the date variable, and if the date mostly explained the changes in attitudes, the part left to the period would be less, which was verified when we removed the date from model 3 and ranked the importance of the variables again (Figure 9).

Figure 8. The variable importance ranking in model 3.

Figure 9. The importance ranking after removing the date variable.

3.4.5. Model 4

We added the whether_tech variable to see the importance rank again (Figure 10). However, the stock variables still rank high in the figure. Possible reason can be the industry variable is related with the stock performance.
Model4: random forest 1(weight) = rf (period, %increase_this_year, %increase, %increase_60days, date, amount_of_float, current_value, open, total_value, recent_trading_volume, price-earnings ratio.TTM., whether_tech)

3.4.5. Model 4
We added the whether_tech variable to see the importance rank again (Figure 10). However, the stock variables still rank high in the figure. Possible reason can be the industry variable is related with the stock performance.

Figure 10. The variable importance ranking in model 4.

3.4.6. Prediction and Model Assessment
We used RMSE indicators to combine the model performance of the 4 models (Table 11). A lower RMSE indicates a better prediction result, and a better performance of the model.

Table 11. Prediction results of the four models.

| Model1 (lm1) | Model2 (lm2) | Model3 (rf1) | Model4 (rf2) |
|--------------|--------------|--------------|--------------|
| RMSE         | 13.91493     | 17.63796     | 10.98379     | 12.84664     |

note Baseline
Use the whether_tech variable instead of industry in comparison with model 1.
Cannot use the industry variable, since rf models reject factors with too many levels (96 levels in the industry variable).
Remove the volume variable in model 1 according to the stepwise selection result.
Add the whether_tech variable in comparison with model 3.

According to the RMSE indicators, we found that model 3 had the best performance.

4. Conclusions
Based on the question-and-answer records of Chinese public firms’ investor surveys, this article examined the changes in companies’ attitudes towards carbon reduction before and after the “Double Carbon” policy. First, our descriptive statistical result shows the following.

There was an increasing trend in the frequency of carbon reduction and environmental protection after the “Double Carbon” goal was proposed and incorporated into the government’s work report, indicating a growing keenness on the topic.
Through sentiment analysis methods, we estimated the sentiment weight of each survey record. According to the weight, through the verification of group differences, we observed that:

1. There was a significant increase in firms’ attitudes towards carbon reduction and environmental protection after the “Double Carbon” goal was incorporated into the government’s work report and consequent relevant policies were added, but the same significant increase was not found after the goal was proposed.
2. A strong significance could be observed in the differences in attitude among the industries. A total of 3122 of the 4560 possible pairs for comparison showed a strong significance in the differences in industries’ attitudes towards carbon reduction and environmental protection.
3. The influence of COVID-19 on attitudes was not observed.
4. Then, in the linear regression models, we observed that:
5. Whether a firm is in a technology industry significantly influences the firm’s attitude.
6. Other significantly related variables were stock value, the increase in stock value since the start of the year, and stock data.
7. COVID-19 significantly influenced firms’ attitudes towards carbon reduction and environmental protection, which was different from the findings in the verification of the significance of group differences.

Finally, we applied random forests to attain the most accurate predictive model. Since the sentiments and emotions of humans are so delicate to estimate, with the predictive models based on non-linguistic variables that were constructed, there were more ways to predict, verify, and assess the firms’ attitudes towards ecological topics.

According to the conclusion, our policy advice is:

1. A goal with consequent specific policies can raise the positive attitudes of firms toward carbon reduction topics, but not the goal alone.
2. Firms’ attitudes toward ecological topics are different from industry to industry, which means that there are different needs and situations in the trend of carbon reduction from industry to industry. Detailed policies with differentiation will be more suitable.
3. COVID-19 influences firms’ attitudes toward carbon reduction and environmental protection, calling back the classic dilemma or trilemma of economic growth, carbon reduction, and a third factor, such as energy consumption or epidemic controls today.

Our database is large in scale and rich in content; it can support more research and exploration tasks. In the analytical work of this article, we calculated the changes in attitudes of Chinese public firms after key time points. However, more research can be conducted. For example, attitudes are also potentially influenced by more firm-level factors, such as CSR, the ownership of the firm (state-owned or private), the corporate culture, and the personalities of CEOs. As for the topic of carbon reduction, an LDA topic analysis model can be used to extract a company’s views and to measure in different directions of emission reduction. In the sentiment analysis, this article distinguished between positively and negatively sentimental words, but with a greater extent of the words imported, the results of sentiment analysis can be more accurate; at the same time, the content of the dictionary can also be further improved. Word2Vec, BERT, etc. can be used to build a dictionary based on the topics of carbon reduction and environmental protection. These are also our next directions.

**Author Contributions:** Conceptualization, C.L., Y.Y. and L.L.; methodology, C.L., Y.W. and L.L.; software, L.L. and Y.W.; validation, J.W. and Y.Y.; formal analysis, C.L., W.L., Z.L., J.Z. and L.L.; investigation, J.Z., J.W. and Y.Y.; resources, C.L. and L.L.; data curation, Q.H., J.W. and J.G.; writing—original draft preparation, C.L., L.L., J.Z., J.W., Y.Y., Z.L., Y.W., Q.H., W.L. and J.G.; writing—review and editing, L.L.; visualization, Y.Y. and W.L.; supervision, L.L.; project administration, L.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article. For more details, see https://github.com/luyuyuyu/gov_mkt_carbon_nlp (Accessed on 3 March 2022).

Acknowledgments: Not applicable.

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

Table A1. The results of model 1.

| Term                                  | Estimate  | Std.Error  | t Value  | p. Value     | Signif Codes |
|---------------------------------------|-----------|------------|----------|--------------|--------------|
| (intercept)                           | 122.8467  | 14.4906    | 8.477671 | 2.36 × 10⁻¹⁷ | ***          |
| date                                  | 0.00641   | 0.000802   | -7.99338 | 1.34 × 10⁻¹⁵ | ***          |
| percentage of increase                | 0.620066  | 0.039568   | 15.6709  | 3.22 × 10⁻⁻²⁵ | ***          |
| amount of float                       | 0.803083  | 0.116003   | 6.922919 | 4.48 × 10⁻⁻¹² | ***          |
| volume                                | 2.55 × 10⁻⁹| 6.16 × 10⁻⁹| 0.414112 | 0.678794     | ***          |
| recent trading volume                 | -0.0001   | 4.80 × 10⁻⁵| -2.18349 | 0.029005     | ***          |
| open                                  | 1.084942  | 0.10173    | 10.66495 | 1.58 × 10⁻⁻²⁶ | *            |
| price-earnings ratio,TTM.            | -0.00116  | 0.000548   | -2.1149  | 0.034443     | ***          |
| total value                           | 8.12 × 10⁻¹²| 1.67 × 10⁻¹²| 4.879455 | 1.11 × 10⁻⁻⁶ | *            |

11–105
industry, compare to white household appliances

| Term                                  | Estimate  | Std.Error  | t Value  | p. Value     | Signif Codes |
| semiconductor                         | -1.32583  | 0.932814   | -1.42132 | 0.15523      | ***          |
| glass                                 | 17.30774  | 4.51258    | 3.835443 | 0.000125     | ***          |
| animal husbandry                      | -2.84401  | 0.881095   | -3.22782 | 0.001248     | ***          |
| ship and marine equipment             | 7.157058  | 5.274547   | 1.356905 | 0.174817     | **           |
| motor                                 | 0.878144  | 1.1123     | 0.789484 | 0.429833     | ***          |
| electricity                           | 0.273886  | 0.990088   | 0.27628  | 0.782067     | ***          |
| power supply                          | 3.588364  | 0.908556   | 3.94524  | 7.84 × 10⁻⁻³ | ***          |
| electronic devices                    | -4.19587  | 1.165698   | -3.59945 | 0.00319      | ***          |
| electronic equipment manufacturing    | 2.439399  | 0.810315   | 3.010434 | 0.00261      | **           |
| electronic components                 | -0.6249   | 0.955577   | -0.65395 | 0.513147     | ***          |
| real estate development               | 4.541732  | 2.479098   | 1.83201  | 0.066956     | **           |
| textiles                              | -1.77311  | 4.72651    | -0.37497 | 0.707684     | ***          |
| non-bank finance                      | 0.707548  | 1.879389   | 0.376478 | 0.706563     | ***          |
| clothing and home textiles            | -3.80677  | 1.030464   | -3.69423 | 0.000221     | ***          |
| steel structures                      | -0.10154  | 0.836296   | -0.12141 | 0.903363     | ***          |
| steel                                 | 3.011207  | 0.846367   | 3.557804 | 0.000374     | ***          |
| port shipping                         | 10.49783  | 2.23662    | 4.693615 | 2.69 × 10⁻⁻⁶ | ***          |
| road rail                             | 2.25131   | 2.450983   | 0.918533 | 0.358344     | ***          |
| optoelectronic device                 | -4.2997   | 1.129387   | -3.80711 | 0.000141     | ***          |
| broadcasting                          | -2.6028   | 8.556085   | -0.3042  | 0.760973     | ***          |
| rail transit equipment                | 50.76793  | 5.629349   | 9.018438 | 1.97 × 10⁻¹⁹ | ***          |
| precious metals                       | 9.219086  | 8.562859   | 1.076366 | 0.281648     | ***          |
| aerospace equipment                   | -6.87912  | 1.93049    | -3.56341 | 0.000366     | ***          |
| aviation airport                      | 4.404922  | 3.31229    | 1.329872 | 0.183566     | ***          |
| synthetic fiber and resin             | 10.09359  | 0.884497   | 11.41167 | 3.98 × 10⁻⁻³ | ***          |
Table A1. Cont.

| Term                          | Estimate | Std.Error | t Value | p. Value | Signif Codes |
|-------------------------------|----------|-----------|---------|----------|--------------|
| internet service              | 5.966019 | 1.417503  | 4.208824| 2.57 × 10⁻⁵| ***          |
| internet technology           | 0.699786 | 1.940025  | 0.36071 | 0.718318 |              |
| internet business             | −2.54472 | 7.420536  | −0.34293| 0.731653 |              |
| fertilizers and pesticides    | −0.88422 | 0.850499  | −1.03965| 0.298507 |              |
| new chemical materials        | 17.6047  | 0.926078  | 18.96   | 5.81 × 10⁻⁸⁰| ***          |
| chemical materials            | 2.414427 | 0.819692  | 2.945528| 0.003226 | **           |
| chemicals                     | 5.166898 | 0.796451  | 6.487399| 8.81 × 10⁻¹¹| ***          |
| chemical and pharmaceutical   | −1.78932 | 0.919407  | −1.94617| 0.051639 |              |
| environmental protection      | 21.78406 | 0.951113  | 22.90376| 1.64 × 10⁻¹¹³| ***         |
| robots                        | −1.46364 | 0.923671  | −1.58459| 0.113066 |              |
| basic metal                   | 0.200088 | 0.838384  | 0.23866 | 0.811371 |              |
| infrastructure                | 0.556408 | 1.685168  | 0.330179| 0.741266 |              |
| computer software             | 15.19635 | 0.954484  | 15.92101| 6.22 × 10⁻⁴⁰| ***          |
| computer hardware             | 3.747518 | 1.074415  | 3.48761 | 0.000487 | ***          |
| furniture                     | 0.243515 | 1.166828  | 0.208377| 0.834935 |              |
| building construction         | 17.80526 | 0.9709    | 18.33893| 7.05 × 10⁻⁷⁵| ***          |
| education                     | −1.79009 | 6.644787  | −0.2694 | 0.787624 |              |
| new metal and non-metal materials | 6.083196 | 0.850658  | 7.151168| 8.72 × 10⁻¹³| ***          |
| metal products                | −4.60701 | 0.828141  | −5.6307 | 2.66 × 10⁻⁸ | ***         |
| forestry                      | 3.114934 | 4.517578  | 0.689514| 0.490503 |              |
| retail                        | 1.407416 | 4.511864  | 0.311937| 0.75509 |              |
| trading                       | −3.13583 | 1.604586  | −1.95429| 0.050672 |              |
| coal                          | 1.563853 | 1.48919   | 1.050136| 0.293661 |              |
| refractory                    | 14.41566 | 1.477046  | 9.759788| 1.75 × 10⁻²²| ***          |
| agriculture                   | −4.29808 | 2.232262  | −1.92543| 0.054181 |              |
| print media                   | 2.806752 | 14.7815   | 0.189833| 0.849402 |              |
| other electrical equipment     | 3.493211 | 1.24306   | 2.810171| 0.004953 | **           |
| other home appliances          | 1.133201 | 1.823094  | 0.621581| 0.53422 |              |
| other building materials       | −0.18492 | 0.978066  | −0.18907| 0.85004 |              |
| other delivery equipment       | 0.859954 | 1.894778  | 0.453855| 0.649935 |              |
| other light industry           | −0.46941 | 5.272863  | −0.08902| 0.929063 |              |
| car                           | 1.765503 | 0.784958  | 2.249169| 0.024506 | *            |
| gas                           | 0.629026 | 1.605851  | 0.391709| 0.695275 |              |
| commercial property management | −0.98464 | 4.160513  | −0.23666| 0.81292 |              |
| biomedicine                   | −4.41707 | 2.675452  | −1.65096| 0.098753 |              |
| petroleum gas                  | 6.115942 | 0.941699  | 6.49458 | 8.40 × 10⁻¹¹ | ***     |
| food                          | −3.0498 | 0.971268  | −3.14002| 0.00169 | **           |
| audiovisual equipment         | 6.564379 | 1.479641  | 4.436468| 9.16 × 10⁻⁶ | ***     |
| transmission and transformation equipment | 2.571072 | 0.966026 | 2.677469 | 0.00742 | **          |
| cement                        | 2.756316 | 1.347438  | 2.045979| 0.040801 | *           |
| water affairs                 | 2.879128 | 1.696806  | 1.696793| 0.089742 |              |
| ceramics                      | 6.746263 | 1.543292  | 4.371346| 1.24 × 10⁻³ | ***     |
| Term                        | Estimate | Std.Error | t Value | p. Value | Signif Codes |
|-----------------------------|----------|-----------|---------|----------|--------------|
| iron ore                    | 6.121013 | 5.631601  | 1.086904| 0.277084 |              |
| railway equipment           | 5.180382 | 2.677895  | 1.934498| 0.053058 |              |
| communication devices       | 3.312028 | 1.418248  | 2.335295| 0.019532 | *            |
| general equipment           | 1.097738 | 0.770044  | 1.425552| 0.154004 |              |
| satellite applications      | 2.709181 | 1.886708  | 1.43593 | 0.151028 |              |
| entertainment supplies      | -2.35337 | 2.753838  | -0.85465| 0.392749 |              |
| logistics                   | -1.19681 | 1.069861  | -1.11866| 0.263291 |              |
| rare metals                 | -1.58956 | 1.16583   | -1.36346| 0.172743 |              |
| rubber products             | 2.218447 | 1.117555  | 1.98263 | 0.047081 | *            |
| consumer electronics        | -3.93873 | 1.951649  | -2.01816| 0.04358 | *            |
| home appliances             | 2.522429 | 1.246603  | 2.023442| 0.043033 | *            |
| leisure service             | 0.435217 | 6.645008  | 0.065495| 0.94778  |              |
| medical service             | -4.71585 | 2.986223  | -1.5792 | 0.114296 |              |
| medical instruments         | -4.50512 | 0.998439  | -4.51216| 6.43 × 10⁻⁸| ***          |
| pharmaceutical business     | -6.10541 | 1.505683  | -4.05491| 5.02 × 10⁻⁵| ***          |
| banking                     | -3.27578 | 1.567455  | -2.08988| 0.036634 |              |
| drinks                      | -2.36368 | 4.164547  | -0.5675 | 0.570529 |              |
| marketing service           | 1.954179 | 2.424267  | 0.806091| 0.420194 |              |
| movies and animation        | -4.28238 | 4.328788  | -0.98928| 0.322531 |              |
| fishery                     | -1.3077  | 8.590432  | -0.15223| 0.879008 |              |
| paper printing              | 1.610445 | 0.896266  | 1.79638 | 0.072567 |              |
| lighting devices            | 2.220598 | 1.615952  | 1.374173| 0.169394 |              |
| traditional Chinese medicine production | -4.44179 | 1.487589 | -2.9859 | 0.002829 | **           |
| jewelry                     | -7.88081 | 3.236608  | -2.4349 | 0.014899 |              |
| professional service        | 8.09088  | 0.875021  | 9.24693 | 2.41 × 10⁻²⁰| ***          |
| professional setting        | -1.25212 | 0.794771  | -1.57545| 0.11516  |              |
| decoration                  | 2.726903 | 1.62272  | 1.680452| 0.092876 |              |
| comprehensive               | 1.854743 | 1.557577  | 1.190787| 0.233743 |              |

| 106 percentage of float in 60 days | -0.03486 | 0.003465 | -10.0612 | 8.63 × 10⁻²⁴ | ***          |
| 107 percentage of float in this year | 0.015725 | 0.000851 | 18.4794  | 5.72 × 10⁻⁷⁶ | ***          |

| periods 108–110, compare to p0 | 2.419141 | 0.353387 | 6.845588 | 7.70 × 10⁻¹² | ***          |
| periodp2                      | 4.221345 | 0.487625 | 8.656947 | 4.98 × 10⁻¹⁸ | ***          |
| periodp3                      | 7.633121 | 0.612796 | 12.35541 | 5.11 × 10⁻³⁵ | ***          |

Significant codes: ·: p > 0.1; :p ≤ 0.1; *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001

Table A2. The whether_tech variable and corresponding industries.

| Name                             | Whether_Tech |
|----------------------------------|--------------|
| 1 banking                        | 0            |
| 2 glass                          | -1           |
| 3 audiovisual equipment          | 1            |
| 4 other building materials       | -1           |
| 5 electricity                    | -1           |
Table A2. Cont.

| Name                              | Whether_Tech |
|-----------------------------------|--------------|
| trading                           | 0            |
| environmental protection          | 1            |
| real estate development           | -1           |
| metal products                    | -1           |
| animal husbandry                  | -1           |
| electronic devices                | 1            |
| building construction              | -1           |
| basic metal                        | -1           |
| commercial property management    | 0            |
| electronic components             | 1            |
| chemical and pharmaceutical       | 1            |
| professional setting               | 1            |
| synthetic fiber and resin         | -1           |
| white goods                       | -1           |
| car                               | 1            |
| transmission and transformation equipment | -1   |
| cement                            | -1           |
| gas                               | -1           |
| chemical materials                | -1           |
| internet service                  | 1            |
| logistics                          | -1           |
| road rail                         | -1           |
| paper printing                    | -1           |
| infrastructure                    | -1           |
| port shipping                     | -1           |
| new metal and non-metal materials | 1            |
| food                              | -1           |
| general equipment                 | -1           |
| traditional Chinese medicine production | 1   |
| water affairs                     | -1           |
| coal                              | -1           |
| fertilizers and pesticides        | -1           |
| petroleum gas                     | -1           |
| drinks                            | -1           |
| rubber products                   | -1           |
| power supply                      | -1           |
| forestry                          | -1           |
| medical service                   | 0            |
| non-bank finance                  | 0            |
| steel                             | -1           |
| rare metals                       | 1            |
Table A2. Cont.

| Name                               | Whether_Tech |
|------------------------------------|--------------|
| 47 aerospace equipment             | 1            |
| 48 professional service             | 0            |
| 49 retail                           | 0            |
| 50 biomedicine                      | 1            |
| 51 new chemical materials           | 1            |
| 52 comprehensive                   | -1           |
| 53 textile                          | -1           |
| 54 chemicals                        | -1           |
| 55 agriculture                      | -1           |
| 56 broadcasting                     | 0            |
| 57 motors                           | -1           |
| 58 railway equipment                | -1           |
| 59 computer hardware                | 1            |
| 60 computer software                | 1            |
| 61 pharmaceutical business          | 0            |
| 62 electronic equipment manufacturing| 1            |
| 63 iron ore                         | -1           |
| 64 clothing and home textiles       | -1           |
| 65 decoration                       | -1           |
| 66 refractory                       | -1           |
| 67 semiconductors                   | 1            |
| 68 communication devices            | 1            |
| 69 other delivery equipment         | -1           |
| 70 marketing service                | 0            |
| 71 steel structures                 | -1           |
| 72 precious metals                  | -1           |
| 73 leisure service                  | 0            |
| 74 ceramics                         | -1           |
| 75 education                        | 0            |
| 76 movies and animation             | 0            |
| 77 entertainment supplies           | -1           |
| 78 other electrical equipment       | -1           |
| 79 medical instruments              | 1            |
| 80 optoelectronic devices           | 1            |
| 81 rail transit equipment           | -1           |
| 82 furniture                        | -1           |
| 83 home appliances                  | -1           |
| 84 robots                           | 1            |
| 85 other light industry             | -1           |
| 86 lighting devices                 | -1           |
| 87 jewelry                          | -1           |
Table A2. Cont.

| Name                                      | Whether_Tech |
|-------------------------------------------|--------------|
| 88 consumer electronics                   | –1           |
| 89 aviation airport                       | 1            |
| 90 ship and marine equipment              | 1            |
| 91 satellite application                  | 1            |
| 92 fishery                                | –1           |
| 93 other home appliances                  | –1           |
| 94 internet business                      | 0            |
| 95 internet technology                    | 1            |
| 96 print media                            | 0            |

References

1. Xing, L.; Shi, L.; Hussain, A. Corporations response to the energy saving and pollution abatement policy. *Int. J. Environ. Res.* 2010, 4, 637–646.
2. Liu, Y.; Zhang, Z. How does economic policy uncertainty affect CO₂ emissions? A regional analysis in China. *Environ. Sci. Pollut. Res.* 2022, 29, 4276–4290. [CrossRef] [PubMed]
3. Yu, J.; Shi, X.; Guo, D.; Yang, L. Economic policy uncertainty (EPU) and firm carbon emissions: Evidence using a China provincial EPU index. *Energy Econ.* 2021, 94, 105071. [CrossRef]
4. Zhao, M.; Lü, L.; Zhang, B.; Luo, H. Dynamic Relationship among Energy Consumption, Economic Growth and Carbon Emissions in China. *Res. Environ. Sci.* 2021, 34, 1509–1522.
5. Nawaz, M.A.; Hussain, M.S.; Kamran, H.W.; Ehsanullah, S.; Maheen, R.; Shair, F. Trilemma association of energy consumption, carbon emission, and economic growth of BRICS and OECD regions: Quantile regression estimation. *Environ. Sci. Pollut. Res. Res.* 2021, 28, 16014–16028. [CrossRef] [PubMed]
6. Li, P.; Ouyang, Y. Quantifying the role of technical progress towards China’s 2030 carbon intensity target. *J. Environ. Plan. Manag.* 2021, 64, 379–398. [CrossRef]
7. Liu, X.; Ji, Q.; Yu, J. Sustainable development goals and firm carbon emissions: Evidence from a quasi-natural experiment in China. *Energy Econ.* 2021, 103, 105627. [CrossRef]
8. East Money Website. Available online: https://data.eastmoney.com/jgdy/tj.html (accessed on 29 December 2021).
9. Li, K.; Mai, F.; Shen, R.; Yan, X. Measuring corporate culture using machine learning. *Rev. Financ. Stud.* 2021, 34, 3265–3315. [CrossRef]
10. Malhotra, S.; Reus, T.H.; Zhu, P.; Roelofsen, E.M. The acquisitive nature of extraverted CEOs. *Adm. Sci. Q.* 2018, 63, 370–408. [CrossRef]
11. Turney, P.D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv* 2002, arXiv:0212032.
12. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. *arXiv* 2002, arXiv:0205070.
13. Cambria, E.; Schuller, B.; Xia, Y.; Havasi, C. New avenues in opinion mining and sentiment analysis. *IEEE Intell. Syst.* 2013, 28, 15–21. [CrossRef]
14. Ortony, A.; Clore, G.L.; Collins, A. *The Cognitive Structure of Emotions*; Cambridge University Press: Cambridge, UK, 1990.
15. The Natural Language Processing Group at the Department of Computer Science and Technology, Tsinghua University (THUNLP). Available online: http://rlp.csai.tsinghua.edu.cn/site2/index.php/13-sms (accessed on 29 December 2021).
16. Xu, L.; Lin, H.; Pan, Y.; Ren, H.; Chen, J. Constructing the affective lexicon ontology. *J. China Soc. Sci. Tech. Inf.* 2008, 27, 180–185. (In Chinese)
17. Yu, J.; Yin, J.; Fei, S. Identifying Synonyms Based on Sentence Structure Analysis. *Data Anal. Knowl. Discov.* 2013, 29, 35–40. [CrossRef]
18. Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32. [CrossRef]
19. Heinrich, V.H.; Dalagnol, R.; Cassol, H.L.; Rosan, T.M.; de Almeida, C.T.; Silva Junior, C.H.; Campanharo, W.A.; House, J.I.; Sitch, S.; Hales, T.C.; et al. Large carbon sink potential of secondary forests in the Brazilian Amazon to mitigate climate change. *Nat. Commun.* 2021, 12, 1785. [CrossRef] [PubMed]
20. Machine Learning Mastery. Available online: https://machinelearningmastery.com/much-training-data-required-machine-learning/ (accessed on 7 March 2022).
21. GitHub. Available online: https://github.com/luyuyuyu/gov_mkt_carbon_nlp/blob/main/raw_data (accessed on 7 March 2022).
22. The World Bank. 2021. World Development Indicators. Available online: https://databank.worldbank.org/source/world-development-indicators (accessed on 16 December 2021).

23. GitHub. Available online: https://github.com/luyuyuyu/gov_mkt_carbon_nlp/blob/main/clean.zip (accessed on 3 March 2022).

24. Russell, S.; Norvig, P. Artificial Intelligence: A Modern Approach, 2nd ed; Prentice Hall: Hoboken, NJ, USA, 2003.

25. Oshiro, T.M.; Perez, P.S.; Baranauskas, J.A. How Many Trees in a Random Forest? Springer: Heidelberg/Berlin, Germany, 2012; pp. 154–168.

26. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev. Data Min. Knowl Discov. 2019, 9, e1301. [CrossRef]