

## *Retraction*

# **Retracted: Analysis and Application of the Ideological and Political Evaluation System of College Students Based on Text Mining**

### **Journal of Sensors**

Received 22 August 2023; Accepted 22 August 2023; Published 23 August 2023

Copyright © 2023 Journal of Sensors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] J. Geng, T. Wang, and S. Zhang, "Analysis and Application of the Ideological and Political Evaluation System of College Students Based on Text Mining," *Journal of Sensors*, vol. 2022, Article ID 4263974, 14 pages, 2022.

## Research Article

# Analysis and Application of the Ideological and Political Evaluation System of College Students Based on Text Mining

Jing Geng,<sup>1</sup> Ting Wang,<sup>2</sup> and Shunyou Zhang<sup>3</sup> 

<sup>1</sup>Baoding Vocational and Technical College, Baoding, Hebei 071000, China

<sup>2</sup>School of Art, Hebei College of Science and Technology, Baoding, Hebei 071000, China

<sup>3</sup>Hebei Institute of Communications, Shijiazhuang, Hebei 051430, China

Correspondence should be addressed to Shunyou Zhang; zhangsy@hebic.edu.cn

Received 29 April 2022; Revised 3 June 2022; Accepted 20 June 2022; Published 29 July 2022

Academic Editor: Yuan Li

Copyright © 2022 Jing Geng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under the background of the new era, the ideological and political theory courses in universities are the key courses to cultivate people by virtue. It is very important for college students in their own ideological and political construction and consciousness. At the present stage, there is little research on the ideological and political current situation of college students, most of them improve the teaching program from the teacher level, and there is no research on the ideological and political evaluation system of college students. In order to understand the current situation of the ideological and political evaluation of college students, we will study it from the two aspects of the school and the society. This paper excavates and analyzes the ideological and political aspects of college students and those in the society: first of all, the ideological and political evaluation of teachers on large social platforms such as Zhihu and Weibo and the moral quality and political consciousness of college students in the society. It was then processed and analyzed using the Python language. Draw the word frequency and word cloud map of the keywords in the evaluation for analysis. Then, use the text data preprocessing method based on the word frequency statistics law (Data Preprocessing Based on Term Frequency Statistics Rules (DPTFSR)). Processing the text data and finally conducting the relevant emotional analysis show the university ideological and political system to understand the ideological and political situation of college students in the new era and to improve the ideological and political education program according to its performance.

## 1. Introduction

In some contemporary research on contemporary ideological and political education, it shows that it needs to explore [1] from the perspective of systematic integration in the process of promoting the high-quality ideological and political development of college students. The moral, intellectual, physical, aesthetic, and labor qualities of college students in daily life are a manifestation of the ideological and political consciousness of college students, and the ideological and political evaluation of college students is influenced by multiple factors. For example, the attraction of university ideological and political theory courses to college students is related to the effectiveness of ideological and political education to a certain extent [2]. In the era of smart education, ideological and political education should be tailored to things, progressive and new according to the situation

[3–5]. Its difficulty is due to current affairs about the ideological and political changes of college students, and the clear perception of this change is teachers. And the comments on social media platforms are [6] for the quality of ideological and political courses. This is related to the status of ideological and political construction.

In fact, in the thousands of evaluations, teachers and society show the ideological and politics of college students from different perspectives, and extracting useful information from a large number of comments is helpful to the objective evaluation of the ideological and politics of college students. Therefore, this paper uses the text mining technology [7–9] to mine and analyze the evaluation of college students and gives reasonable suggestions for ideological and political education according to the results of the analysis system. This helps to reflect the recognition of college students' ideological and political courses, college students'

recognition of teaching programs and content, and the influence of teachers' acceptance of ideological and political courses. First, the text positioned the data objects to be collected and analyzed, and it was found that the question and answer method of the Zhihu website [10] is very helpful for exploring the ideological and political evaluation of college students. Therefore, the data source of this article is the Zhihu website. Firstly, the crawler technology of python is used to collect the data of questions and answers related to the Zhihu website, and then, the data is preprocessed, because only the most basic word segmentation methods and stop words can be completed in the current preprocessing stage of literature mining. The utilization of huge data is also low. Therefore, in this paper, the text data preprocessing method based on the statistical law of word frequency is first carried out [11]. Then, after the low-frequency words are removed in the preprocessing stage, the word segmentation method and the stop words of the basic data are completed. The basic data after pretreatment are obtained. Finally, Python [12] is used again to draw the word frequency map and word cloud map of the preprocessed data for keyword analysis, and emotional analysis technology [13, 14] is used to obtain the social evaluation of contemporary college students' ideology and politics. It is found that the evaluation mainly comes from several aspects, including the sense of honor of the whole country and society, personal values, and moral values.

## 2. Data Collection

Search for questions about topics related to "college students' ideological and political evaluation" on the homepage of Zhihu. There are about 300 questions, and the number of answers below each question is more than a few hundred and less than three or four, and there is a personal evaluation of this answer below each answer. The number of texts is very complex and huge, so this paper will choose three extended topics, such as "contemporary college students' outlook on life and values," "contemporary college students' ideological and political status," and "current college students' personal moral quality evaluation." Because these questions are the direct display of college students' ideological and political consciousness facing the society, they have a direct and valuable reference for understanding college students' ideological and political consciousness. As Zhihu, Weibo, massive open online course, and other website platforms are the main routes for online comment exchange in China, this writing platform can freely discuss and publish its own opinions. Compared with the usual text information, the text of the website has its unique characteristics, and the process of its integration and analysis is more complex. Therefore, it is important to carry out useful means to obtain and sort out useful information from complex website data. Zhihu is a question-and-answer website, on which users can post questions and invite other users to answer. The answers are not standardized. The class will give comments according to their own feelings and thoughts, and other people can also comment and praise under the reply content of the respondent to express different or the same views with

the respondent. This model is more friendly and concentrated, which reflects the social collective's reflection on a certain problem. Weibo is where one person publishes content and others evaluate his views. And there is a need for text collection of data as much and detailed as possible; the model of this platform is very good in line with this demand. Therefore, this paper will collect and analyze the evaluation of related topics of "ideological and political evaluation of college students" in Zhihu.

Due to the wide variety of text mining technologies on the market, in order to find the text mining technologies with high efficiency and high accuracy, the three common mining algorithm technologies first compare the relevant content classification effect of the ideological and political evaluation of college students, and the most suitable and the most relevant one is selected. These three methods are the simple Bayesian method,  $K$ -nearest adjacency reference classification algorithm, and first-order rule learning. More mainstream text mining techniques are shown in Figure 1.

To understand how these methods compare different types of web classification efficiency, the tables in three databases are used. Since the data information of some companies and schools is externally visible, this data can be used first to compare the efficiency of several algorithms. First, the database information of several companies and schools was tested comprehensively, namely, Hoovers28, Hoovers255, and Univ6. For a single web page, associated words, and marker language, the results are shown in Table 1.

Then, compare the runtime with similar recall and precision weight again, shown in Table 2.

The appeal data are plotted as a line of Figures 2 and 3.

These performances are related to the kinds of problems to be dealt with. For school-related web pages, these three methods are not very good for classifying single pages, while for the other two data, the naive Bayes method and  $K$ -nearest neighbor reference classification algorithm are significantly less efficient for handling related words and html titles. Finally, we found that these methods are not applicable to the classification query of Zhihu webpage information, so we will implement another method next.

After determining the web page to be collected, we will collect the data in the web page as a web crawler. In general, there are two ways: one is dedicated crawler software, and the other is to write a code script in the programming language to collect data. In addition, there are many programming languages that can realize the crawler, such as Python, C++, and Java, while this paper uses simple and flexible Python to realize the data collection, and the specific structure framework is shown in Figure 4.

First of all, log into Zhihu to determine the URL of the three questions: "outlook on life, values," "ideological and political status of contemporary college students," and "evaluation of personal moral quality of current college students", and check the answers under each question. The data in the network is basically based on the HTTP protocol, and the data is generally stored in the HTNML web page tags. The flow chart of crawler capture on the Zhihu website based on the ideological and political evaluation of college students is shown in Figure 5.

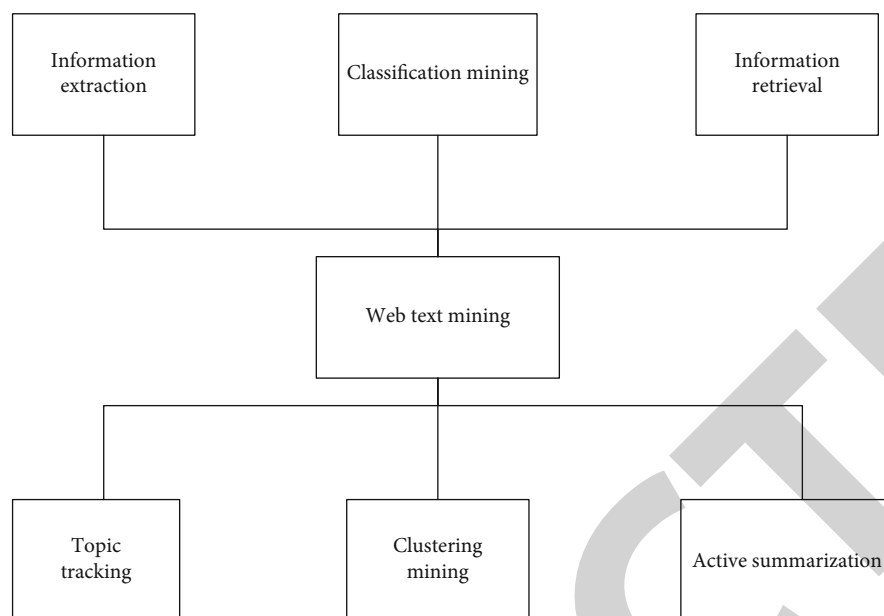


FIGURE 1: Web text mining techniques.

TABLE 1: Test results for the different algorithms under the three databases.

Attribute	Hoovers28			Hoovers255			Univ6		
	NB	KNN	FOIL	NB	KNN	FOIL	NB	KNN	FOIL
Single page	55.1	58.1	31.5	32.5	32	11.6	69.6	83	82.7
Related words	40.1	38.6		18.9	20.4		74.1	86.2	
Tagged words	49.2	49	31.8	24	26.9	12.1	76.3	88	86
HTML title	40.8	43.3	28.7	17.9	22.6	11.5	78.6	81.5	86.3

TABLE 2: Average runtime of the CPU under the three algorithms.

Attribute	Single page			Related words			Tagged words		
	NB	KNN	FOIL	NB	KNN	FOIL	NB	KNN	FOIL
Hoovers28	1.5	5	288.4	2	23		2.3	12	304.3
Hoovers255	3	2.5	315.6	3.5	3.1		3.5	3.9	274.5
Univ6	0.07	0.97	2.8	0.1	2.8		0.27	1.3	10.25

To crawl data, you need to send a Get request to HTTP in advance, and the server will return a Response object after receiving the request. There are many libraries of functions of methods implementing this request in Python such as `re`, `urllib`, and `requests` [15]. This article uses the request library with the GET method in Python. `request.get()` obtains the main information of the web page and makes the GET request for HTTP for obtaining the data through the specified URL. The full parameters of the Get method are as `Requests.get(url, params=None, **kwargs)`.

The url is the url connection of the CNKI question page, `params` is another data in the connection, and `**kwargs` is the 12 empty visit parameters. Create a request object that requests content from the server, and return a request object to represent the server's response, as shown in Table 3.

Finally, we saved the data collected by the crawler as an `xlsx` format d file for later data preprocessing.

### 3. Data Preprocessing

In the digital age, the inherent characteristic of text mining is the scarcity of valuable data. The conflict between a large number of words and a small number of key features results in the lack of effectiveness when a large number of useful information is mentioned and also limits the effect of literature mining, so it is very important to do a good job of efficient data preprocessing before effective analysis of literature information. In order to improve the high performance of the analysis results, the text preprocessing method in this paper is the text data preprocessing method (Data Preprocessing Based on Term Frequency Statistics Rules

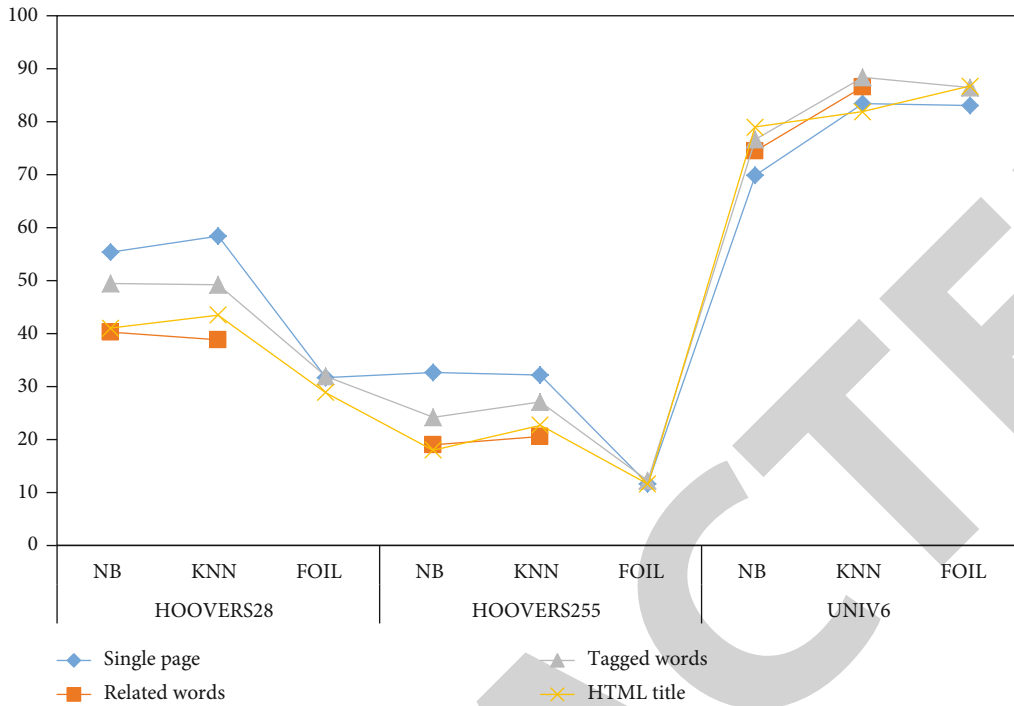


FIGURE 2: Test results of the different algorithms under the three databases.

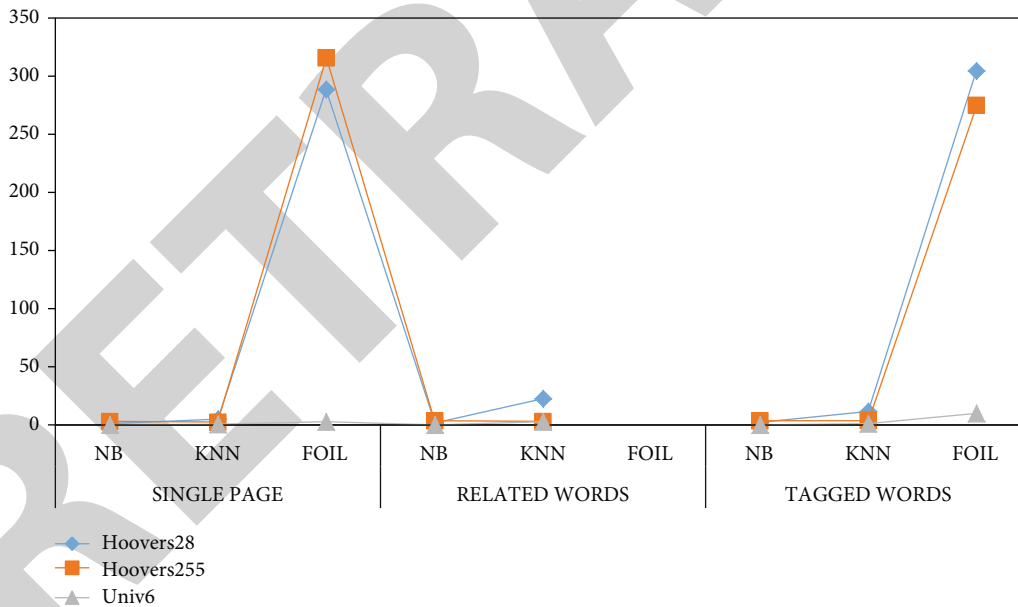


FIGURE 3: Average runtime of the CPU under the three algorithms.

(DPTFSR)) based on the word frequency statistical law, and then, the collected text data is preprocessed by deleting, segmenting, and removing stop words. The basic step is to first initialize the dictionary dict1, dict2, and dict3 and the corresponding storage word frequency of  $TF_n = 1$ ,  $TF_n = 2$ , and  $TF_n > 2$  counters count1, count2, and count3, define the word list TermList and the counter word\_count, and then perform the word separation operation and record the word frequency of each word. Then, classify them according to

different word frequencies, and record the number of words in each frequency. Then, the data is preprocessed based on the word frequency statistics; finally, the different word frequency word sets, corresponding total number of words, and preprocessing list are returned. The following shows the number of appearances of the same word represents the number of appearances of the word  $i$  in the text document  $TF_n$ ,  $n$  is how much of the word appears, while  $TF_k$  (ti) represents the word frequency of the word  $t_i$  which is  $k$

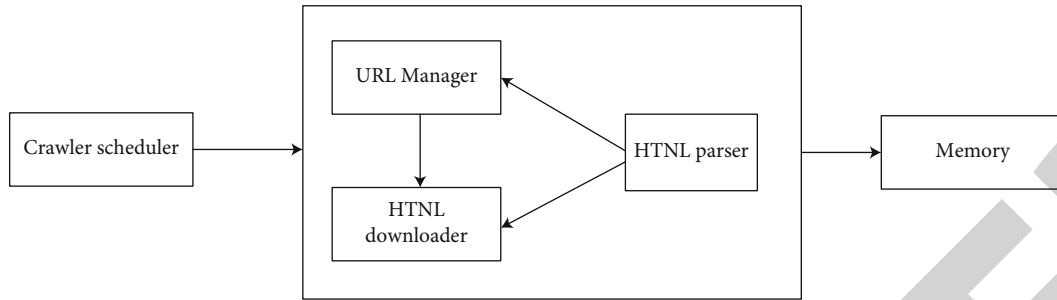


FIGURE 4: The Python crawler frame.

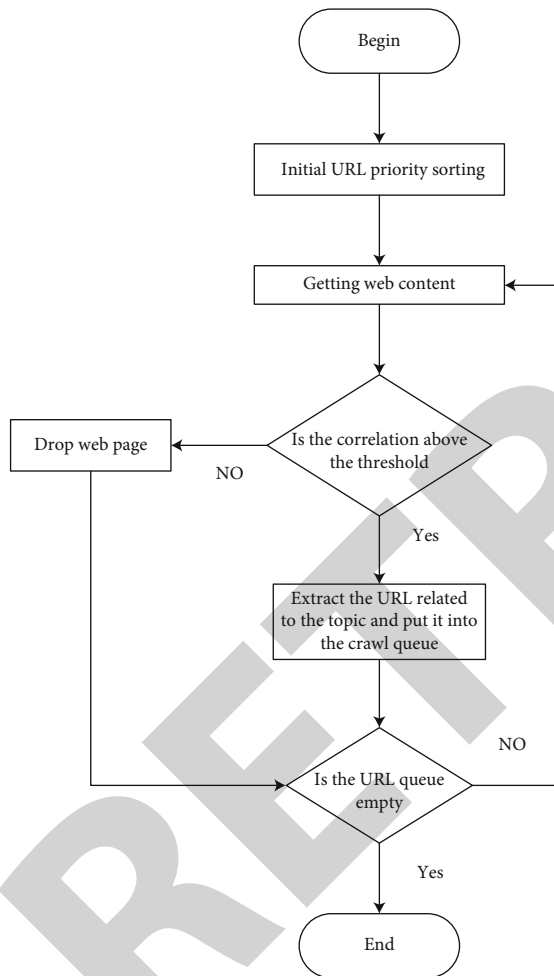


FIGURE 5: Network crawler grasping process.

. The number of words at the same frequency is recorded as  $NTIF_n$ . Using a set of words  $\{ti\}$  in the document, you can meet the expression  $TF_k(t1) = TF_k(t2) = TF_k(tq) = TF_k = TF_k = k$ , so  $ti$  is a set of the same word frequency  $k$ ; the total number is  $q$ . Frequency refers to the frequency produced in the article and the ratio between the frequency and the length of the article, expressed as  $f_n = TF_n/L$ . The word rank is expressed as a  $TR_n$ . When  $TF_k = TF_n$ ,  $TR_k = 1$ ; when  $TF_k = a$  moment,  $TR_k = TR_n$ . You can also use Zipf's

law [16, 17] to push the expression of the same frequency word number, which is one of the three widely used laws of computation in the field of text mining techniques [18].

$$f_n \cdot TR_n = K (K \text{ is a constant}), \quad (1)$$

$$TF_n \leq f_n \cdot L < TF_{n+1}, \quad (2)$$

where  $K$  in formula (1) is a nondeterministic value,  $K = 1/(\ln N_{diff} + 1)$ , changing [19] around a central fluctuation. The following expression for  $NTIF_n$  can be derived from

$$NTIF_n = \frac{K \cdot L}{TF_n \cdot TF_{n+1}}. \quad (3)$$

Put formula (1) into formula (3) to get

$$TF_n \leq \frac{K \cdot L}{TR_n} < TF_{n+1}. \quad (4)$$

Formula (5) is obtained:

$$\begin{cases} TR_{n_{max}} = \frac{K \cdot L}{TF_n}, \\ TR_{n_{min}} = \frac{K \cdot L}{TF_{n+1}}. \end{cases} \quad (5)$$

The number of common frequency words  $NTIF_n$  of the word frequency  $TF_n$  satisfies

$$NTIF_n = TR_{n_{max}} - TR_{n_{min}}. \quad (6)$$

Finally, the above expression can derive the expression of the required number of the same frequency words:

$$NTIF_n = \frac{K \cdot L (TF_{n+1} - TF_n)}{TF_n \cdot TF_{n+1}}. \quad (7)$$

TABLE 3: Common properties of the response object.

Attribute	Describe
r.status_code	Return status of the HTTP request, 200 means a connection is successful, 404 means a connection failure
r.context	Binary form of the HTTP response content
r.encoding	Guess the response encoding mode from the HTTP headers
r.apparent_encoding	Response content encoding method derived from the content analysis
r.text	The string form of the HTTP response content, the url corresponds to the page content

The maximum method is used to refine expressions for the number of words of the same frequency, because NTI  $F_n$  does not fully apply to the case where the word frequency  $TF_n$  takes any value, because it is based on Zipf's law, but Zipf's law does not well reflect the distribution of words with very low word frequency. Next, the maximum value method is used to perfect the expression of  $TF_n = 1, 2$  (formula (8)) same frequency words, where  $TR_n$  and word frequency  $TF_n$  are the corresponding reverse order relationship; the maximum value is sent to determine  $TR_n$ ; the words rank the same with the same, and select the largest word rank; then, the difference between the two adjacent words is

$$NTIF_n = TR_n - TR_{n+1}. \quad (8)$$

And the expression of the number of words with the same word frequency when the word frequency is  $TF_n = 1$  and 2, which are obtained according to the maximum value method, is

$$NTIF_n = \frac{N_{\text{diff}}}{TF_n \cdot TF_{n+1}}, \quad n = 1, 2. \quad (9)$$

$N_{\text{diff}}$  is the total number of different words that appear in the text document. Then, expression (1) yields

$$TR_n = \frac{K}{f_n} = \frac{K \cdot L}{TF_n}, \quad (10)$$

$$TR_n = \frac{K}{f_{n+1}} = \frac{K \cdot L}{TF_{n+1}}. \quad (11)$$

You can find that the expressions of (10), (11), and (8) can derive expression (3). Statistical observation of the whole data shows the results of word frequency and word rank multiplication approach  $N_{\text{diff}}$  when the word frequency equals 2.

$$TR_n \cdot TF_n = N_{\text{diff}}, \quad n = 1, 2. \quad (12)$$

Combine  $N_{\text{diff}}$ 's expressions with (8) for

$$NTIF_n = \frac{N_{\text{diff}}(TF_{n+1} - TF_n)}{TF_n \cdot TF_{n+1}} = \frac{N_{\text{diff}}}{TF_n \cdot TF_{n+1}}, \quad n = 1, 2. \quad (13)$$

Then, the full expression of  $NTIF_n$  obtained from the joint same-frequency words (3) and (9) is as follows:

$$TIF_n = \begin{cases} \frac{K \cdot L}{TF_n \cdot TF_{n+1}}, & n > 2, \\ \frac{N_{\text{diff}}}{TF_n \cdot TF_{n+1}}, & n = 1, 2. \end{cases} \quad (14)$$

According to the obtained processing method of text data, the word set of different word frequency and the corresponding total number of word frequency were counted. Then, we preprocessed the collected word data and finally obtained the preprocessed data list. Figure 6 presents a text mining flow chart based on the word frequency statistical law.

#### 4. Mining and Analysis Process of Ideological and Political Evaluation of College Students

For the data of the ideological and political evaluation text of college students that has been handled well, the value information in the mining and analysis is mainly the analysis of ideological and political keywords and the analysis of college students' emotional tendency. Since the ideological and political performance of college students can be shown from several aspects, according to the top 10 ideological and political evaluation of college students online, we can see the social concern for the ideological and political performance of universities and also reflect the importance of the society to the ideological and political performance of college students from the side. Then, for the positive and negative evaluation of college students' ideological and politics, we summarize what aspects of the ideological and political contemporary universities are missing and what aspects meet the requirements. Through the comparison of the map, we can intuitively see the comparison of the two-level ratings.

**4.1. Keyword Analysis.** The analysis of key words is conducted with word cloud map and word frequency map. For the preprocessed evaluation data, the Jieba package can be called out directly after using the word frequency command to complete the word frequency statistics. Then, the relatively high words frequently appear in the whole evaluation library, and then, the statistical results of the word cloud library are used and visualized to draw the relevant word cloud map and word frequency map, as shown in Figure 7.

As can be seen from Figure 7, in the text data collected, among the words "mission, sense of gain, positive,"

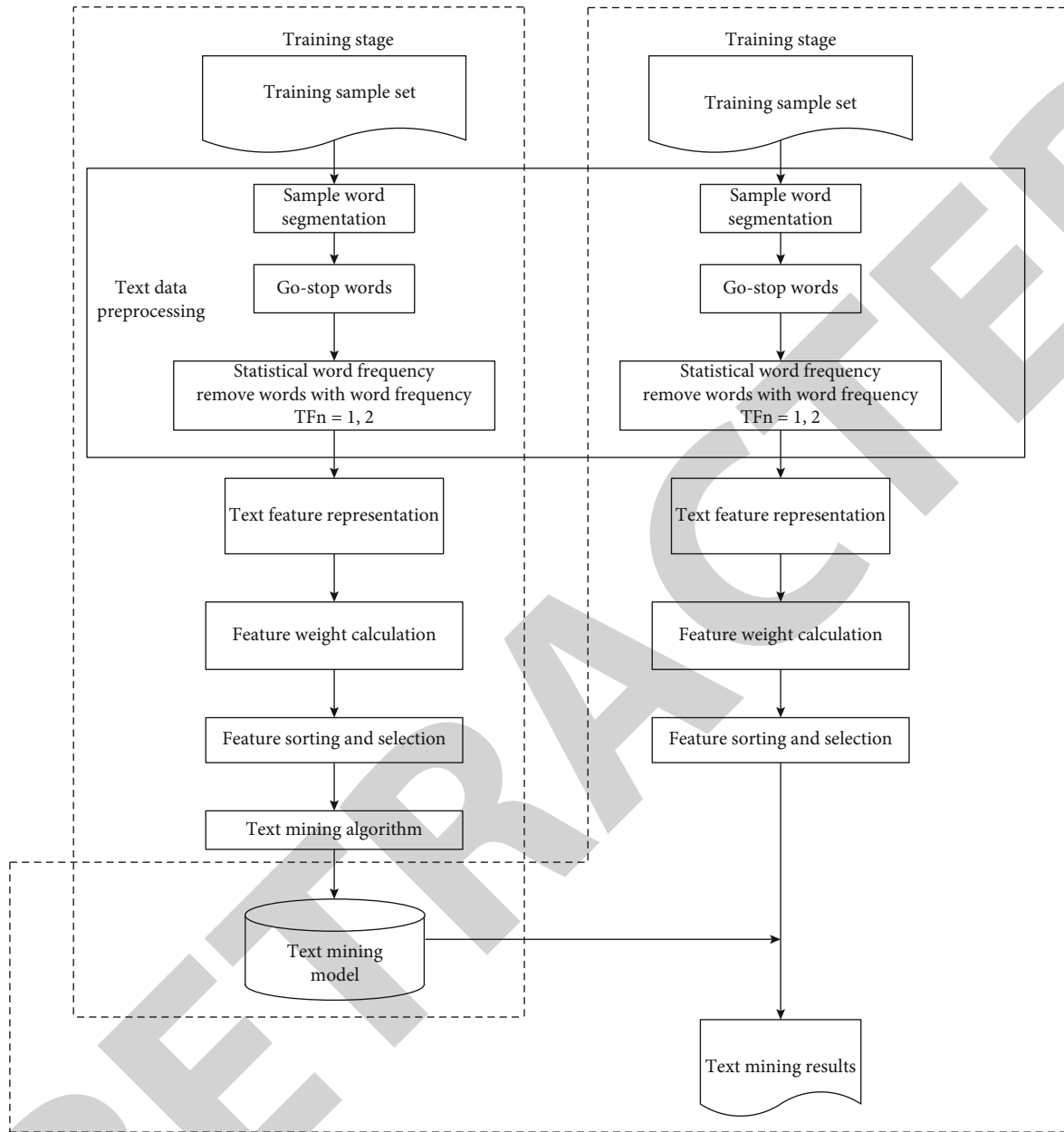


FIGURE 6: Statistical flow chart of word frequency based on text mining.

the most frequently appeared is “positive,” appearing more than 300 times. It can be seen that the evaluation of ideological and political college students in the society attaches more importance to whether college students have the ideological and moral qualities that can fulfill the mission of completing the great rejuvenation of the country. This is also one of the keys to the effectiveness of ideological and political education in colleges and universities. Secondly, the most common ones are the “sense of gain” and the students’ recognition of ideological and political education in practice and the recognition of ideological and political courses. It shows that the students’ sense of gain for ideological and political courses directly affects

the college students’ acceptance of ideological and political courses and thus forms their value system, as shown in Figure 8.

As can be seen from Figure 8, the words “mission, feelings, pressure,” and so on occupy a relatively large area, so the number of occurrence in the social ideological and political evaluation of universities is relatively high. It can be seen that the identity of college students who attach more importance to the whole country is the key. First of all, we must have the overall situation before we can see their own times.

**4.2. Sentiment Analysis.** Classified according to text processing, it can be divided into emotion analysis based on product



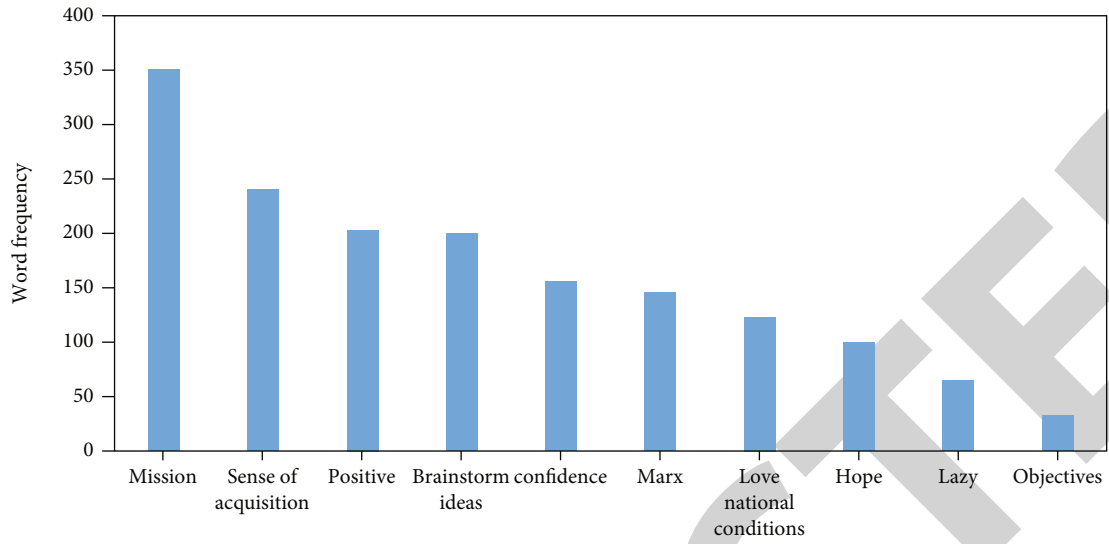


FIGURE 7: Keyword frequency distribution diagram.

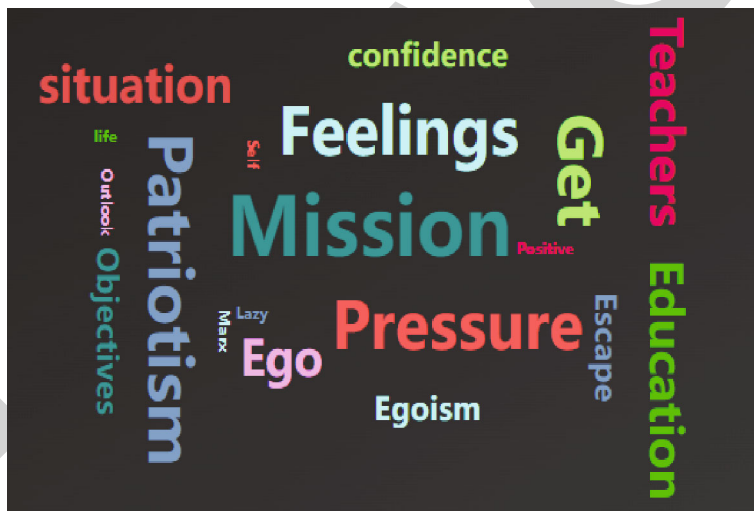


FIGURE 8: Keyword word cloud map.

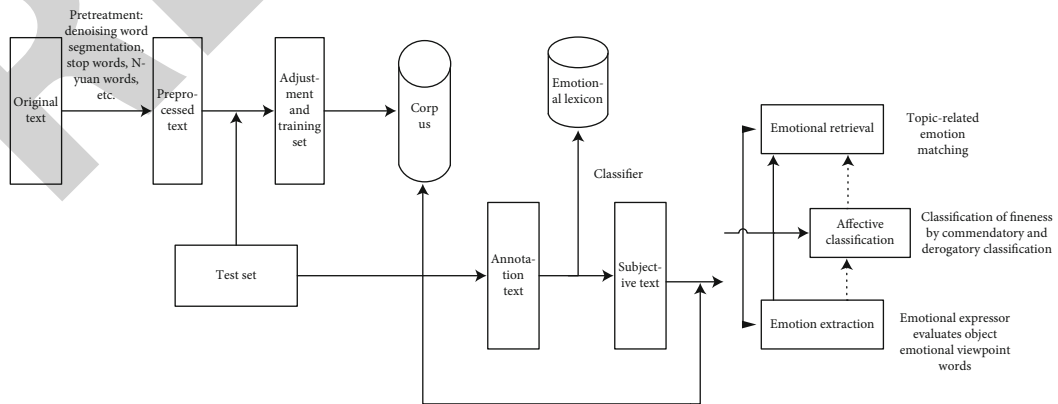


FIGURE 9: Basic flow chart of text emotion analysis

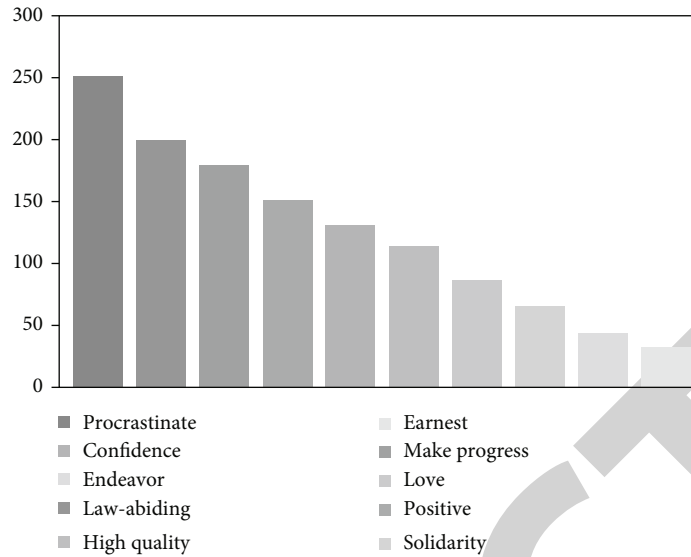


FIGURE 10: Distribution diagram of the positive evaluation word frequency.

evaluation and emotion analysis based on news evaluation. According to the classification of the task, [20] can also include text emotional analysis, emotional retrieval, and emotional acquisition. The basic process of text emotion analysis is shown in Figure 9, including the whole process of crawling of the original text.

#### (a) Dictionary construction

The establishment of emotion dictionary is the basic premise and cornerstone of modern emotion classification. At present, it can be divided into four kinds in real use: general emotional noun, degree adverb, negative word, and domain noun. At present, most of the formation of emotional dictionaries in China uses the expansion of existing electronic dictionaries to make emotional dictionaries. On the other hand, English [21], on the basis of artificially establishing seed adjectives, is used to determine the emotional tendency of words, thus evaluating the true views of opinions. Zhu et al. [22] used the repetition of language meaning to obtain the meaning similarity between the word and the basic emotional related word set and deduced the emotional expression of the word.

#### (b) Construction of the tendency calculation algorithm

Based on the meaning of emotional dictionary feature calculation, it is different from the required practice machine learning method, mainly using the collection of emotional words and grammar library analysis of the special structure and emotional tendency words, using weighted calculation method instead of the traditional human discrimination or only using simple statistical method of emotional classification. Emotional words with different emotional intensity are given different weights and then accumulated according to different weights. Document [23] uses a weighted averag-

ing method (15) calculation, which can help improve the efficiency and accuracy of emotion classification in special fields, such as

$$E = \frac{\sum_{i=1}^{N_p} wp_i + \sum_{j=1}^{N_n} wp_j}{N_p + N_n}. \quad (15)$$

Among these,  $p$  and  $n$  represent the number of words for both positive and negative emotions. The  $wp_i$  is the weight of the number of words indicating the positive emotions, and the  $wp_j$  represents the weight of the number of negative emotion words, mainly based on the different emotional intensity given to the words.

#### (c) Determine the threshold to determine the text orientation

If correct, the weighted calculation is negative, while the zero result has no tendency to be positive or negative. The result evaluation uses the value of accuracy  $F$  to summarize the final results in natural language. As for the combination of emotion words and compared with the traditional computer learning classification mode, although it is full of biased classification mode, it is easier to run with the emotion word set after good practice, and the text in the usual field can quickly classify emotion words.

According to the emotional emotion tendency in the emotion analysis model, the ideological and political evaluation of college students is divided into positive evaluation and negative evaluation, and the word frequency statistical map and word cloud map of positive evaluation and negative evaluation are drawn. See results (Figures 10 and 11).

As can be seen from the positive word frequency distribution map, for the ideological and political evaluation, "patriotism, law-abiding" appeared the most times, with

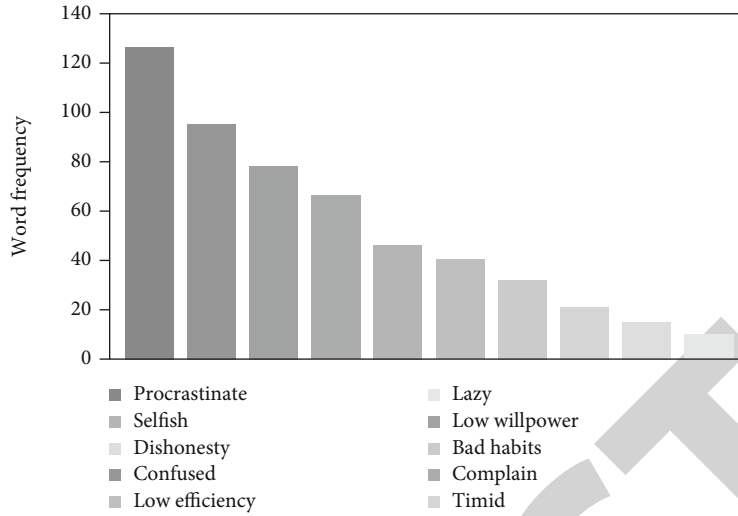


FIGURE 11: Distribution diagram of the negative evaluation word frequency.

TABLE 4: The performance comparison of the SVM classifier before and after the data preprocessing in the ideological and political evaluation topic of the data set.

Special topic of ideological and political evaluation	Traditional SVM model					The SVM model is based on the word frequency statistics rule				
	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Time (s)	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)	Time (s)
Acq	87.89	76.36	70.99	73.58	2279	88.63	78.03	72.53	75.18	611
Crude	94.28	50.61	47.98	49.25	2037	94.55	53.09	49.71	51.34	574
Earn	84.65	84.63	75.02	79.54	2285	86.29	86.99	77.04	81.71	637
Grain	92.94	70.49	61.70	65.80	2196	93.38	72.98	63.22	67.75	605
Interest	94.55	42.75	41.26	41.99	1986	94.28	45.59	43.36	44.45	632
Money	93.92	58.12	54.88	56.45	1959	94.15	59.80	56.74	58.23	553
Ship	96.59	39.74	36.05	37.81	1891	96.72	42.68	40.70	41.67	548
Trade	94.68	45.04	40.41	42.60	2025	94.92	47.73	43.15	45.32	571
Macro average	92.44	58.47	53.66	55.96	2082	92.93	60.86	55.81	58.23	591

more than 200. It can be seen that general people believe that the basic moral quality of college students is passed. It can be seen from the negative frequency distribution of negative words that college students have many bad habits in entering and leaving the society, which shows that the process of college students only accepts the ideological course, which is not implemented into the actual life, which is very unfavorable for personal development. Whether it is positive or negative evaluation is based on the evaluation of college students “knowledge, affection, meaning, action” dimension evaluation, from a comprehensive understanding of the ideological and political performance of college students from these four aspects.

4.3. *Text Mining Comment Analysis.* In order to extract more efficient information about the ideological and political evaluation of college students, two different methods will compare the accuracy, precision rate, recall, and F1 mea-

asures. A large number of articles were selected to verify the efficiency of the two methods.

(a) Data set

This article selects the special topics on ideological and political evaluation and the ideological and political report. Among them, there are 21,578 documents in 135 categories. Documents in eight categories were Acq (1659), crude (405), earn (2775), grain (773, corn and wheat under grain), interest (335), money (502), ship (200), and trade (340). The ideological and political report is divided into four categories: Comp (1162), Rec (1190), Sci (1183), and Talk (975); the experiment is binary in one-to-one categories. The ratio of both the training set to the test set was 7 : 3, as shown in Tables 4 and 5.

In order to compare more clearly, the accuracy, accuracy, recall, and F1 metric of the two models of ideological and political evaluation thematic data set were drawn, respectively, as shown in Figures 12–15.

TABLE 5: The sexual comparison of the classifier before and after the data preprocessing based on the word frequency statistics law on the data ideological and political report.

Ideological and political report	Traditional SVM model					The SVM model is based on the word frequency statistics rule				
	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Time (s)	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)	Time (s)
Comp vs. Rec	87.50	89.12	85.44	87.24	3647	88.78	90.88	88.89	88.89	953
Comp vs. Sci	75.31	76.09	73.15	74.59	3418	77.87	78.28	77.43	77.43	892
Comp vs. Talk	96.02	96.79	95.899	96.33	4065	96.49	96.73	96.77	96.77	1067
Rec vs Sci	73.58	76.09	69.99	72.91	3734	76.99	74.12	75.88	75.88	981
Rec vs. Talk	83.27	86.25	82.78	84.48	3849	86.00	85.04	86.78	86.78	1078
Sci vs. Talk	76.14	80.04	75.23	77.56	4276	79.19	77.94	80.42	80.42	1113
Macro average	81.97	84.06	80.41	82.19	3830	84.22	82.90	84.48	84.48	1014

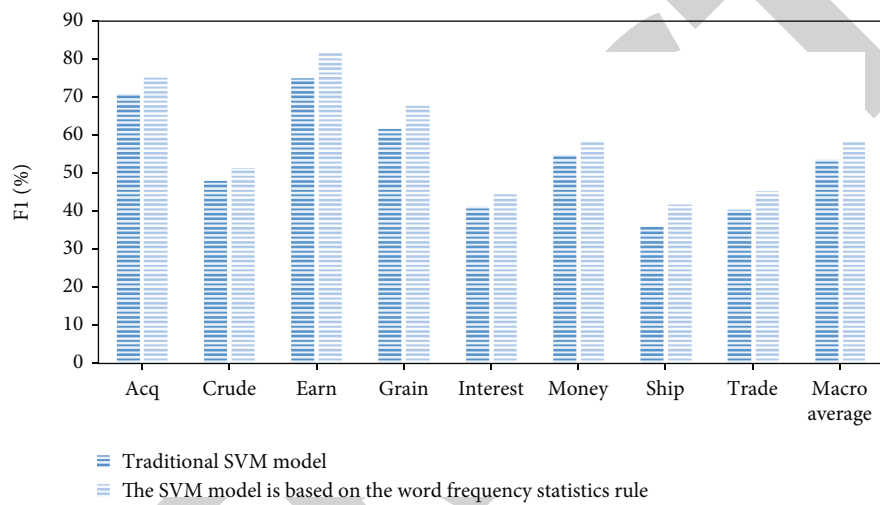


FIGURE 12: Comparison diagram of F1 measures of the traditional SVM and SVM based on word frequency statistics in the ideological and political evaluation topic.

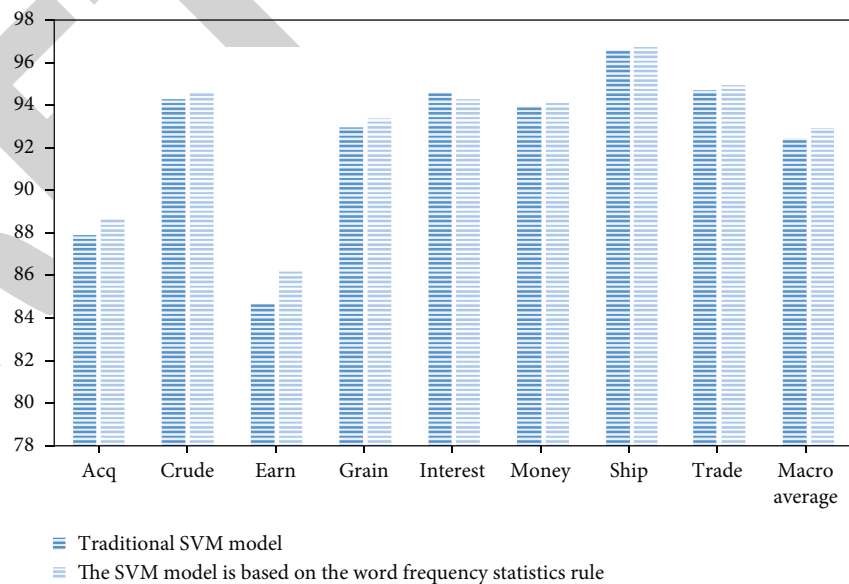


FIGURE 13: Comparison of the traditional SVM and SVM based on word frequency statistics in the ideological and political evaluation topic.

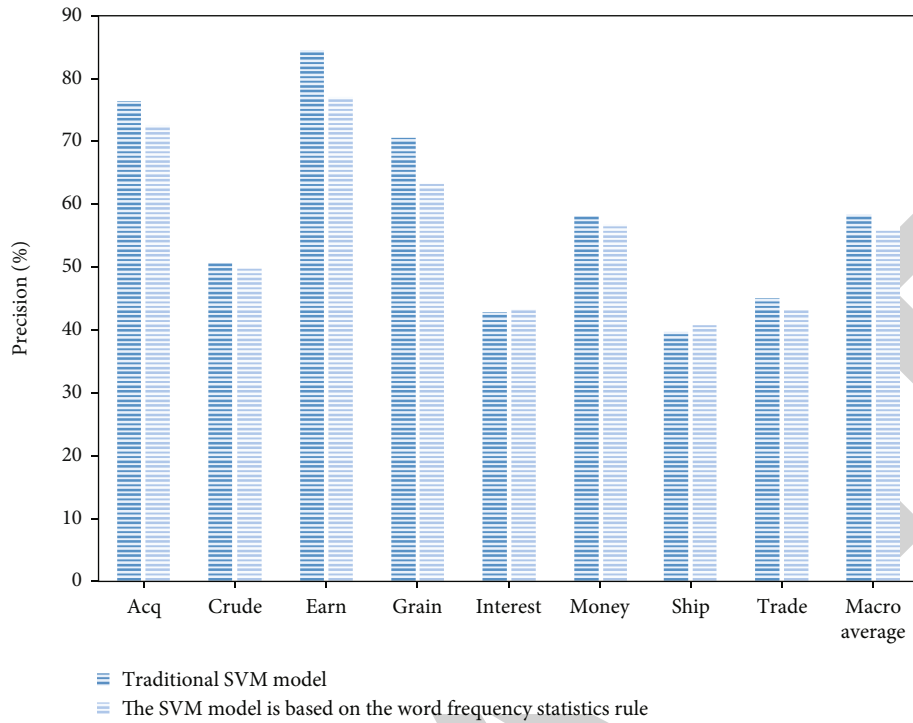


FIGURE 14: Comparison figure of the traditional SVM and SVM based on word frequency statistics in the ideological and political evaluation topic.

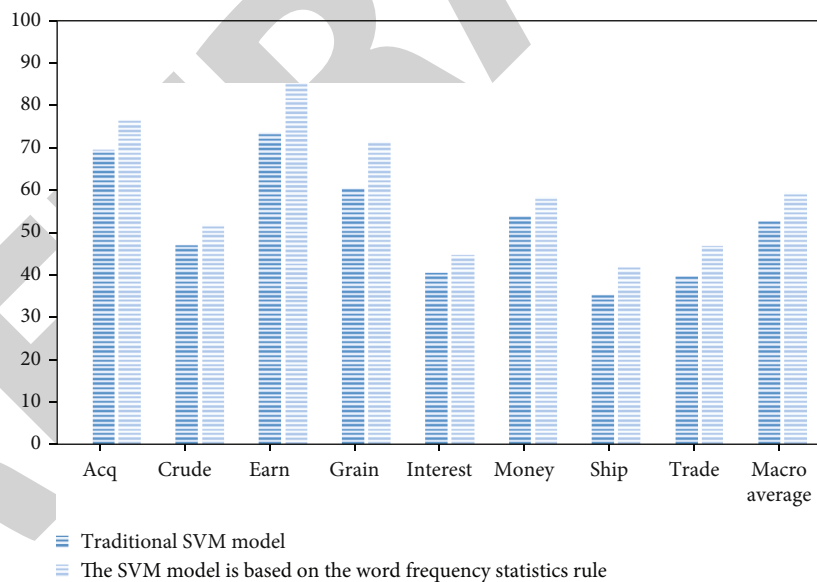


FIGURE 15: Comparison of the traditional SVM and SVM based on word frequency statistics in the ideological and political evaluation topic.

Through experimental comparison, the traditional preprocessing method cannot solve the problems of the data, so this paper finally adopts an improved text preprocessing method based on word frequency rule statistics and then overcome this problem. It can be seen from the figure that the text data preprocessing method based on the text mining law is higher than the traditional performance, and the running time is shorter. Therefore, the preprocess-

ing method based on text mining will be adopted in the ideological and political evaluation of college students based on word frequency statistics. And because the SVM model based on word frequency statistics denoises low-frequency words during preprocessing, it can greatly reduce the feature dimension under the premise of ensuring the accuracy of text mining, significantly reduce the complexity of space-time, reduce the average running time

by more than 70%, and effectively improve the performance of text mining.

## 5. Conclusion

According to the results of the online evaluation and analysis on the ideological and political evaluation of college students on Zhihu, the society has both positive and negative ideological and political evaluation of college students, and most of them are rather positive. The positive view gives affirmation to the moral quality of college students, and contemporary college students have a good performance in terms of public moral quality. It shows that students have a high level of their own knowledge-based cognitive value cognition. The overall ideological and political evaluation of college students is divided into four dimensions, namely, knowledge, emotion, meaning, and practice. College students finally face the society, so they should improve the effectiveness of the quality of ideological and political education from the four dimensions of knowledge, affection, meaning, and action. Ideological and political education can not be simply pre-processed, improve the quality of ideological and political education of college students in the new era [24], optimize the ideological and political education team, and accelerate the construction of the education system. In terms of knowledge dimension, it is believed that college students have formed their own unique values in the process of school education, and most people think that it is positively in line with the characteristics of the times, and the way of thinking to deal with problems is still more dogmatic. In the evaluation of the emotional dimension, the contemporary college students are more qualified in the emotional identity, mind emotion, and self-emotion management and adjustment. It is generally believed that 70 percent of college students have a strong sense of family and social responsibility, and a small number of college students have a weak awareness in this respect. Then, there is the evaluation of the meaning dimension. The results show that contemporary college students have a strong legal concept and abide by the discipline. And daily life can make good use of the legal thinking ability to safeguard the rights and interests that are not infringed. Finally, in the above evaluation, college students lack self-regulation and self-development. Although the general moral quality of college students is good, in the face of personal cognition, development is relatively confused. The evaluation shows that 80% of college students' self-regulation ability is not very good, they have not made their own life planning, they are in a confused state after graduation, their behavior habits, behavior ability, and behavior are in the primary stage, their thinking is generally not very mature, and the law of life is also very irregular, which is closely related to the daily development of students in school.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## References

- [1] Onjani, "Evaluation "research on the quality evaluation of ideological and political education in colleges and universities in the new era"," *Journal of Tropical Crop*, vol. 42, no. 11, p. 3430, 2021.
- [2] H. Tan, "Research on the promotion path of ideological and political theory course in universities in the new era," *Educational Theory and Practice*, vol. 41, no. 33, pp. 36–39, 2021.
- [3] X. Lin, P. Wei, and L. Xu, "Research on the precision of the quality evaluation of ideological and political education in colleges and universities under the background of big data," *Journal of Higher Education*, vol. 29, pp. 170–173, 2020.
- [4] Z. Lei and J. Bing, "Research on the rational connotation and implementation path of college curriculum ideological and political construction," *Journal of Jilin Provincial Institute of Education*, vol. 37, no. 12, pp. 59–62, 2021.
- [5] I. Yu, "Qingfeng. Study on the realization path of ideological and political courses and courses," *Journal of Social Sciences, Jiamusi University*, vol. 39, no. 6, pp. 85–88, 2021.
- [6] X. Guanghui, "Research on ideological and political practice path of college wushu education course," in *2021 National Wushu Education and Health Conference and National Traditional Sports into Campus Seminar Summary Compilation (2)*, p. 299, 2021.
- [7] M. Xin and X. Guifen, "Review of text-mining techniques," *Journal of Jiangsu University (Natural Science Edition)*, vol. 5, pp. 72–76, 2003.
- [8] W. Yan, D. Bin, X. Deng, and J. He, "Electric text mining and classification based on natural language processing technology," *Automation Technology and Application*, vol. 40, no. 10, pp. 60–63, 2021.
- [9] Yin, "Experimental design and implementation of text mining," *Information Technology and informatization*, vol. 9, pp. 94–96, 2021.
- [10] F. Wang, Z. Rui, and G. Haili, "Design and implementation of a distributed crawler based on the Scrapy framework," *Information Technology*, vol. 3, pp. 96–101, 2019.
- [11] Yunxian, Z. Shuliang, L. Yan, G. Lin, Z. Junpeng, and L. Chao, "Text data preprocessing method based on the word frequency statistics law," *Computer Science*, vol. 44, no. 10, pp. 276–282, 2017.
- [12] F. Yanru, "Design and implementation of the network crawler system based on Python," *Computer and Information Technology*, vol. 29, no. 6, pp. 47–50, 2021.
- [13] C. Wang, Z. Hui, M. Xioliang, and W. Yang, "Review of Weibo sentiment analysis," *Computer Engineering and Science*, vol. 44, no. 1, pp. 165–175, 2022.
- [14] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans on Pattern Analysis & Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [15] C. Yusen, "Based on Python," *Information and Computer (Theory Edition)*, vol. 33, no. 21, pp. 41–44, 2021.
- [16] G. K. Zipf, G. Zipf, and G. Zipf, "Book reviews: human behavior and the principle of least effort: an introduction to human

- ecology,” *American Journal of Sociology*, vol. 110, no. 110, pp. 306–306, 1950.
- [17] A. Booth, “A law of occurrences for words of low frequency,” *Information and Control*, vol. 10, no. 4, pp. 386–393, 1967.
- [18] R. Yuen and G. Chao, “Human behavior and the principle of least effort: an introduction to human ecology,” *Language*, vol. 19499, no. 23, p. 26, 1949.
- [19] T. R. Jenkinson, P. A. Mallorie, H. C. Whitelock, L. G. Kennedy, S. L. Garrett, and A. Calin, “Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index,” *Journal of Rheumatology*, vol. 21, no. 9, pp. 1694–1698, 1994.
- [20] Z. Yanyan, Q. Bing, and L. Ting, “Text propensity analysis,” *Journal of Software Science*, vol. 21, no. 8, pp. 1834–1848, 2010.
- [21] M. Hu and B. Liu, *Mining and Summarizing Customer Reviews*, Proceedings of Knowledge Discovery and Data Mining, NY, USA, 2004.
- [22] Y. Zhu, M. Jin, and Z. Yaqian, “The how net-based lexical semantic tendency calculation,” *Chinese Informatics Journal*, vol. 20, no. 1, pp. 14–20, 2006.
- [23] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, p. 130, 2008.
- [24] Z. Jingshui, “Path to improve the ideological and political education quality of college students in the new era,” *Journal of Hubei Open Vocational College*, vol. 34, no. 24, pp. 72–74, 2021.