



Article

Suicide Possibility Scale Detection via Sina Weibo Analytics: Preliminary Results

Yun Gu ¹, Deyuan Chen ^{1,*} and Xiaoqian Liu ^{2,3,*} 

¹ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China

² Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

³ Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: chendy@ucas.ac.cn (D.C.); liuxiaoqian@psych.ac.cn (X.L.)

Abstract: Suicide, as an increasingly prominent social problem, has attracted widespread social attention in the mental health field. Traditional suicide clinical assessment and risk questionnaires lack timeliness and proactivity, and high-risk groups often conceal their intentions, which is not conducive to early suicide prevention. In this study, we used machine-learning algorithms to extract text features from Sina Weibo data and built a suicide risk-prediction model to predict four dimensions of the Suicide Possibility Scale—hopelessness, suicidal ideation, negative self-evaluation, and hostility—all with model validity of 0.34 or higher. Through this method, we can detect the symptoms of suicidal ideation in a more detailed way and improve the proactiveness and accuracy of suicide risk prevention and control.

Keywords: suicidal ideation; machine learning; suicide possibility scale



Citation: Gu, Y.; Chen, D.; Liu, X. Suicide Possibility Scale Detection via Sina Weibo Analytics: Preliminary Result. *Int. J. Environ. Res. Public Health* **2023**, *20*, 466. <https://doi.org/10.3390/ijerph20010466>

Academic Editor: Paul B. Tchounwou

Received: 27 October 2022

Revised: 13 December 2022

Accepted: 20 December 2022

Published: 27 December 2022



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

1. Introduction

Suicide is a conscious act of voluntarily and actively ending one's life [1]. Despite today's level of medical health, according to WHO statistics, nearly 1 million people commit suicide each year worldwide. Approximately 287,000 people commit suicide each year in China, accounting for one third of global suicide deaths and the fifth most-common cause of death, making it an important public health and mental health problem in China and globally [2]. However, the identification of suicidal individuals is still an undoubtedly complex and difficult task [3]; therefore, it is important to find an efficient method for suicide detection.

Current technologies for suicide detection span manifold domains and methods. Traditional approaches use clinical methods with patient–clinic interaction [4] to have subjects fill out specific suicide questionnaires and scales and to have experts analyze the contents of the completed assessments to draw the appropriate conclusions. In other words, clinical treatment seeks an understanding of the psychology behind suicidal behavior [5]. Of course, this approach is extremely dependent on the clinician's own expertise and the face-to-face interaction at the time. The current study uses a judgmental analysis of suicide based on the Suicide Probability Scale (SPS) [6], which is an established and reliable questionnaire for assessing the likelihood of suicide that uses subscale scores and total scores to determine the patient's likelihood of suicide. However, some at-risk people conceal their true situations [3], and practice effects appear after multiple participation in a given assessment [7].

In 2012, a 22-year-old girl named Zou Fan left her last words on Weibo that she was suffering from depression and decided to die. She hanged herself in her dormitory. The police and medical personnel tried their best to save her life but were unable to do anything. In response to the news, netizens were generally saddened and sorry, and more and more people are calling for attention to depressed people to prevent them

from committing suicide. For psychological disorders leading to suicide, the earlier the risk group is detected, the earlier treatment can be provided to avoid the aggravation of suicidal tendencies; however, questionnaires lack timeliness. Moreover, people with suicidal ideation do not tend to fill out questionnaires voluntarily, and questionnaires are not active interventions. Thus, we need more timely and proactive methods to detect suicide. In addition to the traditional method, texts actively generated by risk groups enable better detection [8]. With the advancement of Internet technology, more and more people share their lives, experiences, views, and emotions in those online venues [9], which provides data sources and application platforms for us to monitor suicide.

An increasing number of studies have tried to detect suicide from social media texts [10] and have been able to quickly identify users with suicidal ideation and intervene in a timely manner. Users' posts on social networking sites reveal a wealth of information and their language preferences. Through exploratory data analysis of user-generated content, we can gain insights into language usage and linguistic cues from suicide attempters. Suicide-related keyword dictionaries and lexicons have been manually built to enable keyword filtering [11] and phrase filtering [12]. Gunn and Lester [5] analyzed the Twitter posts of suicide attempters in the 24 h prior to their deaths. Coppersmith et al. [13] analyzed the language use in data from the same platform. Suicidal thoughts may involve strong negative emotions, anxiety, hopelessness, or other social factors. These thought reactions can be expressed in the text, as some words and phrases include "kill", "suicide", "feel alone", "depressed", and "cutting myself". A lexicon based on words such as these can be used as a feature input to identify the possibility of suicide.

The identification of suicidal ideation is mostly based on binary classification models, and suitable feature selection and machine-learning methods have been used to construct classification models. One of the keys is to select the appropriate features. Table 1 shows some of the recent studies on suicide detection based on social media. From the selection of dictionaries, we can find that they [14,15] are based on LIWC dictionaries alone for feature extraction or the fusion of features extracted from Chinese suicide dictionaries [16] and LIWC or construct a new dictionary according to the data situation [17,18] and use it for feature extraction. Most of the modeling methods use machine-learning SVM methods to build discriminative models, but there are also some studies that choose to train neural networks to build models, such as Shing [19] using CNN methods to build models to achieve the desired results. Q. Cong [8] using a variety of modeling methods—including training LSTM networks—to achieve accuracy above 0.9 (although random forest achieved better results in that study).

Table 1. A review of different suicide ideation detection studies on social media.

No	Features Extracted	Modeling Methods	Social Media	References
1	Custom Vocabulary Dictionary	Naive Bayes	Twitter	[17]
2	TFIDF	SVMs	Twitter	[10]
3	Custom Emotion-labeled	Logistic Regression	Twitter	[18]
4	SC-LIWC	SVM	Weibo	[14]
5	Data-driven Dictionary, LIWC, Chinese Suicide Dictionary	SVM, DT	Weibo	[16]
6	TFIDF, LIWC, Sentiment Analysis	Logistic Regression, Random Forest, SVR	Reddit	[15]
7	TF-IDF, LIWC, POS	Random Forest LSTM etc.	Reddit and Twitter	[8]
8	BoWs, empath, readability, LIWC, NRC, LDA etc.	CNN	Reddit	[19]

In summary, the existing studies have been able to detect suicidal ideation to some extent, but there are still some shortcomings and areas for improvement. The traditional questionnaire method and scales can help us identify people at risk with some reliability, but it does not allow timely detection or intervention and requires active completion by the patient, making it highly passive and unable to automatically detect people at risk. In contrast, the automatic detection method can identify suicidal tendencies from text, which allows for a timely and automatic intervention. Therefore, we combine the two methods and use the automatic text detection method to obtain the SPS results. First, for the purpose

of the study, most of the models for suicide detection are still based on classification models, and the results are a binary yes or no. However, for suicide early warning and risk judgment, in most cases, absolute judgments cannot be made, and a risk prediction and assessment of suicidal tendency is required to enable targeted treatment plans. Therefore, the results of this model are no longer categorized but modeled and predicted based on Weibo data for each of the four subscale scores of the Suicide Possibility Scale, which is mature and reliable. Experts can diagnose individuals based on the predicted results to obtain a more reliable suicide likelihood analysis. Second, with respect to the research methodology, unlike other studies that combine features for a single or a few lexical dimensions, our study selects six lexical dimensions to obtain more valuable and comprehensive linguistic features and then selects the optimal set of features to build the model. All aspects of the above appear to enable proactive and automated detection of SPS results and provide a new approach to suicide detection.

2. Materials and Methods

This study aimed to investigate the correlation between the linguistic features of textual social media content and suicide likelihood ideation and provide a machine-learning approach combining feature screening and linear regression to detect the four indicators of the SPS [6]. The working principle of the realization method of this study is as follows.

First, the subjects' Weibo text data were collected over a period of time. Then, several dictionaries were used as the basis for language feature extraction, including the Weibo Basic Mood Lexicon [20], the individualism/collectivism lexicon [21], The Chinese Version of the Moral Foundations Dictionary [22], the Moral Motivation Dictionary [23], the Chinese suicide dictionary [24], and Language Inquiry and Word Count (LIWC) [25]. The relationship between the multidimensional features and the model is judged using the Akaike information criterion (AIC) to filter out the best features for the model. The features are then used as model inputs, and a multiple linear regression algorithm is used to construct a suicide likelihood prediction model. Figure 1 illustrates the above modeling process.

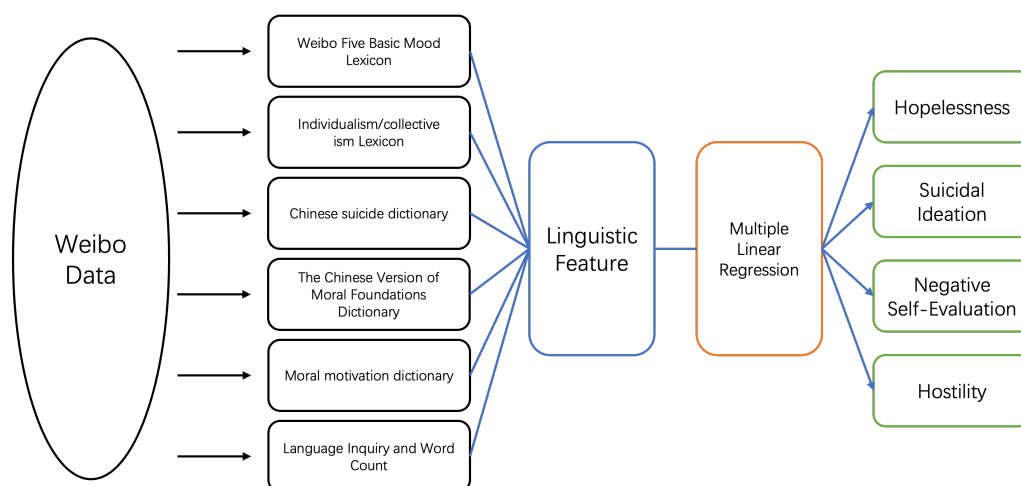


Figure 1. The Model Construction Method.

2.1. Data Collection

The training data for this study were obtained from Sina Weibo, the largest online social media platform in China. We recruited 1062 subjects on our platform and asked them to fill out the Chinese version of the SPS [6], an international questionnaire used to assess suicide attempts in adolescents and adults. The original English version of the questionnaire [26], developed by John G. Cull, Ph.D., and Wayne S. Gill, Ph.D., has good reliability and validity and consists of 36 questions with four dimensions: hopelessness, suicidal ideation, negative self-evaluation, and hostility. The hopelessness dimension consists of 12 questions, the suicidal ideation dimension consists of eight questions, the

negative self-evaluation dimension consists of eight questions, and the hostility dimension consists of seven questions. Each question was scored on a 1~4 scale, which can be simply interpreted as the higher the score, the greater the likelihood of suicide. With the subjects' authorization, a crawler was used to download all their original Weibo data for the 30 days prior to the date the scale was completed. The users were also filtered according to the number of posts in 30 days, and only active users with 10 or more posts were retained. A total of 481 valid samples were obtained, with the average number of posts being 62. The demographic information of the valid samples can be found in Table 2, and the subjects' scores on the SPS scale can be seen in Table 3.

Table 2. Demographics of participants.

Demographics	Subclass	All Weibo Posts (N = 37,474), n (%)
Gender	Male	9775 (26.08)
	Female	27,699 (73.92)
Regions	country	192 (0.51)
	Town	2540 (6.78)
	Prefecture-level city	12,991 (34.67)
	municipality	21,751 (58.04)
Ages	14~20	5744 (15.33)
	20~30	27,624 (73.72)
	30~52	4106 (10.96)
Profession	Laborer	371 (0.99)
	Civil Servant	883 (2.36)
	Military/Police	32 (0.09)
	Researcher/Teacher	4749 (12.67)
	Media Practitioners	1300 (3.47)
	Unemployed	1153 (3.08)
	Student	16,249 (43.46)
	Doctor/Nurse	1111 (2.96)
	Staff	6895 (18.40)
	Freelancers	1473 (3.93)
	Self-Employed	569 (1.52)
	Other	2689 (7.18)

Table 3. Suicide possibility scale scores.

Dimensions	Average	Standard Deviation	Highest	Lowest
Hopelessness	24.53	4.71	39	12
Suicidal Ideation	11.53	3.27	26	8
Negative Self-Evaluation	20.36	4.31	36	9
Hostility	12.91	2.35	21	7
Total Score	69.32	11.82	106	43

2.2. Feature Extraction

2.2.1. Dictionary

In this paper, we used six dictionaries for linguistic feature extraction. Detailed descriptions of the six dictionaries can be found in Table A1, and the following is a brief description of the dictionaries and how the corresponding dictionaries relate to suicide, including

- (1) The Weibo Five Basic Mood Lexicon [20]. It includes an 818-word microblogging basic emotion lexicon constructed by analyzing more than 1.6 million users to measure five basic social emotions (happiness, sadness, disgust, fear, and anger) in microblogs, and verify their validity. Psychological studies have shown that suicide results from a cumulative cause [27] and that the accumulation of negative emotions and repeated outbursts are causative factors of suicide and mediate suicide risk [28].

- (2) The Individualism/Collectivism Lexicon [21] provides a dictionary of collectivism and individualism based on Weibo data. The data for this dictionary also come from Weibo and shows that individualism has a significant positive relationship with suicide rate [29].
- (3) Researchers built the Chinese suicide dictionary [24] from 4653 posts on Sina Weibo and two Chinese sentiment dictionaries, and it has good performance in identifying suicide risk on Weibo.
Some studies have shown that group-oriented (binding) moral intuitions are associated with lower suicide risk, while individual-oriented (individualizing) moral intuitions are associated with higher suicide risk [30]. Therefore, We have also selected two dictionaries related to morality.
- (4) The Chinese Version of the Moral Foundations Dictionary [22] is based on moral foundation theories and the Moral Foundations Dictionary of Graham and colleagues (2009) in the Chinese context.
- (5) The Moral Motivation Dictionary [23] was initially designed by Frimer (2013) for LIWC. It contains 349 words in the agency dimension (e.g., accomplish, defeat, spent) and 146 words in the communion dimension (e.g., accepting, care, kindness).
- (6) The SCLWC, the Simplified Chinese Microblog Word Count tool [31], is a Chinese version of the classic and commonly used LIWC [25] tool for sentiment analysis, allowing a one-stop solution from automatic Chinese word segmentation to psycholinguistic analysis.

2.2.2. The Processing

We will use LIWC as an example to introduce how to extract features from the original text based on the corresponding dictionary. The LIWC program has two central features—the processing component and the dictionaries. The processing feature is the program itself, which opens a series of text files—which can be essays, poems, blogs, novels, and so on—and then goes through each file word by word. Each word in a given text file is compared with the dictionary file [25].

For example, if LIWC were analyzing this sentence:

It was a terrible thing.

The program would first look at the word “it” and then see if “it” was in the dictionary. It is and is coded as a function word, a pronoun, and, more specifically, an impersonal pronoun. All these LIWC categories would then be incremented. Next, the word “was” would be checked and would be found to be associated with the categories of verbs, auxiliary verbs, and past tense verbs. After going through all the words in the text, LIWC would calculate the percentage of each LIWC category. Therefore, for example, we might discover that 2.34% of all the words in a given book were impersonal pronouns and 3.33% were auxiliary verbs. The LIWC output, then, lists all LIWC categories and the rates that each category was used in the given text.

The multidimensional linguistic features were extracted from the post contents of 481 microblog users using the above six dictionaries. They were manually filtered, the full 0-dimensional and irrelevant features were screened out, and a total of 121-dimensional linguistic features were obtained.

2.3. Model Construction with Machine-Learning Algorithms

By comparing various algorithms, we chose the multiple linear regression algorithm as the machine-learning algorithm for the model. For the multiple linear regression algorithm, we used a stepwise regression approach to select a more appropriate set of independent variables.

The basic idea of stepwise regression is to introduce variables into the model one by one, perform an F test after each explanatory variable is introduced, and perform a *t* test on each of the explanatory variables that have been selected. If the explanatory variable introduced first is no longer significant due to the introduction of subsequent explanatory

variables, it is removed. This is undertaken to ensure that only significant variables are included in the regression equation before each new variable is introduced. This is an iterative testing process until neither significant explanatory variables are selected into the regression equation nor insignificant explanatory variables are removed from the regression equation, indicating that the optimal set of explanatory variables is reached at that point. Here, we choose the *AIC* as the criterion for independent variable selection, and when the *AIC* reaches a minimum, we obtain the most available set of independent variables. The basic principles of the *AIC* are as follows:

$$AIC = -2\ln L(\hat{\theta}_L, x) + 2p \quad (1)$$

where $\hat{\theta}_L$ is the maximum likelihood estimate of θ and p is the number of unknown parameters. Stepwise regression is performed for each of the four predicted SPS indicators to obtain the corresponding optimal set of independent variables.

Multiple linear regression analysis is a regression analysis conditional on the given values of multiple explanatory variables and is a method to study the linear relationship between a dependent variable and multiple independent variables. In fact, a phenomenon is often associated with more than one factor, and the optimal combination of multiple independent variables to predict or estimate the dependent variable is more effective and better suited to the needs of our experiments than using only one independent variable for prediction or estimation. Moreover, using regression models, the results that can be calculated by standard statistical methods are unique as long as the models and data used are the same.

The general form of a multiple linear regression model is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_j x_j + \cdots + \beta_k x_k + \mu \quad (2)$$

where k is the number of explanatory variables, $\beta_j (j = 1, 2, \dots, k)$ is the regression coefficient, and μ is the random error after removing the effect of k independent variables on Y .

2.4. Measures of Model Performance

The performances of the constructed suicide probability detection models were evaluated using reliability and validity. Reliability assesses the consistency of a measure, and validity assesses its accuracy.

We used five-fold cross-validation to calculate the Pearson correlation coefficient of the model to verify the validity of the model. The Pearson correlation coefficient is the test statistic that measures the statistical relationship or association between two continuous variables. It is known to be the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives the correlation between linguistic features of textual social media content and suicide likelihood ideation, and it can also demonstrate the reliability of the model. The basic principles of the Pearson correlation coefficient are as follows:

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} \quad (3)$$

where $Cov(x, y)$ denotes the covariance of the sample and σ_x and σ_y denote the standard deviation of the sample.

We calculate the odd-even split reliability of the model. We divide the content of each user's Weibo posts in half according to odd and even, extract features from the odd and even parts, and then build a model according to the same modeling method to predict the four indicators of SPS [6]. Finally, we compare the correlation between the two sets of prediction data obtained to verify the reliability of the model.

3. Results

3.1. Feature Extraction Results

By means of stepwise regression, for each subscale dimension, we obtain the corresponding optimal set of features. For different dictionaries of linguistic features, the features selected for each dimension vary and are roughly distributed as shown in Table 4:

Table 4. Feature Distribution.

Dictionary	Hopelessness	Suicidal Ideation	Negative Self-Evaluation	Hostility
SCLIWC [31]	21	12	18	21
Moral Foundations Dictionary [22]	4	3	4	6
Chinese suicide dictionary [24]	4	3	4	3
Weibo Five Basic Mood Lexicon [20]	1	1	1	1
Individualism/Collectivism Lexicon [21]	1	0	0	1
Moral Motivation Dictionary [23]	1	0	0	0
Total	32	27	19	32

3.2. Split-Half Reliability

After stepwise regression, the extracted multidimensional language model features were filtered to obtain the optimal feature set. The model was constructed by the machine-learning method of multiple regression. A portion of the data is randomly selected to train the new model and split the remaining data in half according to odd and even numbers as a test set, obtain the corresponding predicted values, and compare the correlation between the two to obtain the corresponding model reliability. Table 5 shows the reliability of the model's predictions for the four indicators of the SPS scale and the total scores.

Table 5. Reliability and validity test results of suicide possibility identification model based on microblog text analysis.

Dimensions	Validity	Reliability
Hopelessness	0.34	0.72
Suicidal Ideation	0.35	0.38
Negative Self-Evaluation	0.35	0.47
Hostility	0.36	0.81
Total Score	0.35	0.65

3.3. Criterion Validity

The Pearson correlation coefficient between the predicted and actual scores of each subscale was calculated using a five-fold cross-validation method, which led to the criterion validity analysis (as shown in Table 5). The results showed that the correlation coefficients reached a significant level, which implies that the developed model has high criterion validity.

4. Discussion

The reliability and validity results show that the machine-learning method based on multiple linear regression better predicts the Suicide Possibility Scale and provides a new possibility for the prediction of suicide possibility through non-subjective social media. It is worth noting that the predictive validities of all four indicators of the scale were above 0.34, and the validity of the model for the total score of the scale reached 0.35, indicating that it has great significance in predicting the possibility of suicide through the Suicide Possibility Scale. By looking at the feature extraction results, we find that among the six dictionaries used, SCLIWC, the Moral Foundations Dictionary, and the Chinese suicide dictionary have the most extracted features, indicating that they provide the more dominant and important feature dimensions for accurate prediction of SPS. The reliability results show that suicidal ideation has the lowest reliability results.

The experimental results presented in Table 4 shows that among the set of features extracted based on six dictionaries, three dictionaries retained more features after the stepwise regression of feature filtering, among which SCLIWIC contributed the largest share of features. In the case of research on suicide, related texts typically entail the use of LIWC [32], which is a tool for the statistical analysis of corpora using a wide set of dictionaries. Using this tool has become standard in psychological studies on language [33], particularly studies on the language of suicide victims [13,34,35]. The information it provides is also often used in machine-learning algorithms [19]. LIWC provides a wide range of linguistic category annotations on the text. Michal Ptaszynski [32] found that the analysis of the obtained LIWC study results enabled several valuable insights into the vocabulary used by suicidal users in comparison to that used by non-suicidal users. Therefore, using LIWC categories as additional features helps the model acquire more important features.

The Moral Foundations Dictionary is another source dictionary with a relatively large number of features [22]. This dictionary is based on the LIWC, which is extended and compressed according to the five dimensions of the appeal. The authors developed several ways to measure people's use of five sets of moral intuitions: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity. We can find the category harm present in the set of features of each subscale, which contain a large number of negative words related to harm, while some studies have shown that suicide victims seem to show more self-concern and more negative expressions; use more cognitively exclusive, death-related, and religion-related words; and use fewer work-related words [37]. Therefore, moral-based dictionaries can extract different linguistic features of suicidal and non-suicidal people from Weibo and help in the construction of suicide likelihood models. From Table 4, we can see the Chinese suicide dictionary also contributes important features [24]; it selects initial words from 4653 posts published on Sina Weibo and two Chinese sentiment dictionaries (HowNet and NTUSD), and its performance in identifying suicide risk on Weibo has been confirmed. Its dictionary catalog contains suicide ideation, hopeless, self-regulation, and hostility, which correspond to the four subscale dimensions of the SPS, thus helping us extract the corresponding features more precisely.

By observing the reliability and validity test results of the suicide possibility identification model shown in Table 5, we find that the reliability of suicidal ideation is the lowest. This is due to suicidal ideation having been conceptualized and measured as a state rather than a trait, which would be expected to fluctuate in intensity over time [38]. For a fluctuating state quantity, it is difficult to make the results of each prediction reach a stable state, so there is low reliability.

5. Conclusions

As suicide is an increasingly prominent social issue, the traditional questionnaire method, SPS [6], does not allow for automatic detection or timely intervention, and at-risk individuals often conceal their intentions, which is not conducive to early suicide prevention. There is an urgent need for suicide risk markers that do not rely on self-reports, and this study provides a risk warning of suicide likelihood through social media comments on Weibo. Unlike conventional research methods that use machine learning to directly model suicide classification or traditional scale questionnaire analysis, this study combines machine learning and scale questionnaires, and instead of simply constructing a classification model, four-dimensional subscales of the SPS [6] scale for suicide likelihood discrimination are used for predictive analysis, enabling the results of SPS [6] to be automatically obtained and more timely interventions for risk groups. This study also provides a new way of thinking about existing research methods.

This study has some limitations. First, the amount of data in this experiment was not sufficient, and the time span was short. These factors may affect the reliability to some extent, and by observing the demographic informatics background shown in Table 2, we found that most of the subjects were 20–30 years old, which may have led to an age

sampling bias. In future studies, subjects with more balanced demographic backgrounds should be enrolled, and expanding the time span of the data can allow the model to cover a wider range of subjects. Second, in addition to the original text data of users, social media also contains a large amount of interactive communication information among users, which can also enrich the dimension of features and help the model to better analyze the psychological condition of users, but we were not able to obtain more information this time. In the future, we will consider introducing more aspects of data to help build the model.

Author Contributions: Conceptualization, Y.G., D.C. and X.L.; methodology, Y.G.; software, Y.G.; validation, Y.G.; formal analysis, Y.G.; data curation, Y.G.; writing—original draft preparation, Y.G.; writing—review and editing, Y.G., D.C. and X.L.; supervision, D.C. and X.L.; project administration, D.C. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the Fundamental Research Funds for the Central Universities: E2E41902X2, the Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences, No. E2CX4735YZ.

Institutional Review Board Statement: This research project was approved by Ethics Committee, Institute of Psychology, Chinese Academy of Sciences (project number: H15009).

Informed Consent Statement: Sina Weibo is freely accessible. Informed consent was obtained from all participants involved in the study. The Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences, with the ethics approval number H16003.

Data Availability Statement: To protect the participants' privacy, the original posts used for the analysis are not publicly available but from the corresponding author at a reasonable request.

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

Appendix A

Table A1. Introduction to the dictionary

No	Dictionary	Number of Words	Number of Categories	Categories
1	The Weibo Five Basic Mood Lexicon [20]	818	5	Happiness, sadness, anger, fear, disgust
2	The Individualism/Collectivism Lexicon [21]	117	3	Individualism, collectivism
3	The Chinese Suicide Dictionary [24]	1862	13	Suicide ideation, suicide behavior, psychology of mental illness etc.
4	The Chinese Version of the Moral Foundations Dictionary [22]	295	11	Harm, Fairness, Ingroup, Authority, Purity, Morality General
5	The Moral Motivation Dictionary [23]	950	2	Agency words, communion words
6	The SCLWC, the Simplified Chinese Microblog Word Count tool [31]	7444	71	Emotion, cognition, social contact and personal concerns

References

1. Homan, S.; Gabi, M.; Klee, N.; Bachmann, S.; Moser, A.; Duri, M.; Michel, S.; Bertram, A.; Maatz, A.; Seiler, G.; et al. Linguistic features of suicidal thoughts and behaviors: A systematic review. *Clin. Psychol. Rev.* **2022**, *95*, 102161. [\[CrossRef\]](#)
2. World Health Organization. *Preventing Suicide: A Global Imperative*; World Health Organization: Geneva, Switzerland, 2014; p. 89.
3. Liu, J.; Shi, M.; Jiang, H. Detecting Suicidal Ideation in Social Media: An Ensemble Method Based on Feature Fusion. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8197. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Venek, V.; Scherer, S.; Morency, L.P.; Pestian, J. Adolescent suicidal risk assessment in clinician-patient interaction. *IEEE Trans. Affect. Comput.* **2017**, *8*, 204–215. [\[CrossRef\]](#)
5. O'Connor, R.C.; Nock, M.K. The psychology of suicidal behaviour. *Lancet Psychiatry* **2014**, *1*, 73–85. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Liang, Y.; Yang, L. Study on Reliability and Validity of the Suicide Probability Scale. *Chin. J. Health Psychol.* **2010**, *2*, 225–227. [\[CrossRef\]](#)
7. Mo, W. Various effects and solutions in psychology experiments. *Exp. Sci. Technol.* **2008**, *6*, 118–121.
8. Cong, G.; Ji, S.; Yu, C.P.; Fung, S.f.; Pan, S.; Long, G. Supervised Learning for Suicidal Ideation Detection in Online User Content. *Complexity* **2018**, *2018*, 6157249. [\[CrossRef\]](#)
9. Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horiz.* **2010**, *53*, 59–68. [\[CrossRef\]](#)

10. O'Dea, B.; Wan, S.; Batterham, P.J.; Caelear, A.L.; Paris, C.; Christensen, H. Detecting suicidality on Twitter. *Internet Interv.* **2015**, *2*, 183–188. [[CrossRef](#)]
11. Huang, Y.P.; Goh, T.; Liew, C.L. Hunting Suicide Notes in Web 2.0—Preliminary Findings. In Proceedings of the Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007), Taichung, Taiwan, 10–12 December 2007; pp. 517–521. [[CrossRef](#)]
12. Jashinsky, J.; Burton, S.H.; Hanson, C.L.; West, J.; Giraud-Carrier, C.; Barnes, M.D.; Argyle, T. Tracking suicide risk factors through Twitter in the US. *Crisis J. Crisis Interv. Suicide Prev.* **2014**, *35*, 51–59. [[CrossRef](#)]
13. Coppersmith, G.; Leary, R.; Whyne, E.; Wood, T. Quantifying suicidal ideation via language usage on social media. In Proceedings of the Joint Statistics Meetings, Seattle, WA, USA, 8–13 August 2015; Volume 110.
14. Cheng, Q.; Li, T.; Kwok, C.; Zhu, T.; Yip, P. Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. *J. Med. Internet Res.* **2017**, *19*, e243. [[CrossRef](#)]
15. Aladağ, A.E.; Muderrisoglu, S.; Akbas, N.B.; Zahmacioglu, O.; Bingol, H.O. Detecting Suicidal Ideation on Forums: Proof-of-Concept Study. *J. Med. Internet Res.* **2018**, *20*, e215. [[CrossRef](#)]
16. Huang, Y.; Liu, X.; Zhu, T. Suicidal Ideation Detection via Social Media Analytics. In *Human Centered Computing*; Milosevic, D., Tang, Y., Zu, Q., Eds.; Springer: Cham, Switzerland, 2019; Volume 11956, pp. 166–174. [[CrossRef](#)]
17. Abboute, A.; Boudjeriou, Y.; Entringer, G.; Azé, J.; Bringay, S.; Poncelet, P. Mining Twitter for Suicide Prevention. In *Natural Language Processing and Information Systems*; Springer: Cham, Switzerland, 2014; pp. 250–253. [[CrossRef](#)]
18. Coppersmith, G.; Ngo, K.; Leary, R.; Wood, A. Exploratory Analysis of Social Media Prior to a Suicide Attempt. In Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego, CA, USA, 16 June 2016; pp. 106–117. [[CrossRef](#)]
19. Shing, H.C.; Nair, S.; Zirikly, A.; Friedenberg, M.; Daumé, H., III; Resnik, P. Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA, USA, 5 June 2018; pp. 25–36. [[CrossRef](#)]
20. Yinghong, D.; Hao, C.; Kaisheng, L.; Guoan, Y. Weibo Social Moods Measurement and Validation. *J. Psychol. Sci.* **2015**, *38*, 1141–1146.
21. Ren, X.; Xiang, Y.; Zhou, Y.; Zhu, T. A Mental Map of Chinese Individualism/Collectivism Based on Microblogging Big Data. *J. Inn. Mong. Norm. Univ. Soc. Sci. Ed.* **2017**, *46*, 59–64.
22. Graham, J.; Haidt, J.; Nosek, B. Liberals and Conservatives Rely on Different Sets of Moral Foundations. *J. Personal. Soc. Psychol.* **2009**, *96*, 1029–1046. [[CrossRef](#)]
23. Zhang, Y.; Yu, F. Which socio-economic indicators influence collective morality? Big data analysis on online Chinese social media. *Emerg. Mark. Financ. Trade* **2018**, *54*, 792–800. [[CrossRef](#)]
24. Lv, M.; Li, A.; Liu, T.; Zhu, T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. *PEERJ* **2015**, *3*, e1455. [[CrossRef](#)]
25. Tausczik, Y.; Pennebaker, J. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. Lang. Soc. Psychol.* **2010**, *29*, 24–54. [[CrossRef](#)]
26. Cull, J.G.; Gill, W.S. *Suicide Probability Scale Manual*; Western Psychological Services: Los Angeles, CA, USA, 1982.
27. Carlborg, A.; Jokinen, J.; Nordstrom, A.; Jonsson, E.; Nordstrom, P. Attempted suicide predicts suicide risk in schizophrenia spectrum psychosis. *Nord. J. Psychiatry* **2010**, *64*, 68–72. [[CrossRef](#)]
28. Lin, L.; He, H.; Liu, Y.; Yang, Y.; Mo, J.; Wang, C.; Liu, T. Effect of negative emotions on suicide risk/mediating role of self-injurious behavior. *Heilongjiang Res. High. Educ.* **2019**, *37*, 83–87.
29. Lester, D. Individualism and suicide. *Psychol. Rep.* **2005**, *97*, 576. [[CrossRef](#)] [[PubMed](#)]
30. Silver, E.; Silver, J.; Sigfusdottir, I. Moral Intuitions and Suicide Risk: Results from a National Sample of Icelandic Youth. *Soc. Forces* **2021**, *99*, 1799–1826. [[CrossRef](#)]
31. Gao, R.; Hao, B.; Li, H.; Gao, Y.; Zhu, T. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. In *International Conference on Brain and Health Informatics*; Imamura, K., Usui, S., Shirao, T., Kasamatsu, T., Schwabe, L., Zhong, N., Eds.; Springer: Cham, Switzerland, 2013; Volume 8211, pp. 359–368.
32. Ptaszynski, M.; Zasko-Zielinska, M.; Marcinczuk, M.; Leliwa, G.; Fortuna, M.; Soliwoda, K.; Dziublewska, I.; Hubert, O.; Skrzek, P.; Piesiewicz, J.; et al. Looking for Razors and Needles in a Haystack: Multifaceted Analysis of Suicidal Declarations on Social Media—A Pragmalinguistic Approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11759. [[CrossRef](#)] [[PubMed](#)]
33. Bao, P.; Shen, H.-W.; Jin, X.; Cheng, X.-Q. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes. In Proceedings of the WWW'15 Companion: Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; Association for Computing Machinery: New York, NY, USA, 2015.
34. Litvinova, T.; Litvinova, O.; Seredin, P. Dynamics of an idiostyle of a Russian suicidal blogger. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA, USA, 5 June 2018; pp. 158–167. [[CrossRef](#)]
35. Coppersmith, G.; Dredze, M.; Harman, C. Quantifying mental health signals in Twitter. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Baltimore, MD, USA, 16 June 2014; pp. 51–60.

36. del Pilar Salas-Zárate, M.; López-López, E.; Valencia-García, R.; Aussenac-Gilles, N.; Almela, Á.; Alor-Hernández, G. A study on LIWC categories for opinion mining in Spanish reviews. *J. Inf. Sci.* **2014**, *40*, 749–760. [[CrossRef](#)]
37. Guan, L.; Hao, B.; Liu, T.; Cheng, Q.; Yip, P.; Zhu, T. A pilot study of differences in behavioral and linguistic characteristics between Sina suicide microblog users and Sina microblog users without suicide idea. *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi* **2015**, *36*, 421–425.
38. Reynolds, W.M. *Suicidal Ideation Questionnaire (SIQ)*; Psychological Assessment Resources: Odessa, FL, USA, 1987.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.