

Research Article

Analysis of Public Opinion in Colleges and Universities Based on Wireless Web Crawler Technology in the Context of Artificial Intelligence

WenNing Wu ¹ and ZhengHong Deng²

¹International Cooperation Department, School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China

²School of Automation, School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China

Correspondence should be addressed to WenNing Wu; wuwenning@nwpu.edu.cn

Received 11 November 2021; Accepted 27 December 2021; Published 11 January 2022

Academic Editor: Sang-Bing Tsai

Copyright © 2022 WenNing Wu and ZhengHong Deng. 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.

Wi-Fi-enabled information terminals have become enormously faster and more powerful because of this technology's rapid advancement. As a result of this, the field of artificial intelligence (AI) was born. Artificial intelligence (AI) has been used in a wide range of societal contexts. It has had a significant impact on the realm of education. Using big data to support multistage views of every subject of opinion helps to recognize the unique characteristics of each aspect and improves social network governance's suitability. As public opinion in colleges and universities becomes an increasingly important vehicle for expressing public opinion, this paper aims to explore the concepts of public opinion based on the web crawler and CNN (Convolutional Neural Network) model. Web crawler methodology is utilised to gather the data given by students of college and universities and mention them in different dimensions. This CNN has robust data analysis capability; this proposed model uses the CNN to analyse the public opinion. Preprocessing of data is done using the oversampling method to maximize the effect of classification. Through the association of descriptions, comprehensive utilization of image information like user influence, stances of comments, topics, time of comments, etc., to suggest guidance phenomenon for various schemes, helps to enhance the effectiveness and targeted social governance of networks. The overall experimentation was carried out in python here in which the suggested methodology was predicting the positive and negative opinion of the students over the web crawler technology with a low rate of error when compared to other existing methodology.

1. Introduction

In present, expressing the opinion through reviews has become crucial for public opinion. On a whole, resources are gathered from various sources and fast updated and unique opinion from students across universities and colleges. Technical models are established for significant acquainting of the role of public opinion and enhancing the effectiveness of social governance. Different stage descriptions of opinion subject on the basis on public reviews through opinion that will support for accurate decision-making for social domination. With enormous enhancement of artificial models, user descriptions have become a crucial means to manage

the user characteristics [1]. As wireless communication and artificial intelligence technologies mature, artificial intelligence-based decision support systems have emerged as a new hot research area. Operational research, cybernetics, and behavioural sciences are all included into the decision support system (DSS), which use computer simulation and other technological methods to aid in the decision-making process by providing decision-makers with more accurate and efficient information. Artificial intelligence-based decision support systems play a critical role in education reform in the information era. Based on the opinion provided by the college students, user details can be acquired and used for accurate behavioural analysis of students and their

habits. Crucial details like attitudes, opinions, etc., give efficient data basis [2]. The user description obtained mainly aims on providing significant recommendations in the area of public and business services. The details obtained in the area of social systems mainly aims on three different aspects. The initial focus is on the user classification and characteristics to enhance the necessity of services based on network information. For instances, a model based on user descriptions is established to divide users into groups and perform significant marketing on the basis of each group classification [3]. Considering multiple aspects, user descriptions are utilised to classify user belonging to multiple conversion rates for enhancing the significance of social media procedures. The second is to analyse the emotional behaviour and their wishes. They utilised user willingness of people to manage the actions of social media, utilizing user details models to analyse on their stage characteristics in the procedure of managing social media. The third model is to analyse the details distribution behaviour and guidance opinion of public measuring the user descriptions. For instances, a user-based description and comments of users are utilised in enhancing the effect on antiterrorism-based opinion by public and analysed the outcome. The description model utilised the outcome of the user details characteristics in detail, which provides a vital reference for discovering detailed behaviour.

Moreover, some researchers have gathered user description by exploiting the social media around the guidance of public opinion; every description dimension is significantly independent and lacks in integrated performance. Therefore, suitability and effectiveness of opinion strategy have developed the need for improvement. Based on the existing research, this paper is aiming to govern the social network based on public opinion and based on the public subjects; the model focuses on the attitudes and opinions. In various dimensions of public comments, processing using oversampling provides solution for unbalanced user description and enhances the throughout portrait effect and accuracy.

Particularly in the era of mobile internet, the reviews of students have become spread in an enormous way. To a certain extent, the reviews have provided the negative significance and even become a threat to government organizations safety, moreover, in order to filter the network details by which negative feedback need to be eradicated from the reviews, and further improvise the capability of governments to tackle the crisis. Therefore, taking specialized applications into consideration, analysis of student's opinion should be given extreme importance. As an outcome, such apprehensions have two advantages, which help the government organizations to eradicate the crisis in a timely manner and comprehensively guide the public sectors.

When classification of opinions comes to consideration, the neural network comes to picture, where fast enhancement of CNN has enormous effect on data analysis on the basis of test sample given for classification. This CNN has wide range of advantages by which the performance of training sample has improved and further enhanced the data analysis capability.

With consideration of public opinion, this proposed model utilises the CNN to analysis of reviews and comments in the name of opinion of students. Moreover, this paper focuses on considering the campus environment of students. Based on the information provided in terms of public opinion, the monitoring system needs to be addressed. In order to gather the analysed outcome, the public opinion can be transferred to the server from which the analysis can be carried easily. Among the higher education system, large total of means were utilised to contact participants who are involved in the potential examination on opinion, particularly if the participants and sample are collected from students of various universities and colleges. The data availability of user and applications makes description largely personal. The study examined in this article analysed the effect of personalization of questionnaires and lengthy messages on review responses.

2. Literature Review

Joinson et al. [4] stated that based on the given capabilities of mechanism for getting the opinion, particularly the surveyor ability to utilise the database for analysis, a better deal for personalization is possible. Fan and Yan [5] described that crucial part of invitation design is the user personalization. For instance, Joinson and Reips [6] analysed the three various experiments on email personalization for invitation salutation. On the basis of web-based opinion gathering, user personalization enhanced the rate of response given by student community. Heerwegh [7] categorised 2500 college students into two set. The handling received the objective salutation and the received group treatment including the final and initial names. Outcome depicted that personalization effectively enhances the rate of web survey. Another analysis observed a huge rate of response due to personalization, but they are not significantly statistical. Pearson and Levis [8] identified significant association amongst salutation and preferred age. Including the confidential questions in a review provides various effects. Heerwegh et al. [7] identified that email invitation personalization grabs more answers related to social events. In addition to that, it might be possible to assume that large personalized emails will enhance the rate of response on survey based on webs.

While gathering opinion from the students, most crucial parameter in context to increase the rate of response is the content type sent by the students; Porter et al. [9] stated that survey opinion may be impacted based on the invitational messages. One such instance was investigated by Klofstad et al. [10] when compared with various length of email. Authors stated that there does not exist any significant association amongst a shorter email with little details about the survey. Across various sectors, email survey has been identified as a crucial application for marketing due to low rate of investment and large rate of response. Retti and Chittenden [11] studied the association amongst rate of response and email length. Beside the email length, it is vital to consider other parameters in the invitation email. For example, the email sender is a parameter that may trigger the email rate of response. Joinson et al. [4] identified that the

large rate of response occurred when the student gets the authorization position. One study conducted by Porter and Whitcomb [12] stated the methodology on web survey response and viewing it are significantly affected by the details in the line of review obtained through emails. Having two various group based on the range of involvement with the survey sponsors, researchers did not identify the effective differences among the different factors mentioned in the line of subject mentioned in the contact reason and the sponsor mentioned.

Ducheneaut and Bellotti [13] stated that increase in the rate of responses given by students increases based on the shorter opinions rather than longer ones. However, it is significantly accepted to mention that longer description is better from participants point of view with large information and is more trustworthy. Uncertainty about the length of messages may be dependent on the rate of responses given to questions through research. Fan [14] stated that for adopting the social media, user description methodology is utilised to conduct detailed research on the characteristics in the procedure of social media analysis. Based on the user portraits, study of detailed behaviour and guidance measure is measured. Lu et al. [15] investigated user description on Weibo information and survey to enhance the guiding strategy of terrorism opinion obtained by public and associated the portraits outcome.

Even though certain scholars have analysed social media for gathering user-based information around guiding public opinion in this research model, every dimension obtained from portrait is significantly independent and lacks embedded analysis. Lin and Xie et al. [3] developed the user-based portraits method to categorise Weibo user into sets and carried out marketing accuracy based on every group's characteristic. Zhao [16] utilised portraits methodology to classify users with large conversion proportion and analysed the behaviour of group from various angles in order to enhance the operational effect of social media. Wei et al. [17] studied the behaviour of student from social media from multiple angles. Another study was done by utilizing the wishes and emotional opinions from students. Gao [18] stated that user portraits have become vital means for doing analysis of user characteristics. Yunfei and Weizhu [19] provided significant abstract that gathers user information for further analysis which is considered technically important.

3. Construction of Framework Based on Student Opinion Subject Description

Data gathering consists of many time-consuming and complex activities. These include proxy management, data parsing, infrastructure management, overcoming fingerprinting antimeasures, rendering JavaScript-heavy websites at scale, and much more. Finding a more manageable solution for a large-scale data gathering has been on the minds of many in the web crawling community. Specialists saw a lot of potential in applying AI (artificial intelligence) and DL (deep learning) to web scraping. However, only recently, actions toward data gathering automation using AI

applications have been taken. This is no wonder as AI and DL algorithms became more robust at large scale only in recent years together with advancement in computing solutions. By applying AI-powered solutions in data crawling, we can help automate tedious manual work and ensure a much better quality of the collected data.

User description basically includes four concepts like gathering user information, screening user dimension, analysing data modelling, and presentation of description structure. Gathering information about data in this framework is mainly crawled from various Weibo platform using crawler methodologies. After crawler, it is categorised into text and data types. Because of various formats in the gathered opinion and missing values, various types of details need to be preprocessed individually. Based on the purpose of user description, user categories need to be identified. Based on the guidance of public opinion, three different parameters are selected like attributes, position of comments, and topic of comments; analysis of data modelling is to form the description framework for considered dimensions; description structure presentation is to depict the outcome in a spontaneous way for application and analysis.

The four links presented are processed in the sequential order to gather a flowchart of the student opinion. Figure 1 illustrates the entire procedure of categorizing that involves five steps: acquisition of data, type of data, preprocessing of data, profile method establishment, and profile visualization. Most particularly among them is the data acquisition and division which involves utilization of data crawler methodology to crawl detailed information of the students, content of comment, and other data types of Weibo users to form the source of data of the user description; before proceeding to portrait, previewing of data is very significant task. The process of preprocessing is mainly used to find the missing values; then the user-based model is obtained from different dimensions of basic parameters. The following description outcomes display and fusion using visualization methodology. The method focuses on the methods and principles of three various dimensions in profile; attributes position of comments; and theme of comment.

3.1. Data Preprocessing. Preparation of data is the first and most fundamental phase in the data mining process. A lack of a single data source and a nonstandard format will have a substantial influence on classification results due to noisy data, as well as the existence of incorrect information. Crawling often encounters invalid and erroneous data, which are seldom utilised in the analysis process or cause substantial difficulties in the findings. These data must be preprocessed before they can be used in the analysis. Most of the data preprocessing removes repetitive data, expressive words, short phrases, nonsense, and unclear language. Data preparation involves removing stop words from public opinion data and filtering out noise.

3.1.1. Missing and Duplicate Information. Few data points in the dataset are difficult to comprehend because of the content's grassroots, self-determination, and high repeat

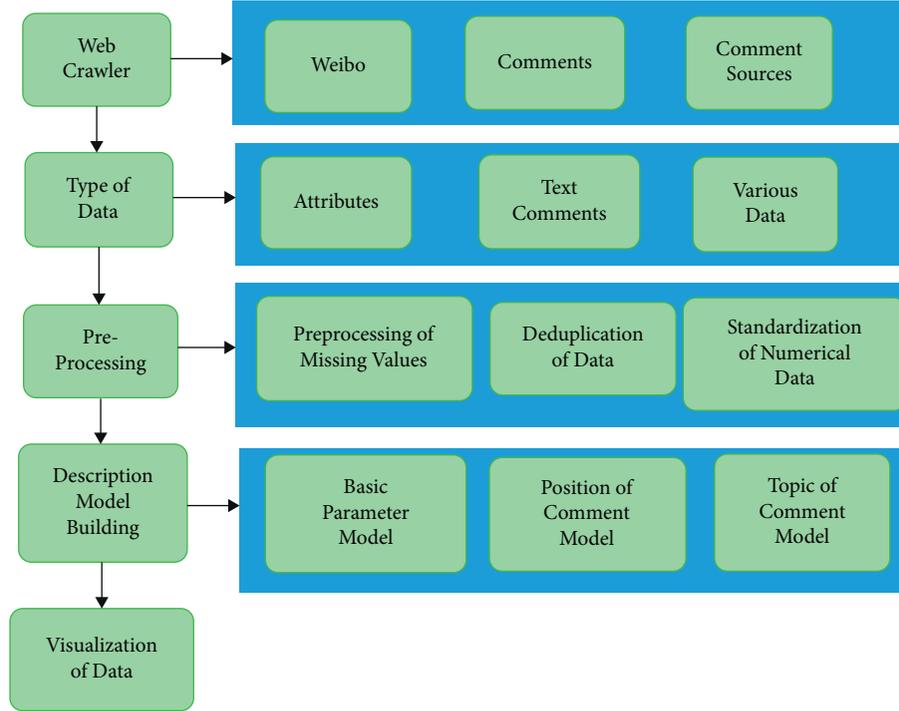


FIGURE 1: Flowchart of the online opinion system.

rate. To put it another way, future researchers will need to know how to weed out the irrelevant data. Data mining and data selection on the basis of their own unique format help filter out duplicates and fill in the value that is been left out by analysing the original data. A stop word is a term that lacks nuance or nuanced description (such as “bar,” “and,” and “have”). These words, by definition, have no use in text processing. Because these words exist in a wide variety of texts with the same frequency, they must be eliminated from the text feature vector. The training and test sets are formed when the experimental data are collected on the basis of attribute details of students which includes gender, name, signature followers, total of fans, time of comments, etc. Among these, signature, name, gender, and time of comment can be strictly utilised as description tags and total followers, etc. The Fame Index (FI) was stated by authors in Ref. [20] that includes considering the user’s popularity. BX illustrates the total user followers; DS depicts the total user’s fans and M illustrates the sample research carried. User participated in the comment event is given by

$$FI = \frac{DS}{BX} + \frac{DS}{M}. \quad (1)$$

While conducting public opinion, sentiment analysis is a vital part that various researchers often utilise it while carrying research and based on the comments given by the participants in opinions. Moreover, sometimes the outcome of the analysis performed is used to handle public opinions and significantly affect the analysis. For instance, the core opinions of students with similar sentiment classification may be entirely opposite. This article chooses the dimension position of comments given by students for content analysis.

The position of comment is a crucial indicator of users. It can be found from the users’ comments or information of events. In this article, opinion position is categorised into negative and positive opinions, another one being neutral. A positive line depicts to a view that intimates the sympathy, positive attitude toward the occurrence of incident while taking review from students. The automatic classification framework for user position is well-known by using neural networks. Opinions obtained from students in the form of comments are divided into test and training groups and then the classification is done using training data. The formed classification methodology is used for training and the proposed model is utilised with testing data. When the performance of the proposed model is as expected, the commentary user positions can be marked and classified with label positions.

3.2. Classification. The neural network-based classification framework requires labelled data. Therefore, prior to model formation, the comment of user details needs to be marked and commented with negative, positive, and neutral mark. Positive depicts the opinion received in a positive manner from student’s opinion; negative illustrates the opinion received negatively from student’s community and incident behaviour. The opinion gathered from the student is in the form of text details and preprocessing is essential before learning classification from the neural network model. The procedure is mainly categorised into variety of steps. It is understood that various CNN methods have different network formations. Every CNN model consists of three layers such as pooling, convolution, and a full connection layer. The typical CNN-based structure is depicted in Figure 2.

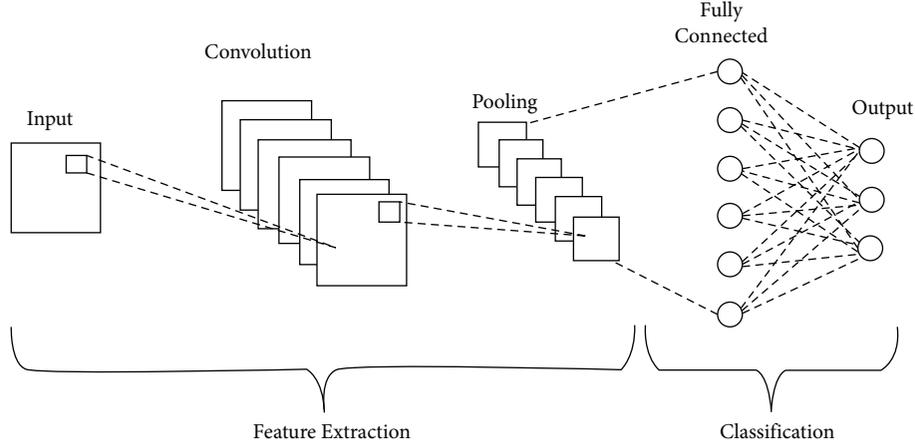


FIGURE 2: CNN structure.

Here, we are adding the extra encoder and decoder can be added along with the channel attention mechanism so that the required features can be easily isolated. Moreover, concepts like weight sharing and local perception under the convolution layer is utilised for feature extraction from the given input data. In this model, local perception depicts to regions continuous function. It is capable to percept the

opinion given by students in the colleges. Sharing of weight focuses to minimize the computing complexity of CNN by avoiding the extraction of feature and establishment of data. Let us assume that $dout_i$ illustrates the analysis of public opinion in terms of i th graph formed from known students' opinion about comments. And it is given by

$$dout_i(a, b) = \delta \left(\sum_{j=1}^m \sum_l \sum_k din_j(a + j, b + k) * filter_{j,i}(l, k) + off_i \right). \quad (2)$$

Moreover, m depicts the number of total students taking part in the opinion; din_j illustrates the feature rate of j th formed graph using features obtained; $filter_{j,i}$ depicts the weight required for the neural network; off_i depicts the offset; and $\delta(a)$ represents the activation function.

Significantly, upon every extracted feature is completed, for minimizing the student's size for gathering the opinion and enhancing the efficiency of training, pooling procedure is essential. Then, a convolutional layer is included to outcome of the final analysis obtained from public opinions, which also subjects on the activation function as follows:

$$dout(i) = \delta \left(\sum_{j=1}^m din(j) * filter_i(j) + off_i \right). \quad (3)$$

Considering the proposed public opinion process, activation function is treated as very essential part. The accuracy is estimated based on the sigmoidal error rate for activation function. Therefore, error function is estimated as given below, where a_j illustrates the input signal and ϵ_k depicts the significant rate of output error in terms of present layer and which is expressed using

$$a_j = \delta(b_k - 1), \epsilon_k = \delta(b_k - 1) * er_{j-1}, \quad (4)$$

where $b_k - 1$ is the outcome signal obtained in the final layer and er_{j-1} is the outcome of last layer. Since the sigmoid

function's monotonicity is from the absolute rate, the following equation is satisfied:

$$\frac{\epsilon_k^2}{a_k^2} = \frac{\delta(b_k - 1)^2 * er_{j-1}^2}{\delta(b_k - 1)^2} < \frac{er_{j-1}^2}{b_{k=1}^2}, \quad (5)$$

which signifies that the function utilised does not cause any weakening of error in output. Moreover, for a CNN with K layers, the output error in the upper limit is expressed as follows:

$$NSR < \sum_{k=1}^K \frac{er_{j-1}^2}{b_{k=1}^2}. \quad (6)$$

To total up, if the error formed from the activation function satisfies and expressed in the given (6), the student's opinion analysis is significantly correct.

The basic processes illustrate the feature extraction as shown in Algorithm 1.

4. Opinion Vector Feature Clustering

In this proposed model, cosine similarity is utilised to estimate the similarity amongst two opinions in term of text. For every text o_j and o_i , their similarity is estimated as follows:

Input: parts of opinion from college and university students after preprocessing. The total feature used K that illustrates the vector length of every feature.

Output: feature Vector

Step 1: on the given opinion, estimation of frequency for every opinion and the outcome is stored by using hash table where the opinion index is given as key and value is the term frequency. The opinion index is estimated as follows:

index = hashfunc(opinion)% K . Hash table is utilised to additional vector which depicts the CNN vector of opinion o_j .

Step 2: on every opinion, estimate the vector frequency of the opinion of, for every opinion o as follows:

for o_j in this opinion O :

for index in this o_j

of (index)+ = 1:

end

end

Step 3: on the opinion given by students, aggregate frequency for every vector is obtained from all the opinion obtained by students and values to the corresponding opinion are accumulated to get the vector of, including of to every node in the network.

Step 4: on every opinion, estimate the CNN feature vector for every opinion using (3).

ALGORITHM 1: CNN-based feature extraction.

$$\text{sim}(o_j, o_i) = \frac{\overline{o_i} \cdot \overline{o_j}}{\|\overline{o_j}\| \|\overline{o_i}\|}. \quad (7)$$

If the threshold is lesser than the similarity, two opinions in form of text are considered as clustering.

5. Result Analysis and Discussion

5.1. Data Source and Acquisition. Utilizing the web crawler methodology, we can obtain significant data by crawl of student's opinion in the form of text. It mainly consists two forms of data. One is the normal details of students, consisting age, gender, number of comments, total likes obtained, etc.; the second is the text opinion of the prevalent Weibo which basically reflects the student's view and the position. These data form the significant source for making analysis on the public opinion from various students.

5.2. Outcome of Portrait Position of Weibo Commentary. This article depicts the function of training the CNN-based model for analysis; marking of position is done manually to the text given by the students. In the marked 500 data available, the position is marked as neutral, negative, and positive. In order to enhance the performance of classification model proposed, the oversampling module was utilised for data oversampling. The outcome area under curve (AUC) range reached 90.5%, which enhanced in association with no sampling. The ROC curves of the proposed classification framework are significant before and after the oversampling process. It can be depicted that the rate of AUC after sampling has been enhanced efficiently, signifying that the proposed framework after training on the data has enhanced.

The model used for training is utilised to classify the 11000 opinions through comments which are crawled. For training, 70% of the data are taken for training and 30% of the data are taken for testing. Among them, the labelled comments like positive, neutral, and negative are 5000, 3000, and 3000 individually. Based on some incident, there exist

campus with some opposing views and significant estimation need to be considered in a timely manner to proceed out the opinion governance.

By taking the information of time of every commentary opinion, it is called that most of the opinion were available on the next day after changing the context of opinion and the entire changes of every position in comment were taken into count by the time of gathering commentary positions. The rate of growth has slowed down gradually, and number of opinions obtained from students a plateau after 10:00 which respect to the time dimension of student's opinion; the duration of sudden enhancement from 7 am to 8 am is the significant period which is illustrated in Figure 3.

The analysis depicts that online public comments obtained from student's opinion describe on the basis of huge data for the purpose of providing the performance accuracy for improvement of public governance, significantly considering the outcome of different portrait dimensions and Weibo influences, position of comments obtained through opinion, and review information like time of comment, subject, etc., propose governance of public opinion strategies for various methodology. While determining the primary subjects, choose an association model of commentary lines and index reputation. The CNN-based classification model is framed on the basis of various comments in content. In student opinion, focus needs to take care of comments which are negatively posted.

Figure 4 illustrates the various comment types in different stages. The proportion of positive opinion is shown in stage 4, non-negative opinion is depicted in stage 3, negative comment is illustrated in stage 2, and neutral comment is illustrated in stage 1. As on a whole, there exists huge positive opinion than negative opinion. Based on the life cycle of the model proposed, the opinion obtained neutrally has gradually improved. The percentage of negative opinion sharply went down from 30% in the initial stage to 20% in the final stages and marginally enhanced by 1% in last phases. Figures 5 and 6 illustrate the total of comments in phase 1 and phase 2 which is positive opinion and negative opinion, respectively. Confusion in the opinion of students

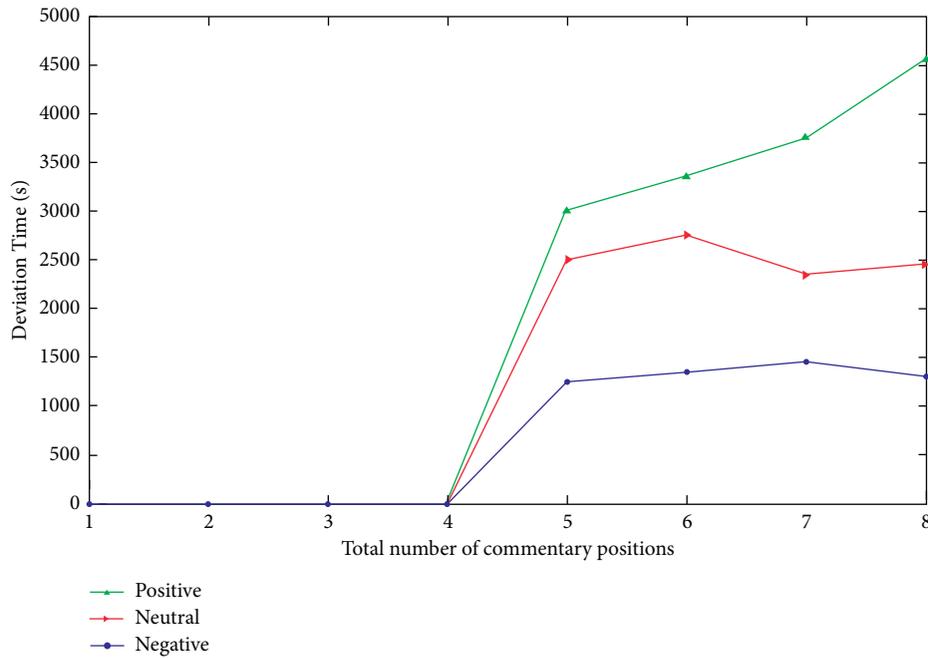


FIGURE 3: Deviation in total number of positions in commentary over a period of time.

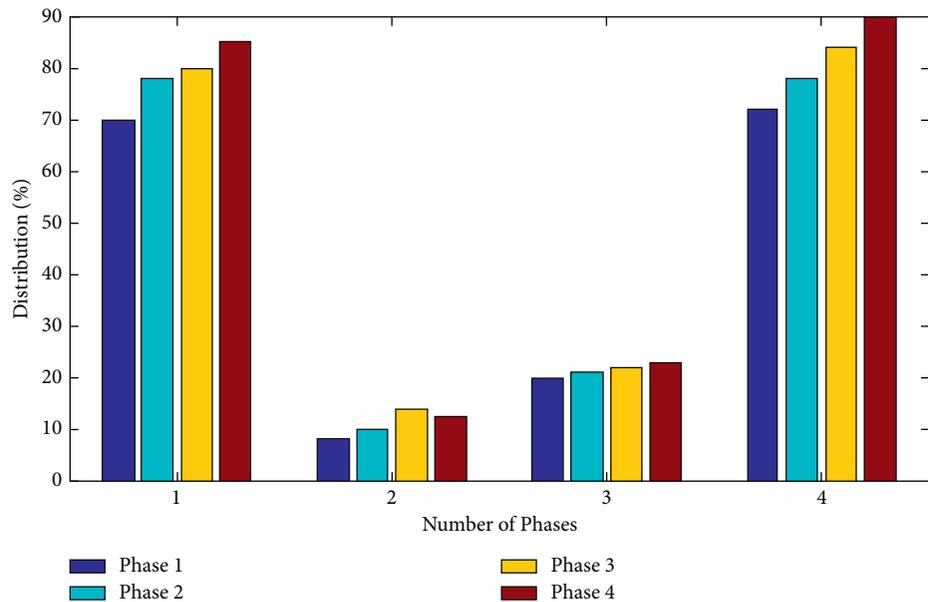


FIGURE 4: Distribution of various types of opinion in four phases of the proposed model.

has some deviation in the comments than the non-negative opinion obtained by students.

The feelings expressed in a phrase are denoted by the nouns and adjectives that accompany them. It is hoped that by emphasising the importance of words and their adjectival forms in sentences, the suggested method would yield more distinct terms for use in future opinion research (Figures 7 and

8 and Table 1). In addition, the sheer volume of feedback on the test set raises the emotional precision by a factor of two because the completion of each phrase signified the strength of oness position. Over opinion analysis, the suggested method has a lower error rate. With the use of various performance indicators, it is possible to demonstrate that the proposed classifier is more efficient than the current techniques [20].

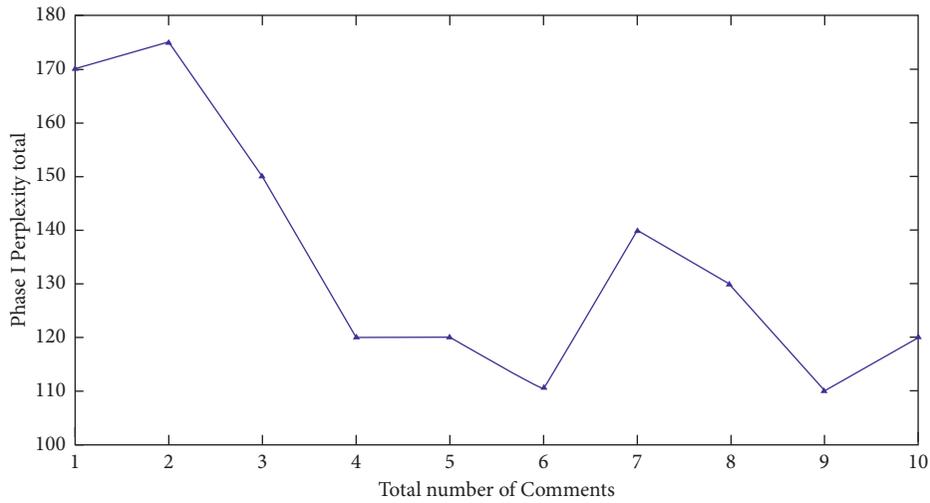


FIGURE 5: Perplexity total of comments in phase 1.

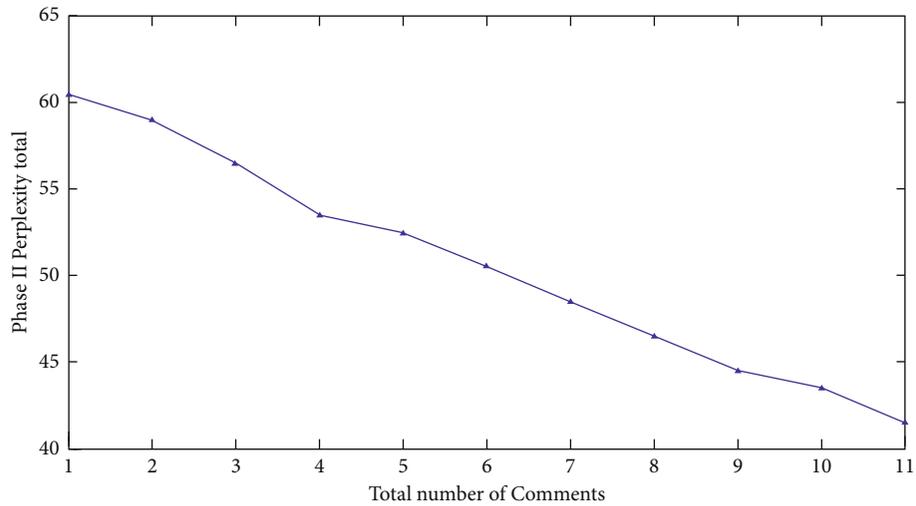


FIGURE 6: Perplexity total of comments in phase 2.

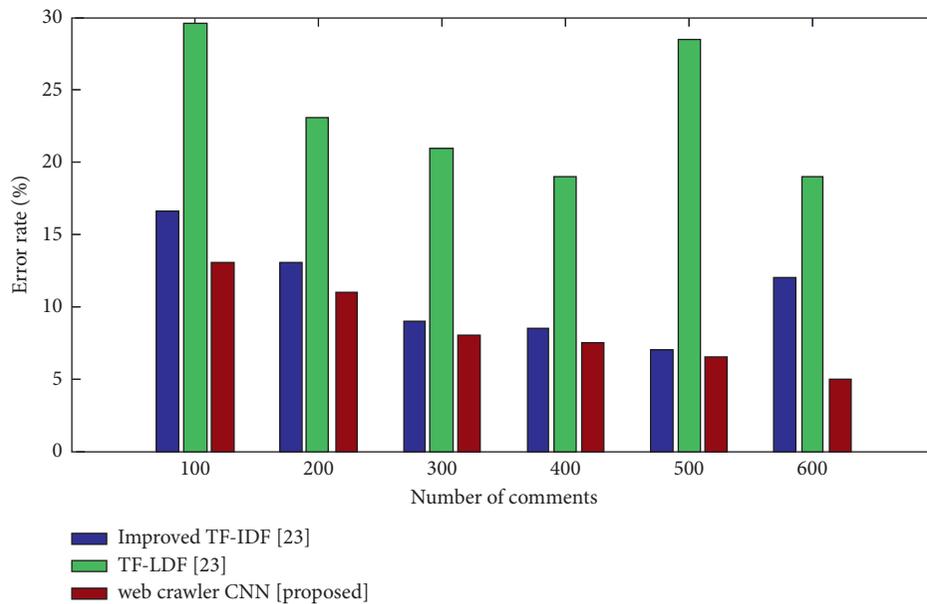


FIGURE 7: Number of comments vs. error rate (positive opinion).

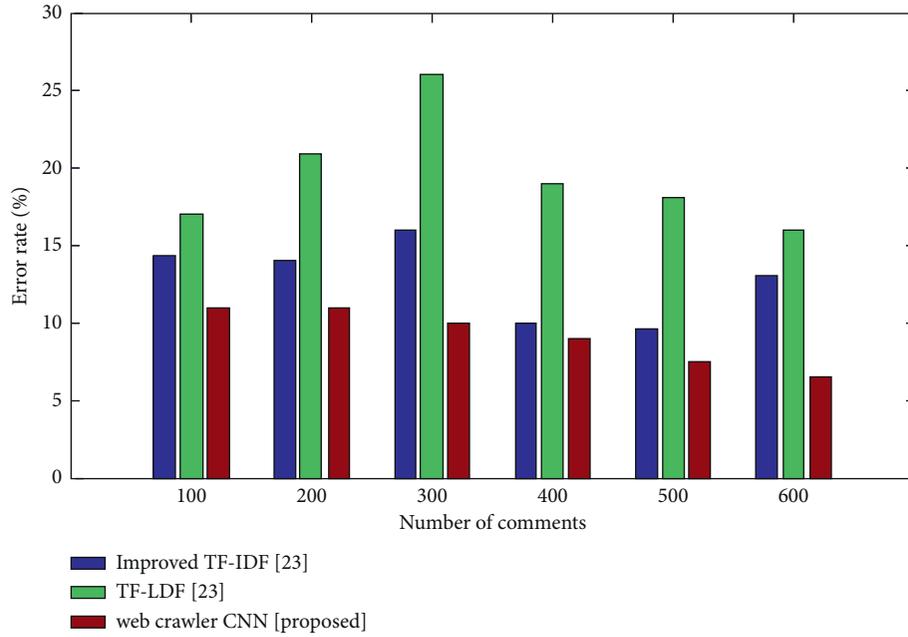


FIGURE 8: Number of comments vs. error rate (negative opinion).

TABLE 1: Opinion analysis.

Number of comments	Error rate (%)					
	Positive opinion		Negative opinion			
	Improved TF-IDF [20]	TF-LDF [20]	Web crawler CNN (proposed)	Improved TF-IDF [20]	TF-LDF [20]	Web crawler CNN (proposed)
100						
200	16.5	29.5	13	14.5	17	11
300	13	23	11	14	21	11
400	9	21	8	16	26	10
500	8.5	19	7.5	10	19	9
600	7	28.5	6.5	9.5	18	7.5
	12	19	5	13	16	6.5

6. Conclusion

In the age of internet, the student’s opinion has been significant in an explosive manner. To certain extent, the college student’s comments have brought an effective approach for making an analysis. Various decadent trends of thought have caused very serious negative effects on the healthy development of country’s society, economy, and culture. At the same time, country’s tendency to one-sidedly attaches importance to data crawling after reform and opening up has also caused us to pay a very heavy price. With enormous development of neural network techniques, certain data analysis based on the training sample has been well addressed. Moreover, this paper proposes the CNN-based framework to complete the public opinion of students of various universities. Furthermore, considering the trouble of student’s opinion, this model proposed an oversampling-based CNN case

study under college environment, and the model monitors the complexity of public opinion which includes opinion collection, submission of analysis results, and guidance. Several methods to be proposed but each of them having the advantages and the disadvantages like it will obtain high range of accuracy over opinion prediction, but the computational time taken for the mining process was high. Moreover, based on the experimental analysis the initial performance accuracy is simulated based on different phases like positive, negative, etc. The gathered dataset is utilised to train the CNN-based monitoring network which is evaluated by involving the ratio of guidance. The simulation outcome depicts that this proposed model outperforms significantly and the student will express higher positive opinion over the web crawler technology; that is, out of 11000, 5000 of them gave positive comments over the web crawler technology. But the main limitation of our proposed work is it is data dependent. It needs large

number of the data to complete the classification process. Our future mission is to analyse the opinion of the student over their education in their schools or colleges by using the optimization technologies.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] G. Gao, "A review of research on user portrait construction methods," *Data Analysis and Knowledge Discovery*, vol. 3, no. 3, pp. 25–35, 2019.
- [2] Y. Qiu and W. Zhang, "Research on group portrait construction method based on network structure and text content," *Library and Information Service*, vol. 63, no. 22, pp. 21–30, 2019.
- [3] Y. Lin and X. Xie, "Microblog group user portrait based on social identity theory," *Information Theory and Practice*, vol. 41, no. 3, pp. 142–148, 2018.
- [4] A. N. Joinson, A. Woodley, and U.-D. Reips, "Personalization, authentication and self-disclosure in self-administered Internet surveys," *Computers in Human Behavior*, vol. 23, no. 1, pp. 275–285, 2007.
- [5] W. Fan and Z. Yan, "Factors affecting response rates of the web survey: a systematic review," *Computers in Human Behavior*, vol. 26, no. 2, pp. 132–139, 2010.
- [6] A. N. Joinson and U.-D. Reips, "Personalized salutation, power of sender and response rates to Web-based surveys," *Computers in Human Behavior*, vol. 23, no. 3, pp. 1372–1383, 2007.
- [7] D. Heerwegh, "An investigation of the effect of lotteries on web survey response rates," *Field Methods*, vol. 18, no. 2, pp. 205–220, 2006.
- [8] J. Pearson and R. A. Levine, "Salutations and response rates to online surveys," in *Proceedings of the Fourth International Conference on the Impact of Technology on the Survey Process*, University of Warwick, Coventry, UK, September 2003.
- [9] S. R. Porter, M. E. Whitcomb, and W. H. Weitzer, "Multiple surveys of students and survey fatigue," *New Directions for Institutional Research*, vol. 2004, no. 121, pp. 63–73, 2004.
- [10] C. A. Klofstad, S. Boulianne, and D. Basson, "Matching the message to the medium," *Social Science Computer Review*, vol. 26, no. 4, pp. 498–509, 2008.
- [11] R. Rettie and L. Chittenden, *Email Marketing: Success Factors*, Kingston Business School, Kingston University, Kingston upon Thames, UK, Occasional Paper Series No 50, 2003.
- [12] S. R. Porter and M. E. Whitcomb, "E-mail subject lines and their effect on web survey viewing and response," *Social Science Computer Review*, vol. 23, no. 3, pp. 380–387, 2005.
- [13] N. Ducheneaut and V. Bellotti, "E-mail as habitat," *Interactions*, vol. 8, no. 5, pp. 30–38, 2001.
- [14] Z. Fan, "Phase analysis of the willingness to adopt digital natives' social media based on user portraits," *Modern Information*, vol. 37, no. 6, pp. 99–106, 2017.
- [15] A. Lu and Y. Zhou, "Portraits and comparison of microblog information and comment users in the context of terrorist incidents," *Information Science*, vol. 38, no. 4, pp. 9–16, 2020.
- [16] S. Zhao, "Social media user portraits with high conversion rate: a research based on in-depth interviews with 500 users," *Modern Communication (Journal of Communication University of China)*, vol. 36, no. 6, pp. 115–120, 2014.
- [17] W. Mingzhu, H. Zhang, and Y. Liu, "Research on the portrait of high-influential people in social media based on multidimensional attribute fusion," *Library, Information, and Knowledge*, vol. 5, no. 5, pp. 73–79, 2019.
- [18] G. Gao, "A survey of user profiles methods," *Data Analysis and Knowledge Discovery*, vol. 3, no. 3, pp. 25–35, 2019.
- [19] Q. Yunfei and Z. Weizhu, "Study for the construction method of group profile based on network structure and text content," *Library and Information Service*, vol. 63, no. 22, p. 21, 2019.
- [20] Y. Yang, "Research and realization of internet public opinion analysis based on improved TF-IDF algorithm," in *Proceedings of the 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*, pp. 80–83, Anyang, China, October 2017.