Analysis of the Impact of Literary Works on College Students' Mental Health Based on Intelligent Data

Xin Li[1*]

[1]Chongqing College of Finance Economics, Chongqing, 402160, China

Corresponding author: Xin Li, xinli098@aol.com

Abstract. In order to explore the impact of literary works on college students' mental health, this paper explores the impact of literary works on college students' mental health through intelligent data analysis methods. Moreover, this paper improves the algorithm and builds the corresponding model, and the effect of the model is verified by experiments. In addition, this paper analyzes the whole process of building the mental state perception model in detail, and provides a model building method for model experiments. Finally, this paper studies several psychological parameters of the control group and the experimental group, including self-esteem, self-confidence, communication ability, and anti-stress ability through experimental research. From the experimental research, it can be seen that there is a certain gap between two groups of objects in terms of self-esteem, self-confidence, communication ability, and anti-stress ability. The mental health improvement effect of the experimental group is significantly better than that of the control group. This verifies that literary works have certain advantages in improving the mental health of college students.

Keywords: intelligent data analysis; literary works; mental health.

1   INTRODUCTION

According to the psychological development topics of individuals at different ages, it helps them solve psychological conflicts, understand themselves and society, and promote personality perfection and individual development. Literary reading provides college students with a relaxed and pleasant learning environment and relieves tension and high-pressure emotions. Moreover, students discover beauty and self-confidence through the interactive participation and experience of literary reading. It can be said that the literary reading activity model is a specific model of developmental psychological counseling for college students. This paper combines intelligent data to analyze the impact of literary works on college students' mental health, and conducts data research from the actual situation to promote the subsequent improvement of college students' mental health.

2   RELATED WORKS

Under the severe social pressure, many college students cannot bear such pressure, and poor mental health even leads to serious consequences [1]. In the past, mental health education was only through the transmission of psychological knowledge, so there was no emphasis on humanistic education. Through the use of literary works to infiltrate healthy psychology can be subtly influenced [2].

Build a multi-disciplinary blending and experiential teaching system to achieve environmental education. At present, the mental health education curriculum is in a situation of being alone and weak and a complete and interconnected curriculum system has not yet been formed [3]. Scientifically formulate teaching plans and content, and attach importance to the supplement and support of other disciplines for mental health education. Deeply promote the incorporation of experiential teaching modes such as literary appreciation, literary criticism, and literary creation, integrate multi-sensory and experiential education into interactive classrooms, and realize the ingenious connection of literary education and mental health education, as well as the organic combination of other education and self-education. From the perspective of subject penetration, excellent literary works contain psychological education resources, and the teaching process should highlight the important position of human emotions, attitudes and values in the teaching process, and emphasize the nurturing infection, subtle influence and unique inner feelings of students [4]. Teachers should accept students' different viewpoints with an open and inclusive attitude, design teaching aesthetically according to the logical structure and goals of teaching content, and maintain an organic combination with students' cognition and psychological experience.

The characteristics of civilianization and openness are beneficial to achieve rapid and effective literary therapy [5]. The online platform provides realistic conditions for college students' aesthetic judgment and emotional appeal. The "zero threshold" of literary criticism and literary creation attracts a large number of students to express their free thoughts on the Internet and open up a space for discourse [6]. In the environment of freedom and inclusiveness of the Internet, college students' literary creation no longer deliberately pursues the aesthetic height and ideological depth of literature [7].

The "experience" in literature education is the fundamental way of mental health education [8]. Only by experiencing life with your heart can you realize the meaning of life and the value of life. Therefore, according to the theme of literary appreciation, the teaching process is divided into five stages: background elaboration, situation highlighting, independent exploration, communication and interaction, and summarization and sharing [9]. Teachers introduce the writing background of literary works and the life of writers, set the problem situation of the course and assign learning tasks [10]. Under the guidance of teachers, students read and appreciate independently, and exchange feelings among students, experience the collision and thinking of different ideas. Finally, teachers share opinions and understandings in response to various problems and confusions faced by students, guide them to reflect on life, experience life process and value, infiltrate the perception of literary works into the content of mental health education, and raise literary reading from perceptual to rational Thinking and Education Session.

3   METHODOLOGY

This paper explores the impact of literary works on college students' mental health through intelligent data analysis methods, improves the algorithm and builds a corresponding model, and verifies the effect of the model through experiments.

4   RELATED CONCEPTS AND TECHNOLOGIES OF DATA MINING

For big data containing massive disordered and incomplete data, it contains a lot of valuable information and knowledge that people did not recognize beforehand. Data Mining is the process of mining value in these unknown data mines by playing the role of a miner (Figure 1).

Data conversion processing can also be called data mapping processing, which generally has three cases. One is the   encoding conversion of text data, because the computer cannot directly process text data, such as calculating the distance between two data. Therefore, the text needs to be coded numerically, and the exponentially changing data can be quickly converted into decimal value data which is convenient for observation and analysis by formula 1.

$$y = \log_m (x+k)$$
(4.1)

Similarly, when the data changes in the form of a power function, it can be processed by the method of opening the nth power, and it can be converted into small numerical data that is easier to observe by using formula 2. y is the value after conversion, and x is the value before conversion.
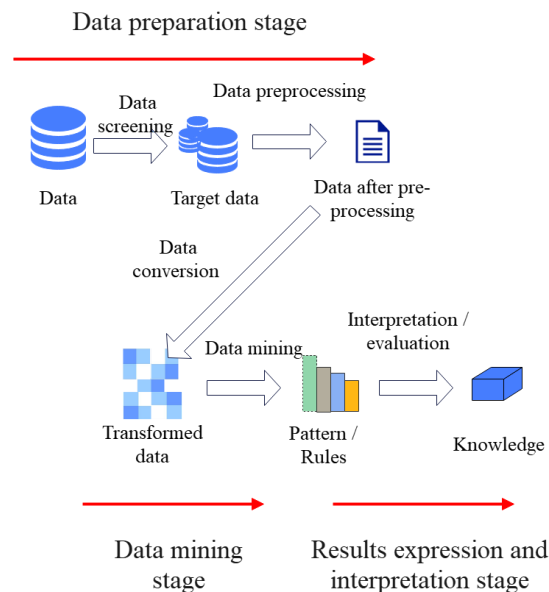
$$y =^n \sqrt{x}$$
(4.2)

Figure 1: Knowledge discovery process.

Pearson's correlation coefficient is a method often used in statistics to measure whether there is a correlation between variables and the magnitude of the degree. Its output range is [-1, 1]. When the negative correlation is present, the output is less than zero, and when the positive correlation is present, the output is greater than zero.

There are two key parameters in the Pearson correlation coefficient analysis. One is the $\rho_{X,Y}$ value for testing significance. When $\rho_{X,Y} < 0.05$, there is a correlation between the two. The second parameter is the correlation degree coefficient r value, which is a coefficient reflecting the degree of correlation. The formulas of these two parameters are as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left(\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right)}{\sigma_X \sigma_Y} \tag{4.3}$$

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)}} \tag{4.4}$$

5   LOGISTIC REGRESSION ANALYSIS

Logistic regression analysis is actually an extension of linear regression analysis, also known as log probability regression, X is said to obey the logistic regression distribution.

$$F(X) = P(X \le x) = \frac{1}{1 + e^{-(x - \mu/\gamma)}} \tag{5.1}$$

$$f(X) = F'(X) = \frac{e^{-(x - \mu/\gamma)}}{\gamma\left(1 + e^{-(x - \mu/\gamma)}\right)^2} \tag{5.2}$$

The distribution function and density function curves are shown in Figure 2.
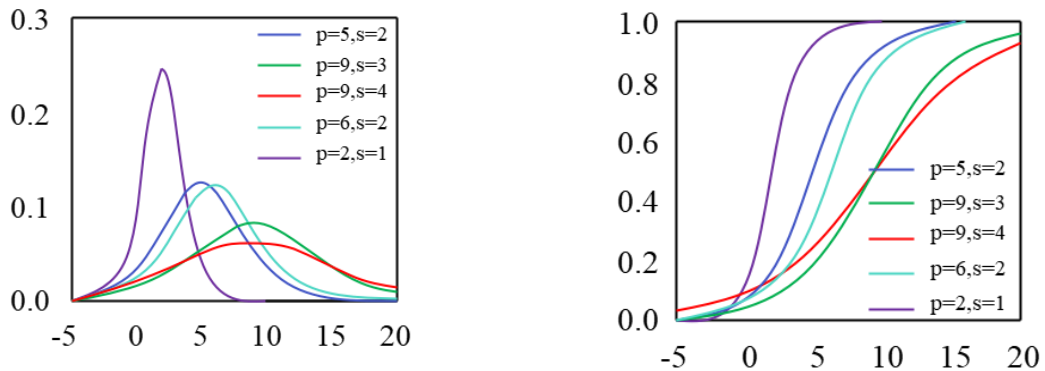

Figure 2: Distribution function and density function curves (a) Distribution function curves, (b) Density function curves.

Binomial logistic regression is suitable for dealing with the situation when the value of the dependent variable Y has only two categories, which is essentially a prediction model of classification probability. The value of variable X is a real number, and the probability model when

the value of the dependent variable Y is 1 is shown in formula 7, and when it is 0, it is shown in formula 8.

$$P(Y=1\,|\,x)=\frac{\exp(\omega\cdot x+b)}{1+\exp(\omega\cdot x+b)}$$
(5.3)

$$P(Y=0\,|\,x)=\frac{1}{1+\exp(\omega\cdot x+b)}$$
(5.4)

In these two formulas, $x\in R^n$ is the input value of the model, $Y\in 0,1$ is the output value of the model, $\omega\in R^n$ is the parameter, and the logistic regression analysis model can calculate the $P_n(Y=0\,|\,x_n)$ and $P_n(Y=1\,|\,x_n)$ values for each dimension x of the input vector, respectively.

The concepts of event probability and logarithm are introduced here. The probability of an event can be expressed by the ratio between the two, namely P/(1-P). On this basis, the probability of the event can be obtained by taking the logarithm of its logarithmic probability. The formula is shown in 9. Similarly, for the binary logistic regression model, the corresponding log probability can be given by formula 10.

$$\log it(P)=\log\frac{P}{1-P}$$
(5.5)

$$\log it(P)=\log\frac{P(Y=1\,|\,x)}{1-P(Y=1\,|\,x)}=\omega\cdot x+b$$
(5.6)

Model parameter estimation is actually the process of finding the optimal solution by the logistic regression model in order to fit the linear function shown in Equation 10. In this process, the method of maximizing the likelihood function can be used to obtain the optimal parameters. The likelihood function of logistic regression can be expressed as:

$$\prod_{i=1}^{N}\left[\pi(x_i)^{y_i}\right]\left[1-\pi(x_i)\right]^{1-y_i}$$
(5.7)

By taking the logarithm, the log-likelihood function is obtained as:

$$L(\omega)=\sum_{i=1}^{N}\left[y_i(\omega\cdot x_i)-\log(1+\exp(\omega\cdot x_i))\right]$$
(5.8)

It can be seen from the above two formulas that after solving the optimal value of $L(\omega)$ with $L(\omega)$ as the objective function, the obtained solution is the maximum value of $L(\omega)$, that is, the optimal value of the parameter ω, thus completing the model parameter estimation. We assume that the optimal value obtained by the parameter o is $\tilde{\omega}$, which can be transformed into the following form after substituting it into the logistic regression model:

$$P(Y=1\,|\,x)=\frac{\exp(\tilde{\omega}\cdot x)}{1+\exp(\tilde{\omega}\cdot x)}$$
(5.9)

$$P(Y=0\,|\,x) = \frac{1}{1+\exp(\tilde{\omega}\cdot x)}$$

(5.10)

6   DATA MINING METHODS

The process of evaluating the results of the classification algorithm is inseparable from the existence of a confusion matrix, which is a computing tool often used in classification algorithms. Figure 3 shows the composition of the confusion matrix:

Forecast classification

		+	-	Total
The actual classification	+	TP (True Positives)	FN (False Negatives) Type II error	TP + FN (Actual Positives)
	-	FN (False Positives) Type I error	TN (True Negatives)	FP +TN (Actual Negatives)
	Total	TP + FP (Predicted Positives)	FN + TN (Predicted Negatives)	TP + FP + FN + TN

Figure 3: Confusion matrix.

There are four parameters in the confusion matrix, which are:

Accuracy: This is also called the overall recognition rate, and it measures the overall ability of the classification model to correctly identify various datasets.

$$accuracy = (TP+TN)/(P+N)$$

(6.1)

Precision: It reflects the proportion of correct classification in the classification results of the classification model for each category, that is, the precision of the model when judging each category.

$$precision = TP/(TP+FP)$$

(6.2)

$$recall = TP/(TP+FP) = sensitivity$$

(6.3)

$$F_1 = \frac{2\times precision\times recall}{precision+recall}$$

(6.4)

7   THE CONSTRUCTION OF MENTAL STATE PERCEPTION MODEL

When the sample data is linearly separable as in Figure 4(a), it can be seen that there can be multiple straight lines or hyperplanes between positive and negative classes to distinguish each

other. However, the purpose of the SVM is to find a line or hyperplane that makes the two classes as distinguishable as possible, as shown in Figure 4(b). To do this, SVM uses a margin maximization method to find this line or hyperplane.
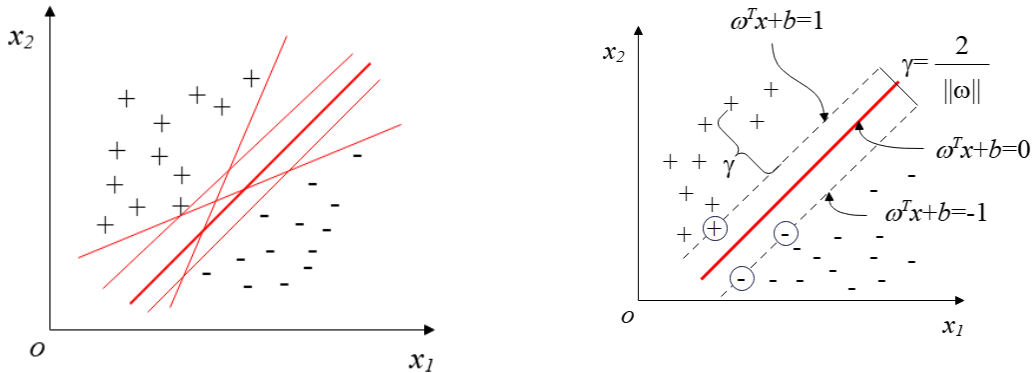

Figure 4: (a) Linearly separable dataset (b) Maximum margin hyperplane.

We set the data set of m samples as $D = \left\{ (X_1, y_1), (X_2, y_2), \cdots, (X_m, y_m) \right\}$, y_m represents the classification identification of the m-th data sample, the feature space of the sample data set contains n feature dimensions, the m-th sample data is $X_m = \left\{ x_m^1, x_m^2, \cdots x_m^n \right\}$.

A linearly separable hyperplane H can be expressed as:

$$H : \omega^T \cdot X + b = 0 \qquad (6.5)$$

Among them, $\omega^T = \left\{ \omega_1, \omega_2, \cdots \omega_n \right\}$ is the weight vector, the classification boundary is:

$$H_1 : \omega^T X + b \geq 1, y = \pm 1 \qquad (6.6)$$

$$H_{-1} : \omega^T X + b \leq -1, y = -1 \qquad (6.7)$$

The decision-making basis of the C4.5 algorithm is based on the information gain rate after each division of the data set. Each time the variable is divided, the direction with the largest gain rate feedback will be selected to execute, and finally the entire decision tree will be obtained. The genetic algorithm in this section is a fitness function based on the distance criterion, which directly relies on the data of the sample itself for calculation, which is intuitive and concise, and the physical concept is clear. The separability of samples is judged by calculating the distance between similar samples and the distance between heterogeneous samples. The corresponding calculation content and formula are as follows:

The overall within-class scatter matrix is:

$$S_w = \sum_{i=1}^{c} P(\omega_i) E \left\{ (X - M_i)(X - M_i^T) \right\}$$

$$= \sum_{i=1}^{c} P(\omega_i) \frac{1}{N_i} \sum_{k=1}^{N_i} (X_k^i - M_i)(X_k^i - M_i)^T \qquad (6.8)$$

The overall inter-class scatter matrix is:

$$S_b = \sum_{i=1}^{c} P(\omega_i) E\left\{ (M_i - M_0)(M_i - M_0)^T \right\}$$

(6.9)

The fitness function is:

$$J = \frac{tr(s_b)}{tr(s_w)}$$

(6.10)

In the formula, C is the number of categories, $M_i = \frac{1}{N_i} \sum_{k=1}^{N_i} X_k^i$ is the mean vector of the i-th category. J is the distance-based criterion fitness function. The smaller the distance between similar samples and the larger the distance between heterogeneous samples, the better the classification effect of the model.

The selection strategy of genetic algorithm is the selection strategy of roulette. This method is a random sampling method with replacement. According to the ratio of the calculated value of each individual fitness function to the sum of the group fitness function value, the probability that it can be selected into the next generation is taken. Therefore, individuals with higher fitness function calculation values are more likely to be retained. The calculation formula is as follows:

$$P_i = \frac{\sum_{1}^{i} J_i}{\sum_{i=1}^{N} J_i}$$

(6.11)

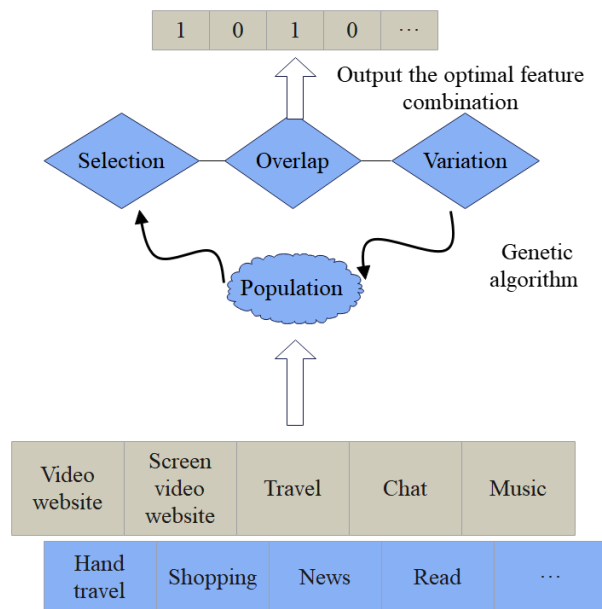The whole process of training the genetic algorithm is shown in Figure 5:

Figure 5: Genetic algorithm training diagram.

8 ANALYSIS MODEL OF THE IMPACT OF LITERARY WORKS ON COLLEGE STUDENTS' MENTAL HEALTH BASED ON INTELLIGENT DATA

This paper takes the mental health of college students as the research subject, and analyzes the influence of psychological factors on their mental health and the relationship between the factors of literary works. The research idea is shown in Figure 6.
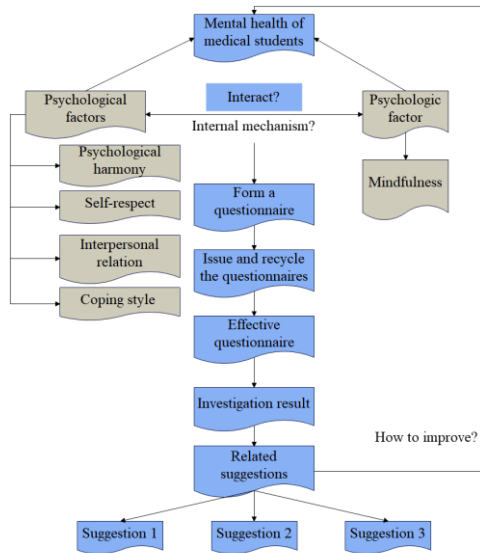

Figure 6: Research ideas.

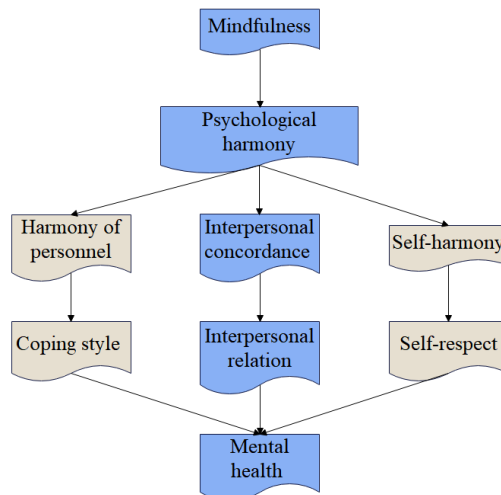The following analysis model is established (Figure 7):

Figure 7: Schematic diagram of the analysis model.

In the early stage, through the mental health test, we can see that the initial mental health level of the two groups is basically the same. For the next two months, the experimental group reads literature, while the control group follows a normal lifestyle.

9   RESULTS

In this paper, several psychological parameters such as self-esteem, self-confidence, communication ability, and stress resistance ability of the control group and the experimental group are studied through experimental research. The experimental results obtained are shown in Figures 8-11.
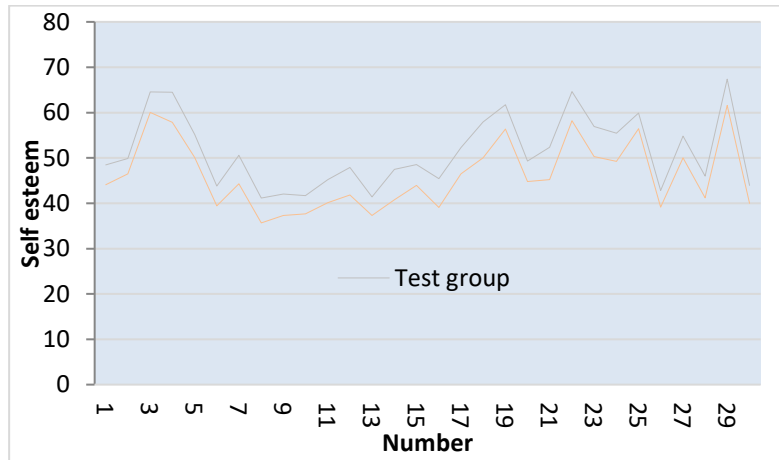

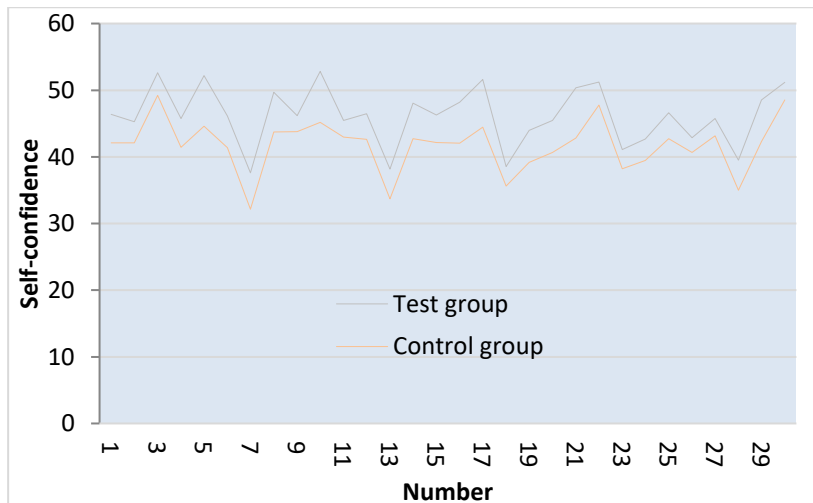**Figure 8**: Statistical diagram of self-esteem of two groups of objects.

Figure 9: Statistical diagram of the self-confidence of two groups of objects.
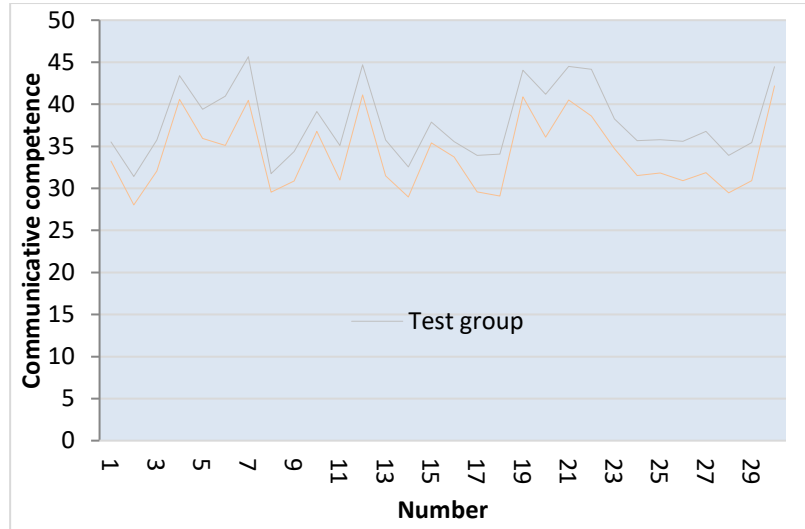
Figure 10: Statistical diagram of the communicative competence of two groups of objects.
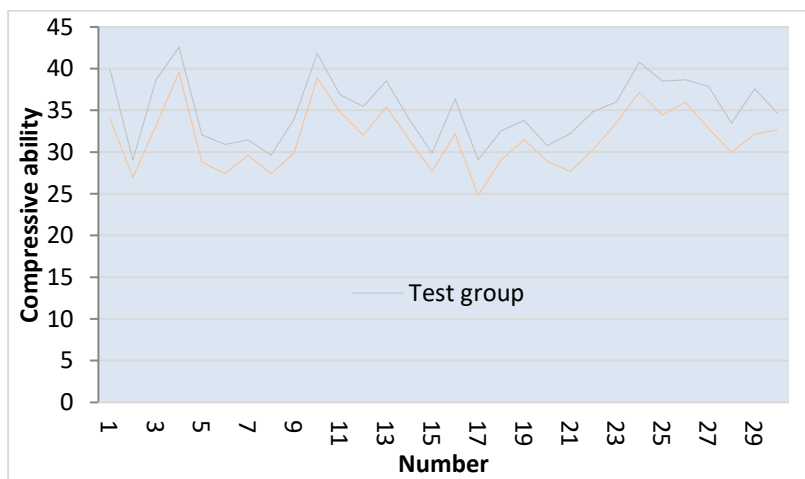

Figure 11: Statistical diagram of the compression resistance of two groups of objects.

10 DISCUSSIONS

At present, college freshmen have various psychological problems, their mental health status is lower than the national average level, the incidence of psychological disorders is high, and their mental health status is not optimistic. There are various factors affecting the mental health of college students, including both internal, and external and objective factors in the process of individual development. It is necessary to implement targeted health education among college students and carry out health education activities to promote the improvement of current college students' psychological quality and comprehensive quality. This paper proposes to explore the impact of literary works on college students' mental health through intelligent data analysis. From

Figure 8 to Figure 11, it can be seen that there is a certain gap between two groups of objects in terms of self-esteem, self-confidence, communication ability, and ability to resist pressure.

11   CONCLUSIONS

Through comparative experimental research, this paper understands and analyzes the mental health status and influencing factors of college freshmen, and finds out the main problems. Moreover, this paper improves the mental health of college students through literary works, and provides a basis for future health education and health promotion. In addition, this paper uses intelligent data analysis on the impact of literary works on college students' mental health, and conducts data research from the actual situation to promote the subsequent improvement of college students' mental health. From the experimental research, it can be seen that there is a certain gap between two groups of objects in terms of self-esteem, self-confidence, communication ability and stress resistance. This verifies that literary works have certain advantages in improving the mental health of college students.

12   RECOMMENDATIONS

When the traditional educational resources and forms can no longer meet the actual educational needs, it is necessary to conduct a new review and thinking about it, and actively seek and use new content and methods. It is an effective way to organically combine literary works and college students' mental health education, and it has also achieved good results in reality. Through the combination of the two, it provides important help for the perfect formation of students' outlook on life, world outlook and values.

13   ACKNOWLEGDEMENT

When the traditional educational resources and forms can no longer meet the actual educational needs, it is necessary to conduct a new review and thinking about it, and actively seek and use new content and methods. It is an effective way to organically combine literary works and college students' mental health education, and it has also achieved good results in reality. Through the combination of the two, it provides important help for the perfect formation of students' outlook on life, world outlook and values.

Xin Li, https://orcid.org/0000-0002-0628-1304

REFERENCES

[1]   Gonçalves-Pinho, M.; Ribeiro, J. P.; Freitas, A., Mota, P.: The use of big data in psychiatry– the role of pharmacy registries, European Psychiatry, 64(S1), 2021 S793-S793. https://doi.org/10.1192/j.eurpsy.2021.2096
[2]   Hong, A.; Kim, B.; Widener, M.: Noise and the city: Leveraging crowdsourced big data to examine the spatio-temporal relationship between urban development and noise annoyance, Environment and Planning B: Urban Analytics and City Science, 47(7), 2020, 1201-1218. https://doi.org/10.1177/2399808318821112
[3]   Liu, J.; Zhai, X.; Liao, X.: Bibliometric analysis on cardiovascular disease treated by traditional Chinese medicines based on big data, International Journal of Parallel, Emergent and Distributed Systems, 35(3), 2020, 323-339. https://doi.org/10.1080/17445760.2019.1606912
[4]   Miller, J.; Atala, R.; Sarangarm, D.; Tohen, M.; Sharma, S.; Bhatt, S.; Cruz, M.: Methamphetamine abuse trends in psychiatric emergency services: a retrospective analysis

using big data, Community mental health journal, 56(5), 2020, 959-962. https://doi.org/10.1007/s10597-020-00563-1

[5]   Moessner, M.; Feldhege, J.; Wolf, M.; Bauer, S.: Analyzing big data in social media: Text and network analyses of an eating disorder forum, International Journal of Eating Disorders, 51(7), 2018, 656-667. https://doi.org/10.1002/eat.22878

[6]   Price, W. N.; Cohen, I. G.: Privacy in the age of medical big data. Nature medicine, 25(1),2019, 37-43. https://doi.org/10.1038/s41591-018-0272-7

[7]   Rudorfer, M. V.: Psychopharmacology in the age of "big data": the promises and limitations of electronic prescription records. CNS drugs, 31(5), 2017, 417-419. https://doi.org/10.1007/s40263-017-0419-y

[8]   Shatte, A. B.; Hutchinson, D. M.; Teague, S. J.: Machine learning in mental health: a scoping review of methods and applications. Psychological medicine, 49(9), 2019, 1426-1448. https://doi.org/10.1017/S0033291719000151

[9]   Wang, Y.; Kung, L.; Wang, W. Y. C.; Cegielski, C. G.: An integrated big data analytics-enabled transformation model: Application to health care, Information & Management, 55(1), 2018, 64-79. https://doi.org/10.1016/j.im.2017.04.001

[10] Wilfling, D.; Hinz, A.; Steinhäuser, J.: Big data analysis techniques to address polypharmacy in patients–a scoping review, BMC family practice, 21(1), 2020, 1-7. https://doi.org/10.1186/s12875-020-01247-1