Eng Int Syst (2026) 5: 703–708
© 2026 CRL Publishing Ltd

Engineering
Intelligent Systems

Using the Decision Tree Algorithmto Explore the Factors Influencing Social Anxiety Among University Students

Na Yao*

Xi'an Vocational and Technical College, Xi'an, Shaanxi 710077, China

This study is conducted to analyze the factors influencing social anxiety among university students using the decision tree (DT) algorithm to provide some bases for its intervention. Based on convenient sampling, questionnaires were distributed to students of a university, and 1,832 valid responses were obtained. The questionnaire included items eliciting demographic information, and the interaction anxiousness scale (IAS). Logistic regression and chi-squared automatic interaction detector (CHAID) were used to analyze the questionnaire data. Data analysis revealed that 28.28% of the students experienced social anxiety. Both the Logistic and DT models showed that family structure completeness, family financial status, leadership qualities, and whether one is an only child were factors influencing the social anxiety of students. The comparison found that the area of the curve under the receiver operator characteristic of the DT model was larger than that of the Logistic model ($Z = 9.264,\ p < 0.01$), indicating that the decision tree algorithm performs well in determining the factors influencing social anxiety. More attention should be paid to students with incomplete family structures and those whose families are not financially secure, so as to help them establish confident and acceptable social behaviors.

Keywords: Decision tree, social anxiety, influencing factor, family

1. INTRODUCTION

University is an important period in terms of an individual's studies and life in general but, at the same time, there is a high incidence of psychological problems among students (Roche et al., 2022), such as depression and anxiety (Downing et al., 2020). Social anxiety is one of them (Zou, 2024). Social anxiety refers to the situation where an individual experiences negative emotions such as embarrassment and/or fear in social situations and therefore avoids socializing (Li et al., 2023). Long-term social anxiety is not conducive to an individual's adaptation to society and can also affect their growth and development (Abinaya & Vadivu, 2024). Psychologists have conducted extensive research on social anxiety (Zabag et al., 2024). Wu et al. (2024) explored the connection between rumination and social anxiety by conducting a six-month follow-up study on more than 300 college students from three universities in Henan Province. They found that rumination and social anxiety among college students were mutually predictable, and any intervention could disrupt the vicious cycle between them. Sousa et al. (2024), through a one-year group study, analyzed the relationship between bullying, victimization, social anxiety, and social withdrawal. They discovered a reciprocal relationship between social anxiety and withdrawal, highlighting the significance of preventive and remedial measures in alleviating social anxiety among adolescents who have experienced bullying. Zhang et al. (Zhang et al., 2024) investigated the prevalence of social

*Corresponding address: No. 251, Yudou Road, Yanta District, Xi'an City, Shaanxi 710077, China. Email: yaonayn@hotmail.com

Table 1 Demographic variables statistics of the study subjects.

		Number	Proportion
Gender	Male	987	53.88%
	Female	845	46.12%
Grade	Freshman	476	25.98%
	Sophomore	458	25.00%
	Junior year	463	25.27%
	Senior year	435	23.74%
Major	Science and engineering	684	37.34%
	Humanities and social sciences	677	36.95%
	Arts and sports	471	25.71%
Place of origin	Town	956	52.18%
	Country	876	47.82%
Whether an only child	Yes	706	38.54%
	No	1,126	61.46%
Whether a student leader	Yes	846	46.18%
	No	986	53.82%
Having a family structure	Yes	1,642	89.63%
	No	190	10.37%
Family financial status	Good	774	42.25%
	Medium	652	35.59%
	Poor	406	22.16%

anxiety among abandoned children and adolescents in China, and found that those with greater levels of online parent-child communication and online social capital were more likely to have high social anxiety. Kaplan et al. (2023) analyzed the connection among body mass index (BMI), social anxiety, and implicit weight bias in female college students, and discovered that an increased BMI was linked to greater levels of social anxiety, and treatment interventions aimed at addressing social anxiety in individuals with high BMIs should consider investigating internalized weight stigma. This paper mainly used the decision tree (DT) algorithm to analyze the influencing factors of social anxiety among university students, in order to better understand the causes of social anxiety among university students and provide some support for the intervention of social anxiety.

2. MATERIALS AND METHODS

2.1 Research Subjects

Convenience sampling (Sharma et al., 2022) was employed to distribute questionnaires to 1,884 university students. After eliminating 52 invalid questionnaires, 1,832 valid questionnaires were obtained, with an effective rate of 97%. Table 1 shows the demographic variables of the study subjects.

2.2 Research Methods

(1) Questionnaire survey

.

The interaction anxiety scale (IAS) developed by Leary is widely used in the measurement of social anxiety (Wu et al., 2024). It consists of 15 questions, and a five-point scoring system is adopted:

1: No match at all,

2: Slight match,

3: Moderate match,

4: Strong match,

5: Total match.

In this study, the Cronbach's α of the scale (Novak & Rebernjak, 2023) was 0.85.

(2) Statistical analysis

Data were analyzed using SPSS 25.0 software (Perdana et al., 2021). Taking social anxiety as the dependent variable, variables with $p < 0.05$ were included in logistic regression (Tu, 2018) and DT for analysis. The DT algorithm used the chi-squared automatic interaction detector (CHAID) (Strzelecka & Zawadzka, 2023). The maximum tree depth was set to 3, and the minimum sample sizes of parent nodes and child nodes were 100/50. Receiver operator characteristic curve (ROC) (Habibzadeh, 2024) was used to analyze the predictive effects of the two models.

3. RESULTS AND ANALYSIS

According to Table 2, the detection rate of social anxiety among students in this university was 28.26%, and the average score was 44.56 points.

According to Table 3, the social anxiety score of females was 3.05 ± 0.71 points, significantly higher than that of males ($t = -4.859$, $p = 0.001$). The social anxiety score of students who are not the only child was 2.95 ± 0.68 points, significantly higher than that of those who were the only

Table 2 Descriptive statistics for social anxiety.

	Total score	Maximum	Minimum	Detection rate /%
Social anxiety	44.56±8.97	75	15	28.28

Table 3 Differential tests for social anxiety.

		Social anxiety score	t/F	p
Gender	Male	2.78 ± 0.71	-4.859	0.001
	Female	3.05 ± 0.71		
Grade	Freshman	2.81 ± 0.67	1.112	0.345
	Sophomore	2.82 ± 0.66		
	Junior year	2.87 ± 0.71		
	Senior year	2.89 ± 0.72		
Major	Science and engineering	2.96 ± 0.66	1.268	0.217
	Humanities and social sciences	2.94 ± 0.68		
	Arts and sports	2.92 ± 0.67		
Place of origin	Town	2.85 ± 0.74	0.711	0.474
	Country	2.88 ± 0.75		
Whether an only child	Yes	2.82 ± 0.76	2.212	0.033
	No	2.95 ± 0.68		
Whether a student leader	Yes	2.86 ± 0.66	-5.021	0.001
	No	3.05 ± 0.62		
Having a complete family structure	Yes	2.84 ± 0.68	8.772	0.001
	No	3.21 ± 0.84		
Family financial status	Good	2.83 ± 0.77	3.845	0.001
	Medium	2.92 ± 0.73		
	Poor	3.04 ± 0.68		

Table 4 Variable assignment.

	Assignment
Gender	Male =0, female =1
Whether the only child	Yes =0, no =1
Whether a student leader	Yes =0, no =1
Having a complete family structure	Yes =0, no =1
Family financial status	Good =0, medium =1, poor =2

children ($t = 2.212$, $p = 0.033$). The social anxiety score of students who were not student leaders was 3.05 ± 0.62, significantly higher than that of students who were student leaders ($t = -5.021$, $p = 0.001$). The score of students with incomplete family structures was 3.21 ± 0.84, significantly higher than that of students with a complete family structures ($t = 8.772$, $p = 0.001$). The social anxiety score of students from poor families was 3.04 ± 0.68, significantly higher than that of students from families with good or medium financial conditions ($F = 3.845$, $p = 0.001$).

The social anxiety of these students was assigned as the dependent variable, with no anxiety = 0 and anxiety = 1. Variables with $p < 0.05$ in Table 3 were taken as independent variables. The assignment is shown in Table 4.

According to Table 5, female students, non-only children, non-student leaders, incomplete family structure, and poor family economic conditions were the main influencing factors of social anxiety among students. Females were more likely to experience social anxiety. Non-only children were more likely to experience social anxiety than the only children. Non-student leaders had a higher risk to experience social anxiety than student leaders. Students with incomplete family structures are more likely to experience social anxiety than those with complete family structures. Students from poor families tended to have a higher likelihood of experiencing social anxiety than those from families with medium or good economic conditions.

The A DT model was established based on the influencing factors, as shown in Figure 1. The explanatory variables were: having a complete family structure, family financial status, the student's leadership qualities, and only-child status. Among them, the most important root node variable was whether a student had a complete family structure, indicated by the detection of 81.58% anxiety among those students who had incomplete families. Also, the social anxiety of students with complete family structures was affected by their families' financial conditions, and the detection rate of social anxiety among students from poorer families was 75.39%. The social anxiety of students with incomplete family structures was affected by whether they

Table 5 Results of logistic regression analysis.

Variable	B	SE	Wald	P	OR	95%CI
Constant	-10.874	2.364	12.214	0.000	0.000	
Female	1.525	0.641	5.487	0.015	4.526	1.321 ~ 16.521
Whether the only child (no)	0.098	0.043	5.251	0.021	1.112	1.015 ~ 1.197
Whether a student leader (no)	0.253	0.102	5.214	0.019	1.214	1.035 ~ 1.526
Whether have a complete family structure (no)	2.352	0.845	7.415	0.006	10.521	1.521 ~ 53.251
Family economic condition (medium)	0.845	0.714	1.058	0.286	2.264	0.415 ~ 10.625
Family economic condition (poor)	1.855	0.776	5.658	0.016	6.325	1.358 ~ 28.864

Note: SE: standard error, OR: odds ratio, CI: confidence interval

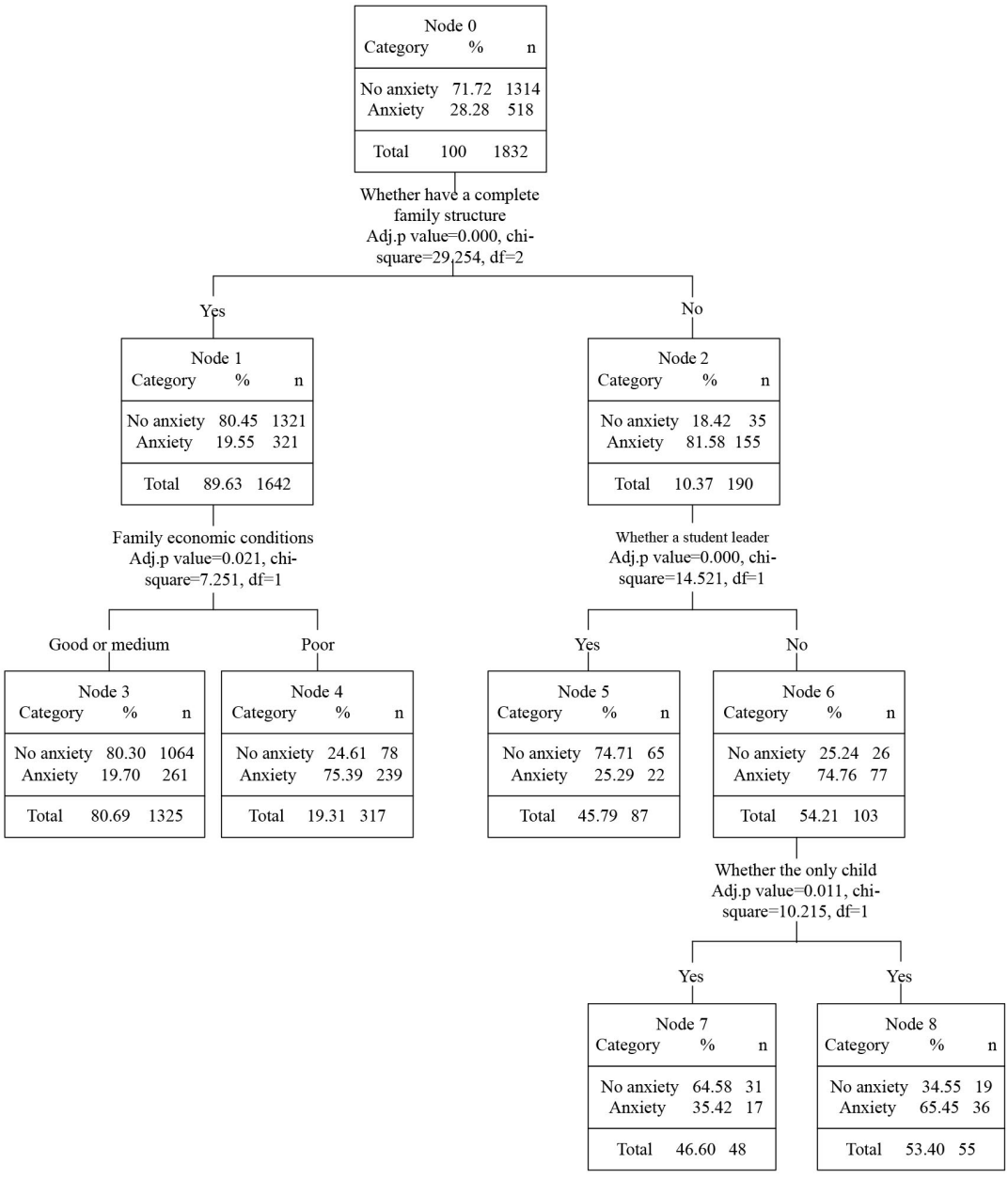


Figure 1 Decision tree model of factors influencing social anxiety among university students.

Table 6 Comparison of classification effects.

	Area under the ROC curve	SE	Asymptotic significance	Asymptotic 95%CI
Logistic	0.896	0.012	< 0.001	0.872 - 0.915
DT	0.967	0.003	< 0.001	0.965 - 0.984

were student leaders, and the detection rate of social anxiety among students who were not student leaders was 74.76%. The detection rate of social anxiety among students who were not student leaders was also related to whether they are the only child, and social anxiety was detected in 65.45% of non-only children.

As shown in Table 6, the DT had a better classification effect in regard to social anxiety, with an area under the ROC curve of 0.967, while the logistic model had an area under the ROC curve of 0.896. The difference between the two models was statistically significant ($Z = 9.264$, $p < 0.01$).

4. DISCUSSION

The physical and mental health of university students is an issue of great concern in society (Labell & Johnson, 2021). Social anxiety is a common psychological problem among university students. It not only affects students' academic performance (Polat et al., 2022) but is also detrimental to their physical and mental development. According to the survey results, the detection rate of social anxiety in the studied university was 28.28%, indicating that this is a significant issue among students.

The analysis conducted using the Logistic and DT models showed that incomplete family structures may cause students to lack sufficient care and attention when facing various pressures, have fewer opportunities to interact with family members, and lack experience in social situations. In addition, incomplete family structures may lead to some biases and discrimination, further exacerbating social anxiety. Students from poor families are restricted by their financial conditions and have a heavier psychological burden when participating in various social activities and attending various social occasions, which makes them more likely to feel anxious in social situations. Students who hold student leadership positions are often involved in more activities and meetings, and therefore have more opportunity to develop their social skills.. Students who do not hold student leadership positions are lacking in this regard and are more likely to be at a loss in the face of unexpected situations, resulting in greater social pressure. Compared with only children, children with siblings may face more conflicts and disputes in family life and may be more likely to be overlooked, resulting in a lack of security in social situations and more severe social anxiety.

In regard to model comparison, the DT model performed better. DT can effectively indicate the interaction between different variables and visually show the importance of different variables in the form of a tree diagram (Zhang, 2024). Moreover, it also shows good results in the classification of social anxiety. However, this study also has several limitations. For instance, the use of the convenient sampling method may limit the generalizability of the results. In terms of the factors influencing social anxiety, only some relatively accessible data were considered. Therefore, future research should be conducted on largersamples to verify the applicability of the results and explore more factors that affect social anxiety to further improve the DT model.

5. CONCLUSION

This study explored the factors influencing social anxiety among university students through questionnaire surveys and the DT algorithm. Logistic regression analysis revealed that female students, non-only children, non-student leaders, incomplete family structures, and poor family financial status were the main factors influencing social anxiety. The DT model showed that a complete family structure was the most important factor influencing social anxiety. In comparison, the classification ability of the DT modelwas superior to that of the logistic regression model. The DT model can be applied in actual mental health programs in universities to provide support for students' mental health.

REFERENCES

1. Abinaya, M., &Vadivu, G. (2024). Identification of Social Anxiety in High School: A Machine Learning Approaches to Real-Time Analysis of Student Characteristics.*IEEE Access*, *12*, 77932–77946.
2. Downing, V. R., Cooper, K. M., Cala, J. M., Gin, L. E., & Brownell, S. (2020). Fear of Negative Evaluation and Student Anxiety in Community College Active-Learning Science Courses.*CBE Life Sciences Education*, *19*(2), 1–16.
3. Habibzadeh, F. (2024). On the use of receiver operating characteristic curve analysis to determine the most appropriate p value significance threshold.*Journal of Translational Medicine*, *22*(1), 12.
4. Kaplan, S. C., Butler, R. M., &Heimberg, R. G. (2023). The Relationship Between Body Mass Index, Implicit Weight Bias, and Social Anxiety in Undergraduate Women.*Cognitive Therapy & Research*, *47*(5), 761–771.
5. Labelle, S., &Johnson, Z. D. (2021). The relationship of student-to-student confirmation in the classroom to college students' mental health and well-being.*Communication Quarterly*, *2021*(1), 1–19.
6. Li, Q., Zhang, X., Yang, X., Pan, N., Li, X., Kemp, G. J., Wang, S., & Gong, Q. (2023). Pre-COVID brain network topology prospectively predicts social anxiety alterations during the COVID-19 pandemic.*Neurobiology of Stress*, *27*, 11.
7. Novak, J., &Rebernjak, B. (2023). There are Many Greater Lower Bounds than Cronbach's α: A Monte Carlo Simulation Study.*Measurement: Interdisciplinary Research and Perspectives*, *21*, 1 - 28.

8. Perdana, A. C., Rachman, R., &Palinggi, M. D. M. (2021). An analysis of population using Multiple Linear Regression Analysis.*IOP Conference Series: Materials Science and Engineering*, *1088*(1), 1–10.
9. Polat, E., Hopcan, S., &Arslanta, T. K. (2022). The Association between Flipped Learning Readiness, Engagement, Social anxiety, and Achievement in Online Flipped Classrooms: a Structural Equational Modeling.*Education and Information Technologies*, *27*(8), 26.
10. Roche, A. I., Holdefer, P. J., &Thomas, E. B. K. (2022). College student mental health: Understanding changes in psychological symptoms in the context of the COVID-19 pandemic in the United States.*Current Psychology*, *43*,13041–13050.
11. Sharma, S. K., Telles, S., Gandharva, K., & Balkrishna, A. (2022). Yoga instructors' reported benefits and disadvantages associated with functioning online: A convenience sampling survey.*Complementary Therapies in Clinical Practice*, *46*, 101509.
12. Sousa, M., Cruz, S., Inman, R., Marchante, M., & Coelho, V. A. (2024). Bullying victimization and bullying perpetration, social anxiety, and social withdrawal in Portuguese adolescents: A reciprocal association model.*Aggressive Behavior*, *50*(4).
13. Strzelecka, A., &Zawadzka, D. (2023). The use of Chi-squared Automatic Interaction Detector (CHAID) analysis to identify characteristics of agricultural households at risk of financial self-exclusions.*Procedia Computer Science*, *225*, 4443–4452.
14. Tu, W. (2018). Analysis on the consumers'behaviors and willingness of online shopping for fresh livestockproducts.*Engineering Intelligent Systems*,*26*(2–3), 107–114.
15. Wu, J., Xiao, W., Liu, B., Yu, J., Zheng, K., Shi, Q., & Ren, Z. (2024). The mediating role of resilience and interaction anxiousness in the effects of physical activity on mobile phone addiction among Chinese college students.*Frontiers in Psychiatry*, *15*, 1–8.
16. Wu, P., Cao, K., Feng, W.,&Lv, S. (2024). Cross-lagged analysis of rumination and social anxiety among Chinese college students.*BMC Psychology*, *12*(1), 1–10.
17. Zabag, R., Rinck, M., Becker, E., Gilboa-Schechtman, Eva., & Levy-Gigi, Einat. (2024). Although I know it: Social anxiety is associated with a deficit in positive updating even when the cost of avoidance is Obvious.*Journal of Psychiatric Research*, *169*, 279–283.
18. Zhang, M., Zhang, C., Jiang, Z.,& Liu, Y. (2024). A latent profile analysis of social anxiety among Chinese left-behind children and adolescents: Associations with online parent-child communication and online social capital.*Child Abuse & Neglect*, *158*, 107102.
19. Zhang, Y. (2024). Data Mining: Analysis of the Influence of College Students' Daily Behaviours on Their Academic Performance. *Engineering Intelligent Systems*, *32*(2), 127–132.
20. Zou, Y. (2024). The Impact of Social Anxiety on College Students.*SHS Web of Conferences*, *193*(9), 1–5.