**International Journal of Cognitive Informatics and Natural Intelligence**
Volume 17 • Issue 1

# Foreign Language Anxiety of College English Teachers and Their Countermeasures

Qianqian Xie, Zhengzhou University of Industrial Technology, China*

## ABSTRACT

It is necessary for English teachers to grasp the causes of students' language anxiety and explore ways to avoid, reduce, and eliminate students' anxiety. This paper discusses the foreign language anxiety of college English teachers in classroom teaching, its possible causes, teachers' awareness of anxiety, and countermeasures. This paper introduces the composition of student affairs analysis system from data layer, analysis layer, application layer, and display layer and combines data warehouse and data mining technology to improve the functions of student information, teacher information, achievement information, course selection information, and course evaluation. On the premise of data mining and data information management, it realizes the construction and application of teaching management data analysis system, using classification model. Apriori algorithm improves the algorithm, uses big data technology to analyze data and design courses, and analyzes the inherent relationship between mental health problems and attributes.

## KEYWORDS

College English Teachers, Countermeasures, Foreign Language Anxiety

## INTRODUCTION

Second language acquisition research has traditionally focused on intelligence and language learning ability, thereby ignoring the impact of emotional factors on individual learning differences. This oversight leads teachers to overlooking students' anxiety about learning a foreign language, resulting in inadequate teaching and limiting the achievement of teaching effectiveness and objectives (Li et al., 2023). Research on the impact of student anxiety on foreign language learning can be traced back to the 1940s, with varying and sometimes contradictory results. Some studies indicated that foreign language anxiety promoted academic performance, others indicated a negative correlation between the two, while still others showed no correlation between the two.

This paper begins by addressing the causes of students' learning anxiety and making targeted adjustments. It introduces a creative approach by combining the "tree" structure improvement strategy with the enhanced Apriori algorithm in the teaching management data analysis system. This optimizes

DOI: 10.4018/IJCINI.335078
*Corresponding Author

and improves the original English teacher's foreign language teaching using two aspects: construction and big data mining. Notably, it significantly improves the execution efficiency of teaching datasets in different environments, indicating that the improved Apriori algorithm can effectively boost the algorithm's execution efficiency. This improvement has the potential of lessening students' anxiety, thereby improving the efficiency and quality of teachers' instruction.

## LITERATURE REVIEW

The study of students' anxiety with regard to learning a foreign language began in the 1940s (Tunç-Pekkan et al., 2023). International language education scholars mainly focus on test anxiety, also known as test-taking anxiety. Lien H.Y. pointed out that language anxiety is the fear and anxiety that learners have when they need to express themselves in a foreign language or a second language (Karakose et al., 2023). Sun B.Q. has done a lot of research on foreign language classroom environment anxiety, which is the focus of foreign language or second language anxiety research today (Li et al., 2023). According to Sun B.Q., language anxiety in foreign language learning encompasses "special psychological activities such as self-perception, views on foreign language learning, feelings of learning a foreign language and learning behaviors caused by the unique language learning process in foreign language classroom learning" (Cao et al., 2023). He divides foreign language classroom anxiety into communication anxiety, test anxiety and negative evaluation anxiety.

Another researcher interested in foreign language learning was the American psycholinguist Krashen, who developed a second language acquisition theory in the 1980s which consisted of five hypotheses, namely: acquisition learning differentiation hypothesis, monitoring hypothesis, input hypothesis, affective filtering hypothesis, and natural order hypothesis. The Input Hypothesis holds that in the process of foreign language learning, the difficulty of information input to learners should be slightly higher than their actual level, which is the famous "i+1" theory (Hassan et al., 2023).

Anxiety scales are widely used in the fields of psychology and education. Anxiety scales can help evaluate an individual's level of anxiety in specific contexts. By self-report or observing the responses of others, scales can provide objective measurement indicators, helping researchers and professionals understand an individual's level of anxiety. The foreign language classroom learning anxiety scale, which measures the breadth and depth of foreign language learning anxiety, includes three aspects of foreign language classroom learning anxiety (Thoyyibah et al., 2023). These are:

1. Communication fear, which refers to the anxiety caused by the inability to express one's meaning clearly while communicating in the target language;
2. Examination anxiety, that is, the anxiety that foreign language learners experience during tests and examinations; and
3. Fear of negative evaluation, which refers to the anxiety that foreign language learners experience due to the fear of embarrassment when learning or using a foreign language.

International research on foreign language classroom anxiety has a long history. Currently, research findings on this topic in various countries mainly focus on the following three aspects: the overall situation of foreign language classroom anxiety, the correlation between foreign language classroom anxiety and achievement, the causes and countermeasures of foreign language classroom anxiety (Prasanth et al., 2023).

In a foreign language classroom where listening and speaking in the target language is the main focus, communication between teachers and students, and between students, often evokes significant fear. The concept of "fear of negative evaluation" was first put forward by Mayumi K. and J. Hüttner. They stated that individuals are afraid of how others might evaluate them, especially if that evaluation could be negative, leading them to seek out ways to escape the evaluation occasion (Manikanta &

Rama, 2023). Guo Yongyu and others have defined anxiety as “an emotional state composed of feelings such as tension, anxiety and fear.” Typically, anxiety experiences are usually accompanied by physiological reactions such as rapid heartbeat, elevated blood pressure, shivering, sweating, dizziness, and so on (Manurung & Lubis, 2023). When students experience such drastic reactions, they will be unable to learn.

In the field of foreign language educational psychology, Zhang Qingzong has analyzed the relevant theories and models of foreign language learning and put forward some new models and methods of foreign language teaching (CheshmehSohrabi & Mashhadi, 2023). From the perspective of learners’ individual differences, the factors affecting foreign language learning are studied in three aspects: physiological differences, cognitive differences, and emotional differences (Baek & Doleck, 2023).

## RELATED MATERIALS AND METHODS

### Big Data Mining

Data mining is a non-trivial process of extracting potential, effective, novel, useful, and ultimately understandable knowledge from a large amount of noisy, incomplete, fuzzy and random data (Sankaran et al., 2023). It is the process of selecting an appropriate classification algorithm for the input training data to build a classifier (Liang et al., 2023) and of training necessary data to include data sets and their related class labels. The function of the classifier is to map the input data to the corresponding class label. It is assumed that each data in the training data is represented by a dimension attribute vector, and each data belongs to a predefined class label. This stage can be regarded as a process of building a function.

Bayesian classification is a classification algorithm based on Bayesian theorem, which can be used to handle supervised classification problems (Klimova et al., 2023). Bayesian classification is one of the commonly used classification algorithms in data mining, which can be used for data classification, prediction, recognition, and other operations (Alice Chen et al., 2023). In the field of data mining, Bayesian classification is commonly used for tasks such as text classification, spam filtering, and medical diagnosis (Batool et al., 2023). By establishing a probability model and inferring unknown information from known information, Bayesian classification can effectively handle a large number of features and complex relationships (Guleria & Sood, 2023). It can also adapt to constantly changing data by constantly updating prior probabilities. Bayesian classification assumes the description of the measured values of the attribute set is the class label. It is used to express the relationship between the two. In the training stage, each combination of sum is learned, so that the probability value is maximized as the class label.

Rule based classification is a technology that uses training data to extract a set of if-then rules as classifiers, and it then classifies unknown data sets based on this. Association rules are like expressions, in which association rules can be measured by support and confidence. Support determines the frequency in a given data set, and confidence represents the frequency in the contained transactions. They are defined as follows:

$$s\left(x \rightarrow y\right) = \sigma\left(x \cup y\right) \Big/ n \tag{1}$$

$$c\left(x \rightarrow y\right) = \sigma\left(x \cup y\right) \Big/ \sigma\left(x\right) \tag{2}$$

This represents the total number of transactions, the total number of occurrences of the itemset, and the total number of occurrences. The data is then given as input to the Reduce function, which performs the processing. After all the Reduce functions are executed, the user will get the full result of the job and return to the user a fraction of the final result of the job. The MapReduce computing model is shown in Figure 1.

## C4.5 Decision Tree Algorithm for Data Analysis

The C4.5 decision tree algorithm is a classic data analysis and machine learning algorithm used to construct decision tree models (Nosrati et al., 2023). It was proposed by Ross Quinlan in 1993 and is an extended and improved version of the ID3 algorithm (Castelletti et al., 2020). The basic idea of C4.5 algorithm is to construct a decision tree model that can best classify or predict target variables by analyzing the features in the dataset (Liu & Tang, 2023). This algorithm is suitable for classification problems and can handle data containing discrete and continuous features.

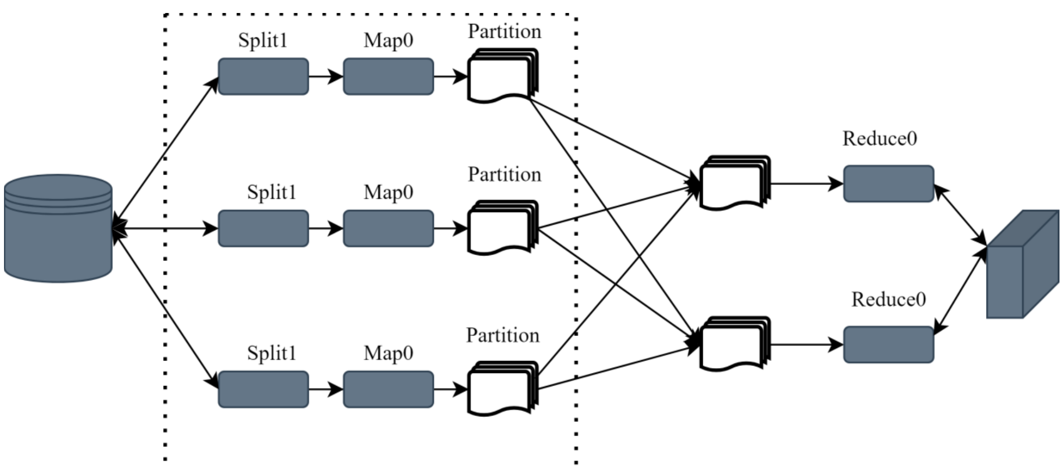
The main steps of the C4.5 algorithm are as follows:

1. Feature selection: Based on information gain or information gain ratio criteria, select the best feature as the partitioning feature for the current node. Information gain measures the degree of uncertainty reduction of the target variable under given feature conditions.
2. Building a decision tree: Divide the dataset into subsets based on the selected features and recursively construct subtrees. For each subset, repeat steps 1 and 2 until the termination condition is met, such as all samples belonging to the same category or having no more features to choose from.
3. Pruning: To prevent overfitting, pruning techniques can be used to prune the constructed decision tree. By deleting some subtrees or merging adjacent nodes, the decision tree can be simplified and generalization performance can be improved.

The advantages of the C4.5 algorithm when compared to ID3 algorithm mainly includes two aspects:

1. Processing continuous features: The C4.5 algorithm can process continuous features by discretizing them into several discrete value intervals and treating them as discrete features for processing.

**Figure 1. MapReduce calculation model**



2. Processing missing data: The C4.5 algorithm can handle samples containing missing data. By estimating the possible values of missing data and weighting them according to probability, the integrity of the dataset is maintained.

The C4.5 decision tree algorithm has high interpretability and wide application in data analysis. It can be used for tasks such as classification, prediction, and feature selection. By constructing a decision tree model, we can extract rules and patterns from the data, helping us better understand the data and make decisions. At the same time, the C4.5 algorithm also provides a foundation and reference for other decision tree-based algorithms, such as random forests.

### Distributed Storage and Computing Technology

In the context of big data, the distributed storage system has become a very popular and inevitable means of storage (Gaskins et al., 2023). Since the size of the data set often exceeds the storage capacity of a single machine by dozens or hundreds of times, the method of partitioning and storing on each separate computer must be adopted. This storage mode is called distributed storage, and the system that can realize this storage mode is a distributed storage system.

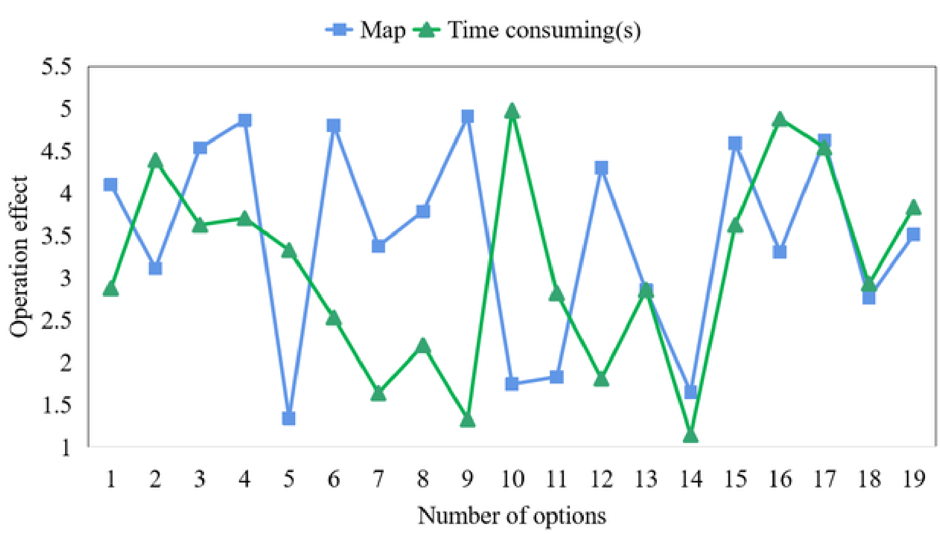
Logistic regression analysis is used to study the influence of the independent variable data set on the dependent variable data set. The data type can be either categorical (unordered categories) quantitative (ordered values). However, the choice of data analysis method depends on the number of options within the dependent variable.

In the logistic regression analysis model, a nonlinear relationship between the dependent variable and the independent variable can be expressed as:

$$p\left(y=1 \mid x\right)=\frac{e^{z}}{1+e^{z}} \tag{3}$$

Figure 2 shows the experimental results when the above operation effect is recorded and the calculation time is counted.

**Figure 2. Algorithm MAP value and time-consuming comparison**



### Centralized Monitoring Information Extraction for Big Data Mining

The sampling data of the signal centralized monitoring system is divided into normal data and abnormal data (Kamakshamma & Bharati, 2023). The division is based on whether the equipment is running normally or not. Once the fault information occurs, the system will extract the abnormal data deviating from the normal mode, analyze it, and judge the cause of the fault. Data understanding collection is based on understanding the knowledge representation of data mining schemes, thus forming data sets used in data mining. Data objects can be represented from three aspects. Namely, when the maximum likelihood method is used to solve this likelihood function, we get a cross-entropy loss function:

$$J\left(\theta\right)=\frac{1}{T}\sum\nolimits_{t=1}^{T}\sum\nolimits_{c\leq j\leq c,j\neq 0}\log\left(\omega_{t}+_{j}|\,\omega_{t}\right) \tag{4}$$

$p\left(w_{t}+_{j}|\,w_{t}\right)$ in the formula is a normalized probability in the whole dictionary. However, in the log analysis system, our log sequence does not have strong timing, and the log sequence reflects the user goal orientation in this period of time. Therefore, the objective letter of word2vec needs to be improved as follows:

$$\frac{1}{t}\sum_{t=1}^{T}\sum_{j\geq M_{tf\_}idf}\log p\left(\omega_{t+j}\mid\omega_{t}\right) \tag{5}$$

Among them is the TFIDF weight threshold of the log sequence. By calculating the TFIDF weight of each log sequence, it is assumed that the original training window is c, which is the minimum value of the largest C logs in the sequence TFIDF. We can use this to define the distribution density. The calculation method of the distribution density is to first take v consecutive line segments from to, and then calculate the distribution density of the valid data distributed on it, as shown in Equation (6):

$$p\cdot ij=\frac{\sum\nolimits_{p=i}^{j}N_{p}}{\sum\nolimits_{p=i}^{p=j}l_{mp}}\left(i=j-v,0<j\leq m-v\right) \tag{6}$$

Type-the length of the line segment; —— all effective points on the line segment.

Prediction accuracy P indicates the probability that the client may like the items in the recommendation list and it can be used to indicate the accuracy of the recommendation system. The recommended prediction accuracy of the system is:

$$p=\frac{1}{m}\sum_{u=1}^{m}p_{u}=\frac{1}{m}\sum_{m}^{m}\frac{|\,rl_{u}\cap tl_{u}\,|}{rl_{u}} \tag{7}$$

where it is the total number of partitions in the data set; it is subscript; it indicates all items with positive feedback predicted by the model in the data set; it indicates all items in the dataset that are actually positive feedback.

Data preparation requires certain operations to improve data quality and make the existing form of the dataset suitable for data mining. These operations generally include: continuous attribute discretization, attribute selection, data cleaning, data completion, data integration, and so on.

## Classification Model Predicts English Teaching Efficiency

Students' test scores are not only an important indicator to examine students' learning achievements in one semester, but also a key indicator to evaluate the quality of teachers' instruction in that semester (Suki Antely et al., 2023). Besides, test scores can also be used to as feedback on the influence of various teaching activities on students' scores, they can help schools to carry out teaching research, and help improve teaching strategies (Merlin & Prem, 2023). An Apriori algorithm can be used to better analyze test scores.

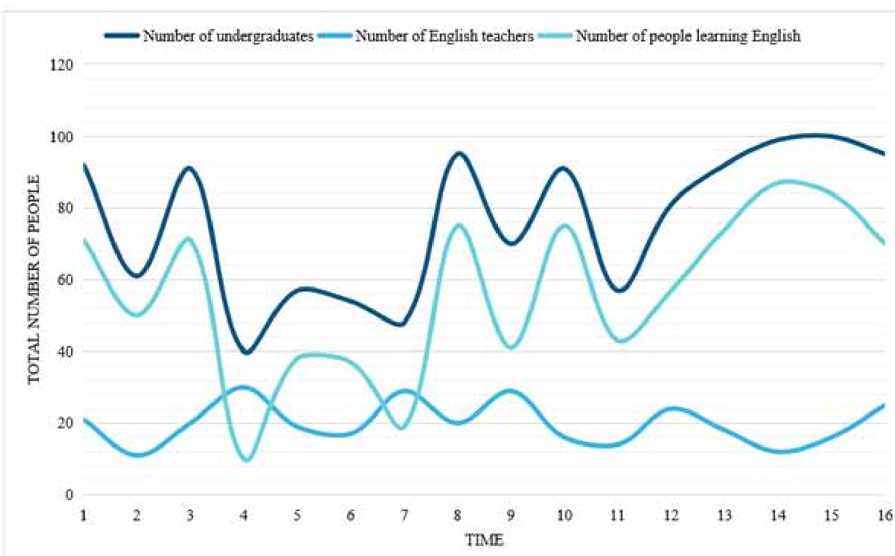
In practical applications, the Apriori algorithm contains a set of zero or more items and represents association rules through the expression "item 1 → item 2". After determining that the support and confidence are greater than the minimum threshold, the algorithm calculates the improvement value between two itemsets to determine whether the rule is useful (Onwuegbuzie et al., 1999). The calculation formula for lift is as follows:

$$lift\left(a \to b\right) = \frac{c\left(a \to b\right)}{s\left(b\right)} \tag{8}$$

where $s$ is the support of itemsets $a$ and $b$ in the confidence of $c$ at the same time.

An itemset needs to analyze the data frequency, and based on the frequency of the data set, it then realizes the quantitative analysis of the confidence and support of the data. Some representative grid-based clustering algorithms are: STING, WaveCluster, CLIQUE, and so on. Among them, the STING (Statistical Information Grid) algorithm uses grid-based multi-resolution clustering technology to divide the spatial area into a certain number of multi-level rectangular units, and the division of grid granularity will affect the final clustering effect. Figure 3 shows the data from the library for the number of English books borrowed in one day by students and English teachers, and the link between the ratios of the total number of books borrowed by all undergraduates and the books borrowed by English learners.

**Figure 3. The proportion of English learners vs. all undergraduates who borrowed books**



Comprehensively weighing these three indexes, the F-score (comprehensive evaluation) is selected as the new evaluation index. F-score is the harmonic average of $p$ and $r$ and its function is to balance the influence of accuracy and recall. The harmonic average of $p$ and $r$ is:

$$f = \left(1 + \beta^2\right) \cdot \frac{p \cdot r}{\beta^2 \cdot \left(p + r\right)} \tag{9}$$

where $\beta$ is used to adjust the weight. When $\beta = 1$, the weights of precision $p$ and recall $r$ are the same. At this time, the expression of comprehensive evaluation $f$ is:

$$f = \frac{2pr}{p + r} = \frac{2tp}{2tp + fp + fn} \tag{10}$$

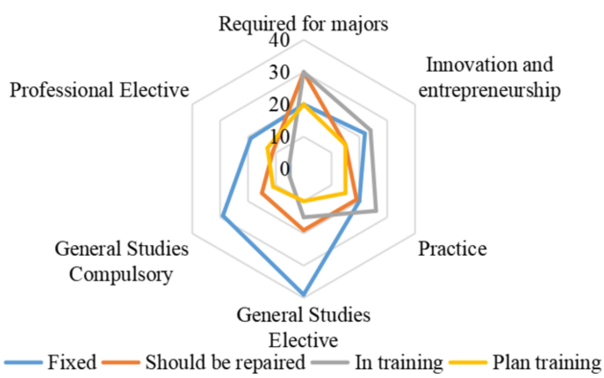
For some imbalanced datasets, some small sample categories are very important even though the number is small. Therefore, in order to balance the impact of precision and recall, the harmonic mean F-score of and F-score is selected to evaluate the performance of the model.

## ANALYSIS OF RESULTS AND DISCUSSION

This paper uses data mining techniques to analyze the anxiety levels of foreign language students, evaluate their mental health from a large amount of data, identify factors affecting students' learning of foreign languages, and to aid teachers in overcoming instructional challenges. Specifically, the study employs the C4.5 decision tree algorithm for analyzing a selected teacher's course data. The decision tree is trained and tested on the test data set using data fitting and learning methods, and the actual data is classified and predicted with the correct decision tree. Furthermore, the study considers the chronological arrangement in university curricula when extracting association rules, with the option of additional pruning operations based on time. Additionally, analyzing statistics on teachers' practices in different types of foreign language courses will help to provide valuable insights into their teaching approaches and situations. The student statistics from the course study are shown in Figure 4.

Figure 4 shows a metric that measures of English course learning, showing the credits that different types of students deserve and have earned. In the algorithm, the data set of anomaly mining

**Figure 4. Student course study statistics**



analysis is composed of three factor score tables, after data cleaning and data standardization, as shown in Figure 5.

According to the research and analysis of the improved algorithm's data in this paper, the average score of each item in the Foreign Language Classroom Anxiety Scale is calculated, and it is concluded that the score of each item is between 1 and 5 points. The final statistical results are shown in Table 1.

Based on the Apriori algorithm, students' grades are analyzed. Assuming that a database D contains students' grades in a class, the grades of 10 courses taken by 10 students are used, which are represented by T1, T2, T3, T4, T5, T6, T7, T8, T9, and T10, respectively, as shown in Table 2.

From the calculated results, it can be seen that there is a certain relationship between the impact of courses T1, T2 and T6, and according to confidence ({t1 t6}) = > {t2}=20%/30%=67%, if both T1 and T6 fail, the probability of failure of T2 is 67%. This is mainly because svm-aco-cf can find the global optimal parameters by using the ACO algorithm. Compared with svm-cf and k-rec sys CF, the

**Figure 5. Teachers' mental health test**

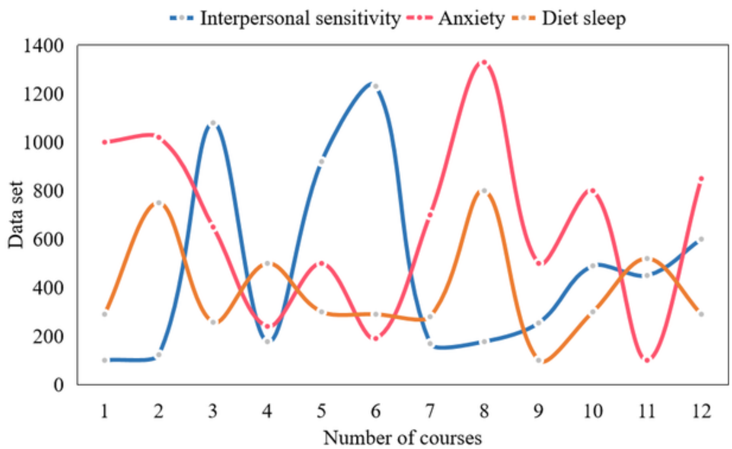



**Table 1. Language classroom anxiety scale**

| Topics of the Foreign Language Classroom Anxiety Scale | Score | Number of People Tested |
|---|---|---|
| I never felt confident speaking English in an English class. | 3.51 | 125 |
| I don't worry about making mistakes in English class. | 2.33 | 301 |
| I started to panic when I had to speak English in class without any preparation. | 4.51 | 156 |
| In English class, I get so nervous that I forget what I know. | 1.52 | 140 |
| I feel confident when I speak English in an English class. | 3.21 | 132 |
| I feel apprehensive when I don't understand what the teacher says in English. | 4.11 | 147 |
| I don't get nervous when talking to native English speakers. | 2.51 | 263 |
| I felt overwhelmed by the rules of the English language to learn. | 3.21 | 347 |
| I do not feel pressured to prepare for the English study. | 3.11 | 503 |
| I became both nervous and vague when speaking English in English class. | 1.23 | 406 |

Table 2. Survey data sheet

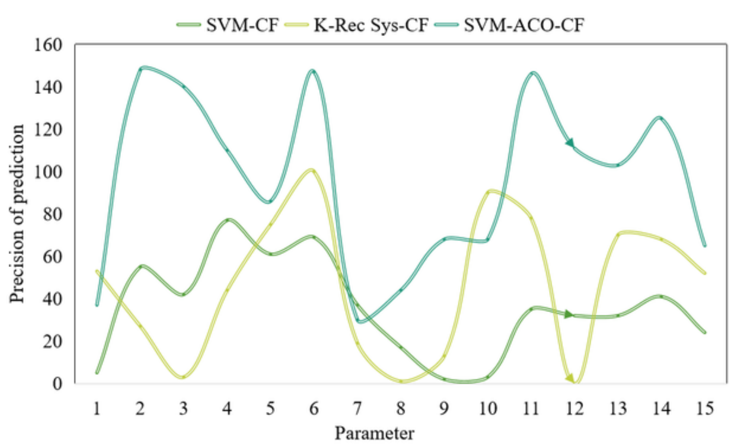
| Course | Support | Item Set | Fail Rate |
|---|---|---|---|
| T1 | 60.5% | 3 | 20% |
| T2 | 35.6% | 6 | 67% |
| T3 | 48.4% | 5 | 40% |
| T4 | 67.3% | 3 | 16% |
| T5 | 55.3% | 4 | 13% |
| T6 | 47.3% | 5 | 30% |
| T7 | 78.6% | 2 | 30% |
| T8 | 23.4% | 7 | 16% |
| T9 | 67.9% | 3 | 27% |
| T10 | 71.3% | 2 | 21% |

parameter setting of svm-aco-cf is more accurate, which improves the classification accuracy. Also, with the increase of the number of recommended items n, the prediction accuracy of the three different classifiers will also decrease. Figure 6 depicts the changes in prediction accuracy for each project.

Among them, the SVM-ACO-CF classifier has the highest prediction accuracy and K-Rec Sys-CF has the lowest prediction accuracy.

This paper analyzes student behavior through the steps of exploring student data, preparing student data, establishing a student model, and selecting a suitable prediction model, so as to realize the effective application of college student data. The selected model is evaluated through the three aspects of accuracy rate, F-score, and Kappa coefficient to ensure the balance of the selected student data and the accuracy of the built data model. During the process of data mining and data model establishment of the student data from Zhejiang Agriculture and Forestry University, the study found that there is a certain deviation between the model prediction and the experimental results. Some of the possible reasons for the prediction deviation could include incomplete feature set, insufficient data volume, and human factors.

Figure 6. Change of forecast accuracy of each project



This rule shows that students with excellent results in Physical Education also have excellent results in College English. However, objectively, this rule is not in line with reality, because the content of these two courses are not directly related, so the promotion degree of this rule lift<1. Finally, after filtering the table according to this criterion, the association rule table obtained is shown in Table 3.

It is evident that using the Apriori algorithm to calculate the support and confidence in association rules is a valid and accurate approach. However, during the process of calculation, numerous irrelevant rules may emerge. This error may be caused by issues with the score data, or it may be caused by the imperfect Apriori algorithm in the simulation process. Addressing these problems will require algorithm optimization in subsequent steps.

In this paper, a teaching management data analysis system is constructed and realized by mining the data of the teaching management data analysis system. Based on statistics pertaining to teachers' work status in the teaching affairs system, the data is analyzed regarding teaching quality, student recognition, and the application of teaching information. The teaching supervision department conducts data statistics through the educational affairs system. Based on the above association rule analysis, the data mining and data information processing of the teaching management data analysis system provides information support for making decisions regarding teaching management. These resources contribute to the enhancement of teaching management practices in colleges and universities.

## CONCLUSION

This study reveals the limitations of traditional second language acquisition research that has often overlooked the influence of emotional factors, particularly student anxiety, on foreign language learning. In the past, teachers often focused on only students' intellectual and language learning abilities while teaching and neglected the impact of student anxiety. This resulted in poor teaching outcomes and limited the achievement of teaching objectives.

**Table 3. Final mining table**

| ID | Rules | Support | Confidence | Lift |
|---|---|---|---|---|
| 1 | 13_a=>21_a | 0.43 | 0.51 | 1.03 |
| 2 | 21_a=>11_a | 0.55 | 0.69 | 1.02 |
| 3 | 13_a=>24_b | 0.43 | 0.38 | 1.53 |
| 4 | 15_a=>21_a | 0.23 | 0.52 | 1.04 |
| 5 | 3_f=>29_a | 0.61 | 0.44 | 1.07 |
| 6 | 13_b=>21_f | 0.33 | 0.52 | 1.08 |
| 7 | 12_a=>19_f | 0.45 | 0.33 | 1.03 |
| 8 | 20_a=>14_d | 0.64 | 0.67 | 1.05 |
| 9 | 13_b=>18_d | 0.43 | 0.56 | 1.02 |
| 10 | 15_a=>21_d | 0.51 | 0.43 | 1.09 |
| 11 | 1_b=>8_f | 0.32 | 0.28 | 1.06 |
| 12 | 4_d=>13_b | 0.23 | 0.37 | 1.03 |
| 13 | 5_b=>2_f | 0.41 | 0.46 | 1.07 |
| 14 | 1_f=>31_d | 0.57 | 0.35 | 1.01 |
| 15 | 18_b=>19_f | 0.35 | 0.47 | 105 |

This study proposes a strategy to optimize and improve the instruction methods of teachers who teach English as a foreign language by combining the "tree" structure enhancement strategy with the improved Apriori algorithm in the teaching management data analysis system. Specifically, the study advocates for the analysis and optimization of teaching data from two angles: construction and big data mining. This approach aims to enhance teaching efficiency and quality in the following steps.

Firstly, through the "tree" structure improvement strategy, we can better understand the causes of student anxiety and make targeted adjustments. This strategy involves making an in-depth analysis of students' anxiety, helping teachers understand its root causes, and adopting appropriate teaching strategies to reduce students' anxiety and improve learning outcomes.

Secondly, we will integrate the improved Apriori algorithm into the teaching management data analysis system to optimize the processing and analysis of teaching data. This improved algorithm can significantly improve the execution efficiency of teaching datasets in different environments, thereby better responding to large-scale teaching data mining tasks. By effectively processing and analyzing teaching data, teachers can more accurately understand students' learning status and needs. This allows them to promptly adjust teaching content and methods, thereby improving teaching efficiency and quality.

The results of this study indicate the feasibility of using these improvement strategies to optimize and improve teaching. The improved teaching management data analysis system can better meet the needs of teachers by providing accurate and real-time teaching data analysis and feedback. At the same time, it effectively alleviates students' anxiety, empowering teachers to better respond to students' emotional needs, thereby improving teaching effectiveness and learning outcomes. This study provides new ideas and methods for second language acquisition research and holds significant practical guidance for teaching.

## AUTHOR NOTE

The author has no conflicts of interest to disclose. The figures and tables used to support the findings of this study are included in the paper. This research was not supported by any funds. The author would like to sincerely thank those researchers whose techniques have contributed to this research.

## REFERENCES

Baek, C., & Doleck, T. (2023). Educational data mining versus learning analytics: A review of publications from 2015 to 2019. *Interactive Learning Environments*, *31*(6), 3828–3850. doi:10.1080/10494820.2021.1943689

Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, *28*(1), 905–971. doi:10.1007/s10639-022-11152-y

Cao, C. H., Liao, X. L., Jiang, X. Y., Li, X. D., Chen, I. H., & Lin, C. Y. (2023). Psychometric evaluation of the depression, anxiety, and stress scale-21 (DASS-21) among Chinese primary and middle school teachers. *BMC Psychology*, *11*(1), 209. doi:10.1186/s40359-023-01242-y PMID:37452365

Castelletti, F., La Rocca, L., Peluso, S., Stingo, F. C., & Consonni, G. (2020). Bayesian learning of multiple directed networks from observational data. *Statistics in Medicine*, *39*(30), 4745–4766. doi:10.1002/sim.8751 PMID:32969059

Chen, M. R. A., Hwang, G. J., Majumdar, R., Toyokawa, Y., & Ogata, H. (2023). Research trends in the use of E-books in English as a foreign language (EFL) education from 2011 to 2020: A bibliometric and content analysis. *Interactive Learning Environments*, *31*(4), 2411–2427. doi:10.1080/10494820.2021.1888755

CheshmehSohrabi, M., & Mashhadi, A. (2023). Using data mining, text mining, and bibliometric techniques to the research trends and gaps in the field of language and linguistics. *Journal of Psycholinguistic Research*, *52*(2), 607–630. doi:10.1007/s10936-022-09911-6 PMID:36125692

Gaskins, J. T., Fuentes, C., & De La Cruz, R. (2023). A Bayesian nonparametric model for classification of longitudinal profiles. *Biostatistics (Oxford, England)*, *24*(1), 209–225. doi:10.1093/biostatistics/kxab026 PMID:34296256

Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, *28*(1), 1081–1116. doi:10.1007/s10639-022-11221-2 PMID:35875826

Hassan, M. M., Zaman, S., Mollick, S., Hassan, M. M., Raihan, M., Kaushal, C., & Bhardwaj, R. (2023). An efficient Apriori algorithm for frequent pattern in human intoxication data. *Innovations in Systems and Software Engineering*, *19*(1), 61–69. doi:10.1007/s11334-022-00523-w

Kamakshamma, V., & Bharati, K. F. (2023). Adaptive-CSSA: Adaptive-chicken squirrel search algorithm driven deep belief network for student stress-level and drop out prediction with MapReduce framework. *Social Network Analysis and Mining*, *13*(1), 90. doi:10.1007/s13278-023-01090-z

Karakose, T., Polat, H., Yirci, R., Tülübaş, T., Papadakis, S., Ozdemir, T. Y., & Demirkol, M. (2023). Assessment of the relationships between prospective mathematics teachers' classroom management anxiety, academic self-efficacy beliefs, academic amotivation and attitudes toward the teaching profession using structural equation modelling. *Mathematics*, *11*(2), 449. doi:10.3390/math11020449

Klimova, B., Pikhart, M., Benites, A. D., Lehr, C., & Sanchez-Stockhammer, C. (2023). Neural machine translation in foreign language teaching and learning: A systematic review. *Education and Information Technologies*, *28*(1), 663–682. doi:10.1007/s10639-022-11194-2

Li, Q., Xie, Z., & Zeng, G. (2023). The influence of teaching practicum on foreign language teaching anxiety among pre-service EFL teachers. *SAGE Open*, *13*(1). doi:10.1177/21582440221149005

Li, X., Guan, X., Wang, J., Zhang, Y., Ma, M., Pu, T., Jobe, M. C., & Ahmed, M. Z. (2023). The effect of pre-service teachers' family adaptation on anxiety: A moderated mediating effect. *Sustainability (Basel)*, *15*(14), 10796. doi:10.3390/su151410796

Liang, J. C., Hwang, G. J., Chen, M. R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, *31*(7), 4270–4296. doi:10.1080/10494820.2021.1958348

Liu, K., & Tang, C. (2023). Privacy-preserving Naïve Bayes classification based on secure two-party computation. *AIMS Mathematics*, *8*(12), 28517–28539. doi:10.3934/math.20231459

Manurung, S. K., & Lubis, R. S. (2023). Effectiveness of decision support system for hydroponic plant nutrient selection using Apriori algorithm method. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 8*(1), 22-28. 10.33395/sinkron.v8i1.11975

Merlin, N. G., & Prem, M. V. (2023). Efficient indexing and retrieval of patient information from the big data using MapReduce framework and optimisation. *Journal of Information Science*, *49*(2), 500–518. doi:10.1177/01655515211013708

Nosrati, V., Rahmani, M., Jolfaei, A., & Seifollahi, S. (2023). A weak-region enhanced Bayesian classification for spam content-based filtering. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(3), 1–18. doi:10.1145/3510420

Prasanth, N. N., Karajgi, A., Sood, A., & Raja, S. P. (2023). Flood areas prediction in Bangladesh using Apriori algorithm. *International Journal of Hybrid Intelligence*, *2*(2), 119–127. doi:10.1504/IJHI.2023.129298

Sankaran, S., Sankaran, K., & Bui, T. (2023). Student satisfaction with R vs. Excel in data mining and business analytics: A Herzberg's motivation-hygiene theory perspective. *Decision Sciences Journal of Innovative Education*, *21*(2), 68–82. doi:10.1111/dsji.12285

Suki Antely, A., Jegatheeswari, P., Bibin Prasad, M., Vinolin, V., Vinusha, S., Rajakumar, B. R., & Binu, D. (2023). Sparse FCM-based map-reduce framework for distributed parallel data clustering in E-Khool learning platform. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, *31*(01), 1–23. doi:10.1142/S0218488523500034

Thoyyibah, T. (2023). Identifying factors affecting the relationship between department and graduation level of informatics engineering students using Apriori algorithm: A case study at Pamulang University. *Piksel: Penelitian Ilmu Komputer Sistem Embedded and Logic*, *11*(1), 159–170. doi:10.33558/piksel.v11i1.6933

Tunç-Pekkan, Z., Ölmez, İ. B., & Taylan, R. D. (2023). An online laboratory school research on pre-service mathematics teachers' experiences and mathematics teaching anxiety. *Education and Information Technologies*, *28*(5), 5739–5761. doi:10.1007/s10639-022-11404-x PMID:36373042

*Qianqian Xie, graduated from Henan Normal University with a master's degree, is currently a teacher and associate professor at the Vocational Education College of Zhengzhou Industrial Application Technology College. Her research focuses on English and American literature and English education. Her ORCID iD is 0009-0009-2363-7980.*