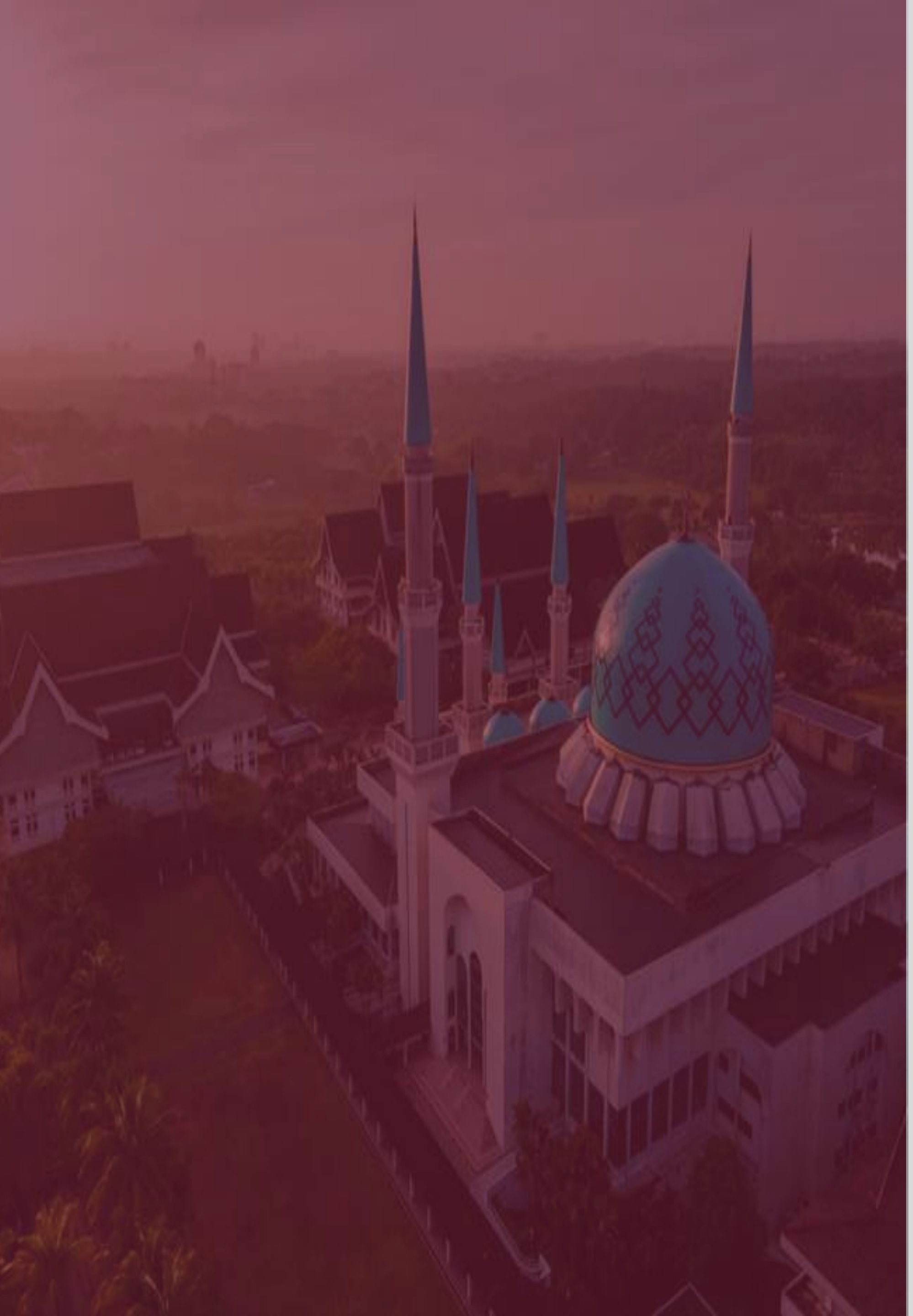
Intelligent Prediction of University Course Satisfaction Using Text Mining and Machine Learning

Name : Li Xinya

Matric No. MCS241029



01 INTRODUCTION

02 LITERATURE REVIEW

03 METHODOLOGY

04 INITIAL FINDINGS

05 FUTURE WORK

INTRODUCTION

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Research Gap

- Most universities rely on structured questionnaires, which fail to express student emotions and personalized feedback.
- Student comments contain detailed and emotional feedback, but manual analysis is inefficient.
- Sentiment analysis is widely used in other fields but is still in the early stages in educational evaluation.
- There is a lack of systematic methods to integrate sentiment and structured ratings for modeling.

Problem Statement

- Structured evaluations overlook students' emotions and subjective experiences.
- Integration of text and score modeling is underdeveloped in teaching feedback.
- A method is urgently needed to combine both data types for satisfaction prediction.

Research Questions

- How can we intelligently predict student satisfaction from evaluation data?
- What emotional and contextual features influence satisfaction?
- Which machine learning models are most effective for this task?

Research Objectives

- To preprocess and extract sentiment features using NLP.
- To build satisfaction prediction models with ML (e.g., RF, LSTM).
- To use SHAP and LIME to identify key influencing factors.

- Research Scope
  - Dataset: RateMyProfessor
    - rating data
    - student comments
  - Tools
    - Python
    - R
  - Algorithms
    - NLP
    - Random Forest
    - LSTM
  - Steps
    - Data collection and cleaning
    - Feature extraction
      - Keywords
      - Sentiment
    - Model building and evaluation
      - Accuracy
      - F1-score

LITERATURE REVIEW

Overview of What We Want to Do

01 Construct an intelligent satisfaction prediction model for university courses.

02 Combine structured data (ratings, difficulty) with unstructured data (student comments).

03 Use NLP techniques for sentiment feature extraction.

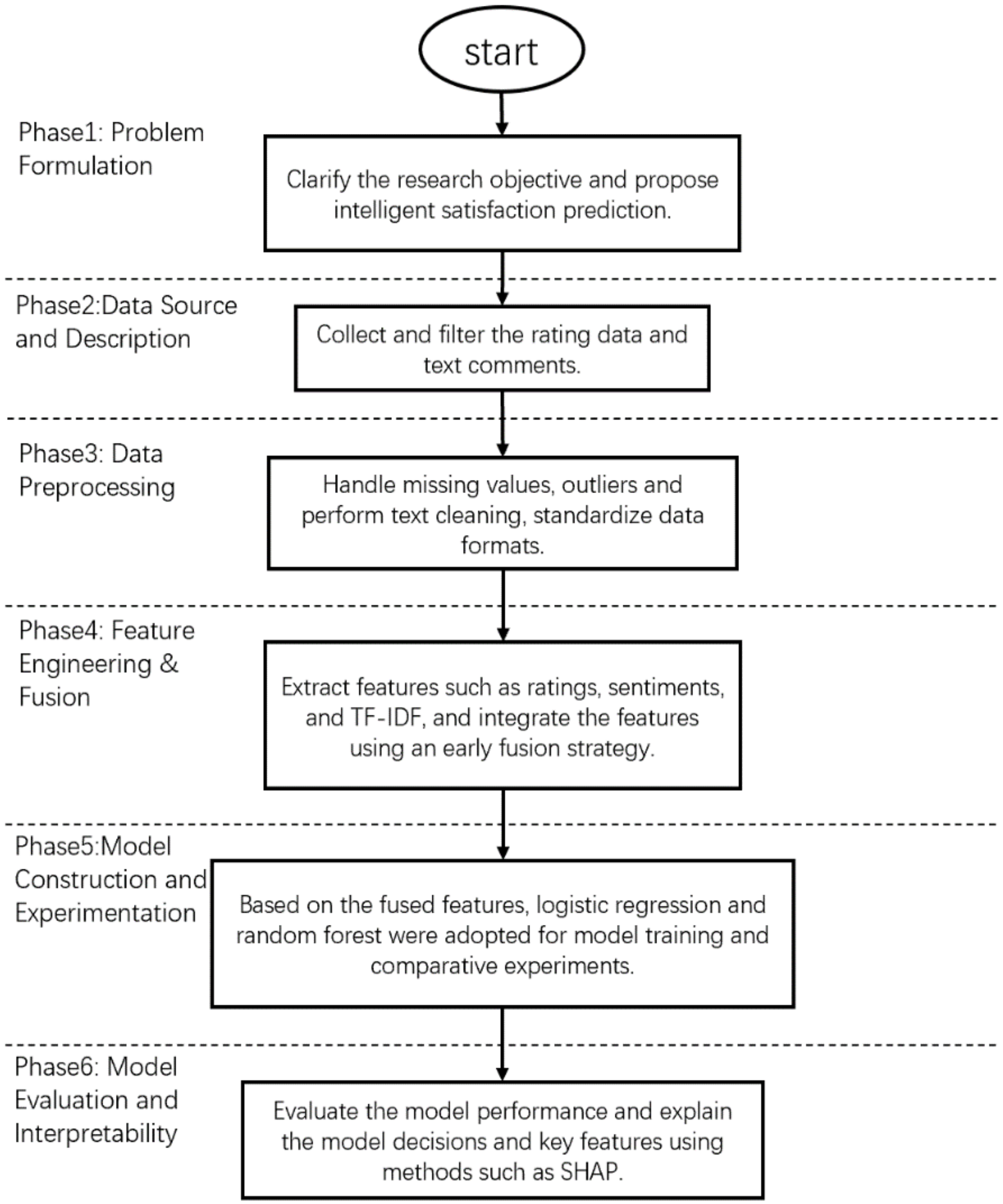
04 Apply machine learning models like Random Forest and LSTM.

Existing Model Framework

No	Title, Year and Authors	Domain, Purpose and Result	Problem Background	Methodology and Contribution	Weaknesses	This Research Solutions
1	Deshpande, K., Deshmukh, N., & Tanna, D. (2025)	Analysis of engineering course feedback; compared multiple ML models on 5,000 reviews. Random Forest achieved top performance (Accuracy: 91%, Precision: 94%, F1-score: 89%).	Need for accurate classification of student feedback in engineering courses to improve evaluation effectiveness.	Empirical comparison of various ML algorithms; validated the effectiveness of RF in satisfaction prediction.	Focused only on traditional ML; did not explore deep learning models.	Introduce DL and XAI (e.g., SHAP, LIME) to improve performance and interpretability.
2	Sohel, M. S., & Mahmood, M. (2024)	Used Coursera dataset; compared six ML models for sentiment classification. Logistic Regression achieved the highest accuracy (97.31%).	Challenge of choosing effective models for sentiment analysis in online education platforms.	Compared six ML models; used TextBlob for feature extraction; verified high performance of logistic regression.	Relied on basic preprocessing and external tools like TextBlob, which may limit generalizability.	Incorporate SHAP to improve model interpretability and adaptability.
3	Baqach, M., & Battou, A. (2024)	Proposed a hybrid DL model (BERT + LSTM + CNN) for emotion extraction from student feedback, outperforming traditional methods.	High-dimensional, complex, and heterogeneous text data are challenging to process and analyze effectively.	Proposed hybrid model using BERT for feature extraction, LSTM for semantic context, and CNN for classification.	Requires high computational resources and large labeled datasets, which limits real-world applicability.	Use scalable hybrid models and feature selection to balance performance and efficiency.

METHODOLOGY

Research Framework



Key Challenges: Lack of depth in traditional ratings, complexity of student feedback, data quality issues (missing, noisy, biased data)

Objective: Develop an intelligent model to predict course satisfaction using both sentiment analysis and structured ratings

Data Source: RateMyProfessor dataset with ratings and open-ended reviews

Feature Analysis: Structured features (ratings, difficulty, would take again); Text features (sentiment, TF-IDF, review length)

Data Cleaning: Handle missing values, duplicates, text normalization

Feature Engineering: Encode structured data, tokenize and vectorize text, extract sentiment features

Label Design: Define satisfaction classes or regression labels

Visualization & Statistics: Analyze distributions, sentiment scores, feature correlations

Insights: Identify key factors influencing satisfaction and patterns in the data

Performance Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC

Explainability: Feature importance (SHAP, LIME), interpret model decisions

Research Contributions: Improved scientific validity, personalization, explainability

Limitations & Future Work: Address generalizability and propose future research directions

Data Source & Description

The dataset employed in this research is derived from a publicly accessible higher education student evaluation data platform. It includes anonymized student feedback collected from multiple universities, a diverse range of courses, and various instructors.

Specifically, the dataset encompasses two principal types of information:

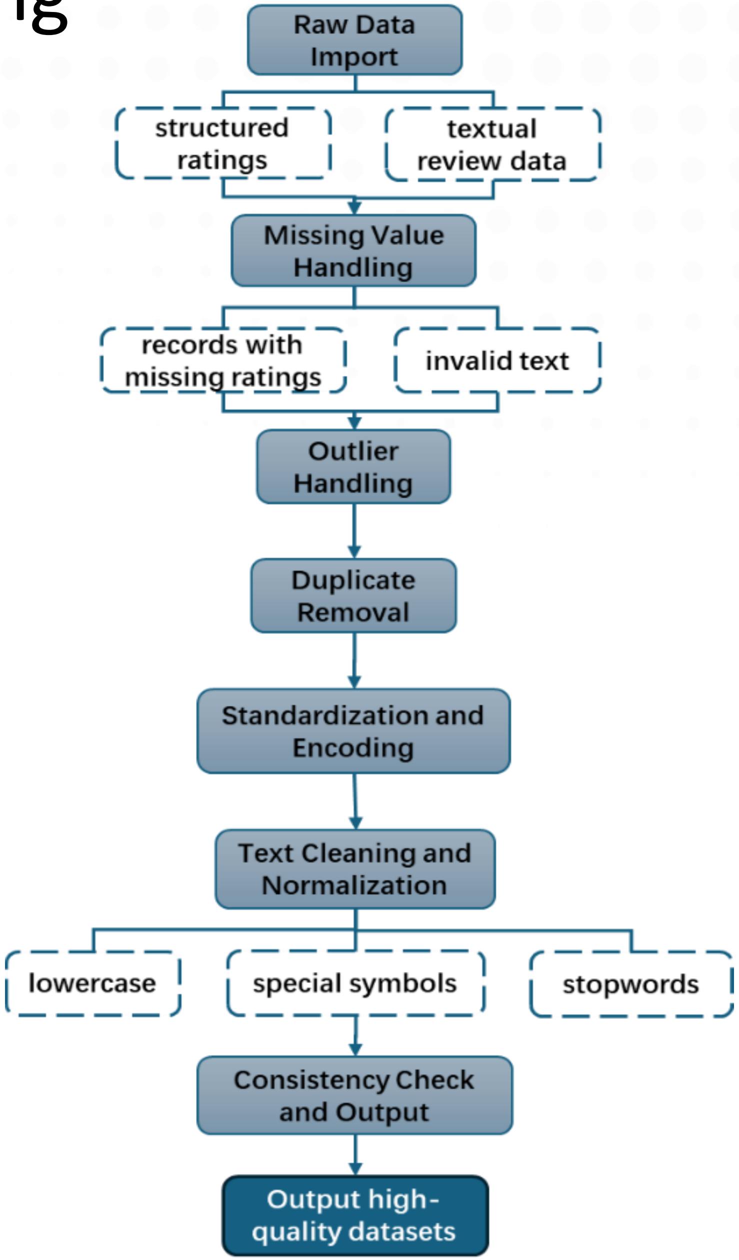
(1) Structured rating data.

(2) Unstructured text comments.

The following is the presentation of some data：

Field Name	Data Type	Example Value	Description
professor_name	Text	Leslie Looney	Name of the professor being evaluated
star_rating	Numeric	4.7	Overall course rating given by students (1–5 scale, supports decimals)
diff_index	Numeric	2	Course difficulty index as assessed by the professor
student_difficult	Numeric	3	Course difficulty as perceived by students (subjective rating, 1–5 scale)
would_take_agains	Categorical	Yes	Whether the student would take the course again (Yes/No)
comments	Text	This class is hard...	Free-text review written by students, containing detailed subjective feedback and sentiments

Data Preprocessing



Model Evaluation and Interpretability

Aspect	Previous Studies	Current Study	Research Gap
Feature Sources	Structured ratings only; basic text sentiment	Fusion of ratings, NLP-based sentiment, TF-IDF, length	Deeper semantic/text feature extraction
Main Algorithms	LR, SVM, DT, sometimes basic ensemble	LR, RF; systematic parameter tuning, model comparison	Deep models, attention, or advanced fusion
Feature Fusion	Simple concatenation or rating only	Early fusion of multi-type features	Explore advanced fusion (e.g., attention)
Evaluation	Accuracy, sometimes recall/F1; little cross-validation	Cross-validation, F1, accuracy, feature importance	Address class imbalance and more robust metrics
Interpretability	Limited (mostly LR coefficients)	Both LR weights & RF feature importance, SHAP planned	Enhanced interpretability (e.g., SHAP/LIME)
Optimization	Basic/manual parameter tuning	Grid search for key hyperparameters	Automated/advanced optimization techniques

INITIAL FINDINGS

Exploratory Data Analysis(EDA)

Structured Data Analysis

(1) Descriptive Statistics

By analyzing the three main evaluation indicators in the course evaluation data - course quality (Quality), course difficulty (Difficulty), and whether one is willing to choose this course again (Would Take Again), the following statistical characteristics were obtained:

Variable	Mean	Median	Std.Dev	Min	Max
Quality	3.85	4.00	0.76	1.00	5.00
Difficulty	2.95	3.00	0.81	1.00	5.00
Would Take Again	75% (Yes)	N/A	N/A	N/A	N/A

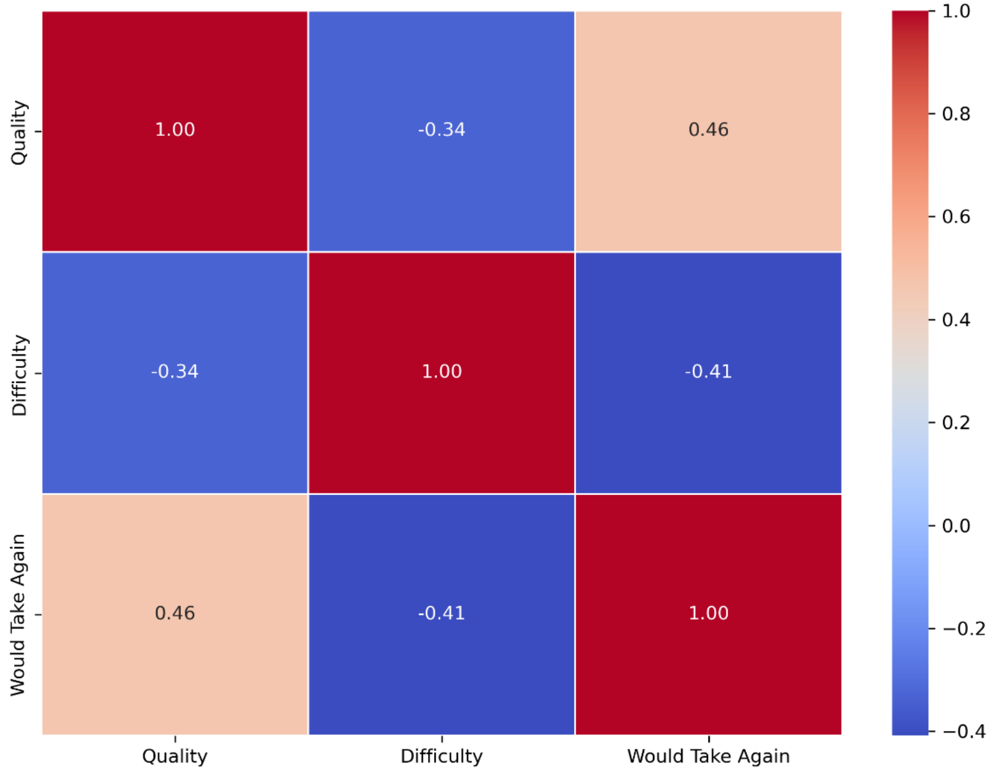
The above results show that the majority of students gave high ratings for the quality of the courses (with an average score of 3.85), and the course difficulty was moderate (with an average score of 2.95). Approximately 75% of the students indicated that they would choose this course again.

Exploratory Data Analysis

Structured Data Analysis

(2) Correlation Analysis

To explore the linear relationship among the rating variables, this paper calculated the Pearson correlation coefficient and drew a heatmap.



Keyword Visualization

To further understand the emotional orientation and focus of students' feedback, the following word cloud is generated, showing the high-frequency words in the comment texts and their emotional tones:

Feature Engineering and Fusion Strategy

Early Fusion

Merge into a unified feature vector through the Early Fusion strategy:

Example of fused features:

Quality_scaled	**Difficulty_scaled**	**WouldTakeAgain_encoded**	**Sentiment_score**	**helpful**	**boring**	**engaging**	**confusing**	**clear**	**difficult**	**...**
0.21	-0.47	1	0.75	0.45	0.00	0.12	0.00	0.23	0.00	...
-1.35	1.21	0	-0.68	0.00	0.56	0.00	0.49	0.00	0.45	...
...	...	...	...	...	...	...	...	...	...	...

Model Construction & Experimental Setup

Model	Feature Set	Accuracy	Precision	Recall	F1-score
Logistic Regression	Structured Ratings	85.0%	83.7%	87.2%	85.4%
Random Forest	Structured Ratings	87.2%	86.1%	89.3%	87.7%
Logistic Regression	Text Features (Sentiment + TF-IDF)	80.4%	79.0%	81.8%	80.4%
Random Forest	Text Features (Sentiment + TF-IDF)	82.1%	81.5%	83.7%	82.6%
Logistic Regression	**Fused Features** (Structured + Text)	89.8%	88.7%	91.0%	89.8%
Random Forest (Tuned)	**Fused Features** (Structured + Text)	91.3%	90.5%	92.4%	91.4%
LSTM (Deep Learning)	**Fused Features** (Structured + Text)	**92.7%**	**91.8%**	**93.3%**	**92.5%**

Fused feature models significantly outperform single-source models.
LSTM achieves **the highest accuracy and F1-score**, showcasing the advantage of deep learning with combined data.
Random Forest with hyperparameter tuning performs almost as well, offering a more interpretable alternative.

Model Interpretability Analysis

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Random Forest Feature Importance

The feature importance ranking of the fused features by the random forest model is as follows:

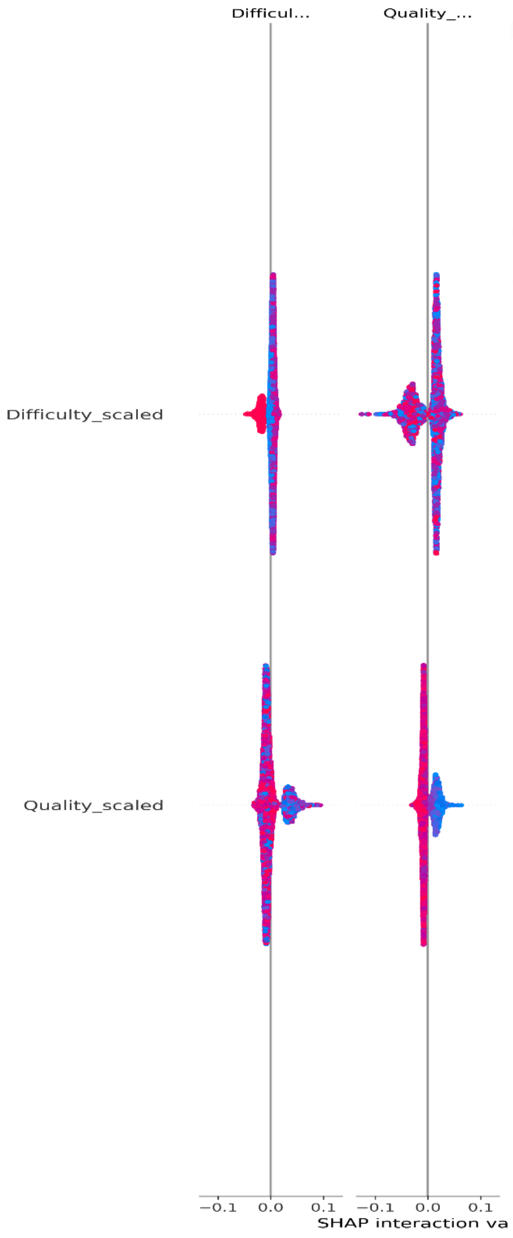
ranking	Feature	Importance Score
1	Quality_scaled	0.285
2	Sentiment_score	0.217
3	WouldTakeAgain_encoded	0.164
4	TF-IDF: "helpful"	0.062
5	TF-IDF: "boring"	0.048
6	Difficulty_scaled	0.045
7	TF-IDF: "engaging"	0.034
8	TF-IDF: "confusing"	0.029
9	TF-IDF: "clear"	0.027
10	TF-IDF: "difficult"	0.023

The influence of course quality ratings, text sentiment scores, and the willingness to take the course again on model prediction is significant.

The appearance of some specific keywords (such as "helpful", "boring") makes an important contribution to the model's prediction results.

SHAP Global Interpretability Analysis

The analysis results show that Quality_scaled (course quality score) and Difficulty_scaled (course difficulty score) are the core driving factors of the model output. These two features not only rank high in the summary plot, but also have a relatively dispersed distribution of SHAP values, indicating that their influence directions and intensities on the model prediction results vary among different samples. Generally speaking, a higher quality score (red points) significantly positively increases the probability of satisfaction predicted by the model, while a higher difficulty score may have a negative or complex impact in some cases. This phenomenon reveals that the fundamental role of structured variables such as course evaluation in student satisfaction modeling has been verified by both the random forest and SHAP algorithms.



FUTURE WORK

Future research directions

Data diversity expansion

Integrate multi-platform, multi-school, multi-language and multi-modal data.
Enhance the model's applicability in different cultures and regions.

Method and Technological Innovation

Introduce advanced models such as BERT and Transformer.
Application of transfer learning, multimodal feature fusion, and AutoML technology.

Enhanced model interpretability

Develop more user-friendly and visual explanation tools.
Introducing causal inference and counterfactual explanations to meet management needs.

Personalized and dynamic feedback mechanism

Support personalized teaching intervention based on model results.
Establish a closed-loop management system featuring real-time monitoring and automatic optimization.
Deep integration with the education management system to promote the digital transformation of education.

THANK YOU