CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion and Implications

5.1.1 Summary of the Performance of the Multi - source Feature Fusion Model

This study constructed and compared the performance of mainstream models such as Logistic regression, random forest, and LSTM in the task of predicting college course satisfaction based on structured scoring features, text sentiment features, and the fusion of both. The experimental results showed that there were significant differences in the predictive capabilities of the models corresponding to different feature inputs. Specifically, models that solely used structured scores (such as course quality, course difficulty, and willingness to take the course again) or solely used text sentiment features (such as VADER sentiment scores and TF-IDF keywords) achieved moderate prediction performance in terms of accuracy, precision, recall, and F1 score. However, after fusing the two types of features, the overall performance of both traditional machine learning models and deep learning models improved significantly.

For instance, when the LSTM model was only fed structured scoring features, its accuracy and F1 score were 87.2% and 87.7% respectively; while using only text features, they were 82.6% and 80.4%. When structured scores and text sentiment features were combined, the accuracy and F1 score of LSTM increased to 92.7% and 92.5% respectively. Similarly, the fusion feature versions of random forest and Logistic regression models outperformed their single-feature counterparts (as shown in Table 4.4.3). This result fully validates the significant value of multi-source feature fusion strategies in enhancing the generalization ability and prediction accuracy of models, especially in subjective and information-rich educational evaluation tasks such as course satisfaction.

5.1.2 Analysis of Key Features and Model Interpretability

To prevent the model from becoming a "black box", this study introduced interpretability tools such as feature importance analysis, SHAP (Shapley Additive Explanations), and LIME (Local Interpretable Model-Agnostic Explanations) to deeply analyze the specific impact of various features on the model's prediction results. Taking the random forest model with fused features as an example, the ranking of feature importance shows that the course quality score (Quality_scaled) and the text sentiment score (Sentiment_score) are the most core variables affecting the prediction of satisfaction, with importance scores of 0.285 and 0.217 respectively, far higher than other features. In addition, binary encoding of "whether willing to take again" and keywords such as "helpful", "boring", and "engaging" also occupy important positions.

The SHAP global interpretation results show that the positive increase of Quality_scaled and Sentiment_score can significantly increase the model's prediction probability of "satisfaction", while negative keywords such as "boring" and "difficult" significantly lower the prediction value. For example, in the case of LIME local interpretation, when a student's text comment is "helpful and clear, but a bit boring", the model will treat "helpful" as a positive feature and increase the prediction value, while "boring" is a negative feature and reduces the prediction value. Through the above multi-dimensional interpretive analysis, the model not only has strong predictive ability but also achieves the transparency of the prediction logic, which helps managers and teachers understand students' real feedback and provides data support and direction guidance for subsequent optimization measures.

5.1.3 The practical application of models in educational management

Based on the above-mentioned fusion model and the results of the interpretability analysis, the research findings of this paper have significant application value in the practical management of higher education. Firstly, the fusion model can accurately identify the key factors influencing student satisfaction, such as high-quality courses, positive emotional evaluations, and important positive and negative keywords. Through the quantitative analysis of course ratings, emotional scores, and high-frequency keywords in the text, educational administrators can not only "know which courses and teachers have high satisfaction", but also "know why", providing a scientific basis for course improvement, teacher motivation, and personalized feedback to students.

For instance, when a course is found to have "high difficulty and low emotional score", the management can adjust the course structure, reduce the difficulty, or enhance classroom interaction in a targeted manner; for courses with negative high-frequency words such as "boring", teachers can optimize teaching methods based on student suggestions to increase the course's appeal. Further, the interpretability output of the model helps to dynamically monitor student experiences, enabling early detection and intervention of problems, and promoting the formation of a continuous quality improvement mechanism based on data.

5.2 Limitations

Although this study has made positive progress in the prediction of college course satisfaction and the interpretability of the model, it still inevitably has certain limitations. Firstly, the research data mainly comes from the RateMyProfessor platform, and the data source and sample structure are relatively single, making it difficult to comprehensively reflect the diversity of different institutions, disciplines, and regions. The samples of spontaneous online evaluations are highly subjective, and extreme opinions are more likely to be recorded, leading to a skewed data distribution; some samples also have issues such as missing variables or invalid text. Additionally, sentiment analysis tools like VADER have certain limitations in adaptability and accuracy when dealing with educational jargon, polysemous words, slang, and complex contexts, which may affect the objectivity and effectiveness of sentiment feature extraction.

In terms of methods and experimental design, the generalization ability of the model needs further verification, especially lacking transfer tests on external datasets and in different cultural backgrounds. Deep learning models like LSTM have high requirements for computing resources and are not yet suitable for large-scale real-time deployment in actual teaching systems. During the feature fusion process, abnormal scores or meaningless texts can introduce noise and affect the stability of discrimination; sentiment analysis algorithms have limited processing capabilities for complex expressions such as irony and metaphor, which affects the recognition of true emotions in some samples. Moreover, the experimental design adopted relatively simplified strategies for parameter tuning and feature engineering, with a relatively single model evaluation index and mainly based on single stratified sampling, lacking robustness tests such as K-fold cross-validation or external independent data. The above limitations suggest that subsequent research should strengthen the integration of multi-source data, improve the accuracy of sentiment modeling, and continuously improve in aspects such as parameter optimization, model evaluation systems, and external validation to enhance the scientificity, applicability, and promotion value of the intelligent prediction model for course satisfaction.

5.3 Future Work

Although this study has achieved positive results in the intelligent prediction of course satisfaction, there is still room for further exploration and optimization in terms of data, methods, and practical applications. The future research directions can be summarized as follows:

a) Expansion of Data Diversity and Representativeness

The current data mainly comes from a single evaluation platform, with limited diversity and representativeness. In the future, data from multiple educational platforms, universities, or different countries and regions can be integrated, and feedback information in multiple languages and dimensions (structured data and unstructured text, audio, etc.) can be collected to enhance the model's universality and adaptability, and better apply it to diverse educational scenarios.

b) Innovation in Modeling Methods and Techniques

Although this study has integrated structured scores and text sentiment features, there is still room for improvement in deep modeling and multimodal feature fusion. In the future, more advanced deep learning architectures (such as BERT, Transformer, etc.) can be introduced, combined with transfer learning, multimodal fusion (such as text, voice, facial expressions, behavioral data, etc.) and automated feature engineering (AutoML) technologies to further enhance model performance and development efficiency.

c) Enhancement of Model Interpretability and Practical Application

The current model has improved transparency with the help of tools such as SHAP and LIME, but the interpretability and visualization reports tailored to the actual needs of educational management still need to be strengthened. In the future, more user-friendly and operational explanation tools can be developed, and methods such as causal inference and counterfactual explanations can be introduced to help managers deeply understand the key factors affecting satisfaction and convert model results into specific teaching improvement suggestions.

d) Personalized Application and Dynamic Feedback Mechanism

Subsequent research can focus on personalized teaching interventions based on model outputs, formulating targeted improvement measures for students, courses, and teachers with different satisfaction levels. At the same time, it is recommended to establish a dynamic monitoring and continuous feedback mechanism, automatically optimizing model parameters based on real-time data to achieve closed-loop management of satisfaction prediction, and explore in-depth integration with teaching management systems to promote the digital and intelligent transformation of educational management.

In conclusion, future research on intelligent prediction of course satisfaction will continue to develop in the directions of diversified data, intelligent methods, enhanced interpretability, and deeper practical application. Only by constantly breaking through existing limitations and closely integrating with educational reality can the scientificity, effectiveness, and management value of intelligent evaluation tools be improved, providing stronger support for educational decision-making and student growth.

5.4 Summary

This chapter provides a comprehensive summary and in-depth discussion of the research results, systematically analyzing the model's experimental performance, feature interpretability, practical management value, and research limitations. It also offers specific prospects for future research directions. In the task of intelligent prediction of course satisfaction, the integration of structured scores and text sentiment features significantly improves the model's prediction accuracy and practical applicability. By introducing explainability analysis methods such as SHAP and LIME, not only is the model's transparency and decision traceability enhanced, but it also provides data-driven precise decision support for educational administrators in colleges and universities.

At the same time, this study objectively reflects on the shortcomings in data sources, method selection, and experimental design, emphasizing the necessity of integrating multi-source heterogeneous data, method innovation, improving model robustness, and deeply integrating with actual educational scenarios. In the future, with the increase in data scale and diversity, as well as the continuous progress of artificial intelligence methods, course satisfaction prediction models will play a greater role in intelligent educational evaluation, personalized teaching intervention, and management decision-making.

Overall, the multi-source feature fusion and explainable modeling methods proposed in this paper not only effectively enhance the prediction ability of course satisfaction in colleges and universities but also provide a theoretical basis and technical path for promoting the scientific, data-driven, and intelligent management of education. It is hoped that the results of this study can provide useful references for subsequent related research and practical applications, and promote the continuous innovation and deep integration of educational big data and artificial intelligence in higher education.